The Role of Information in Marketing Strategy

by

Yunfei Yao

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor J. Miguel Villas-Boas, Chair
Professor Ganesh Iyer
Associate Professor Yuichiro Kamada
Assistant Professor Quitzé Valenzuela-Stookey

Spring 2023

The Role of Information in Marketing Strategy

Abstract

The Role of Information in Marketing Strategy

by

Yunfei Yao

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor J. Miguel Villas-Boas, Chair

This dissertation studies the role of information in marketing strategy. Specifically, it focuses on two key questions: How can consumers optimize their information search behavior to make better decisions? And in response to this, how can firms increase their profits by influencing consumers' information acquisition processes? The dissertation consists of three chapters of independent work.

Consumers frequently search for information before making decisions. Since their search and purchase decisions depend on the information environment, firms have a strong incentive to influence it. In the first essay, I endogenize the consumer's information environment from the firm's perspective. We consider a dynamic model where a firm sequentially persuades a consumer to purchase the product. The consumer only wishes to buy the product if it is a good match. The firm designs the information structure. Given the endogenous information environment, the consumer trades off the benefit and cost of information acquisition and decides whether to search for more information. Given the information acquisition strategy of the consumer, the firm trades off the benefit and cost of information provision and determines how much information to provide. This paper characterizes the optimal information structure under a general signal space. We find that the firm only smooths information provision over multiple periods if the consumer is optimistic about the product fit before searching for information. Moreover, if the search cost for the consumer is high, the firm designs the information such that the consumer will be certain that the product is a good match and will purchase it after observing a positive signal. If the search cost is low, the firm provides noisy information such that the consumer will be uncertain about the product fit but will still buy it after observing a positive signal.

When considering whether or not to buy a product, the consumer can often evaluate different attributes of it. Sometimes, the consumer chooses which attribute to search for because of exogenous reasons (e.g., one attribute is more important than others). However, the consumer often is unclear which attribute is more important ex-ante. Assuming that a product has

two symmetric attributes, the second essay characterizes the optimal search strategy of the consumer by endogenizing the optimal attribute to search, when to keep searching for information, and when to stop searching and make a decision. We characterize the search region by a set of ordinary differential equations for moderate beliefs and by a system of equations for extreme beliefs. We find that it is always optimal for the consumer to search the attribute about which she has the higher uncertainty due to the faster speed of learning. The consumer only searches for the more uncertain attribute if she holds a strong prior belief about one of the attributes, and may search for both attributes otherwise. We then show how firms can influence consumers' search behavior and increase profits by informative advertising. The firm does not advertise if the consumer's prior beliefs about both attributes are extreme. Otherwise, the firm advertises the better attribute if the consumer is optimistic enough about the worse attribute, and advertises the worse attribute if the consumer is less optimistic about it.

As consumers become increasingly concerned about their privacy, firms can benefit from committing not to sell consumer data. However, the holdup problem prevents them from doing so in a static setting. The third essay studies whether the reputation consideration of the firm can serve as a commitment device in a long-run game when consumers have imperfect monitoring technology. We find that a patient enough monopoly can commit because its reputation will be permanently destroyed if consumers observe the data sale. The persistent punishment provides the monopoly a strong incentive not to deviate. In contrast, reputation may fail to serve as a commitment device when there are multiple firms. The penalty for selling data is smaller as consumers cannot know which exact firm sold the data. Also, other firms can hurt the reputation of a particular firm even if it does not sell data. We find some sufficient conditions under which the incentive to deviate is so strong that firms lose the commitment power. Reputation failure in the presence of multiple firms persists when we consider endogenous or asymmetric monitoring.

To my parents

Contents

List of Figures

List of Tables

Acknowledgments

"Miguel!" During the dinner of my campus visit as an admitted student, one of the current students asked us, "Who is your favorite theorist of all time?" After some contemplation, one student chose Friedman, while another settled on Samuelson. My answer, however, emerged without any hesitation.

Six years later, I stand in profound gratitude to my dissertation chair, J. Miguel Villas-Boas. Your guidance surpassed my wildest expectations, providing a perfect balance of detailed advice and independence. Being your student has been an unparalleled privilege – an opportunity I wouldn't exchange for a Nobel Prize.

I must extend my heartfelt thanks to Ganesh Iyer, Yuichiro Kamada, and Quitzé Valenzuela-Stookey. Ganesh, the moment you emailed me about my admission to Berkeley will forever remain the happiest moment of my academic life. Your steadfast support, encouragement, and availability, even on weekends, have been truly invaluable. I feel comfortable talking to you about my research at any stage, and you have imparted so much wisdom over the years. Yuichiro, your astute guidance has significantly enriched my research. I am genuinely grateful for your willingness to delve deeply into the model during our enlightening conversations. Each time we discuss research, I invariably expand my understanding and discover new perspectives. Quitzé, your commitment to organizing theory activities has invigorated the economics theory community at Berkeley. I have also gained invaluable knowledge from your insightful comments during both individual and group meetings. I wish your arrival at Berkeley had been sooner.

I am indebted to numerous faculty members for your substantial contributions to my Ph.D. journey: David Ahn, Giovanni Compiani, Federico Echenique, Haluk Ergin, Brett Green, Przemyslaw Jeziorski, Shachar Kariv, Zsolt Katona, Kei Kawai, Sarah Moshary, Olivia Natan, Leif Nelson, Chris Shannon, Steve Tadelis, and Dong Wei. Your perceptive feedback and unwavering support have been instrumental. I am also grateful to Stanford economics faculty members Gabriel Carroll, Paul Milgrom, Michael Ostrovsky, Yuliy Sannikov, and Ilya Segal for welcoming me into your exceptional theory classes, which laid a solid foundation for my research.

My deep appreciation goes to my recommenders Eva Ascarza, Puneet Manchanda, Ya'acov Ritov, and especially Fred Feinberg. Your support has been pivotal to my academic journey, and you have consistently served as my inspirational role models. Fred, you played a decisive role in my admission to Berkeley. Your faith in me since my undergraduate days has truly shaped my life. I cannot fathom my academic trajectory without your influence.

I extend my warmest thanks to Melissa Hacker, Adam Lim, and Lisa Sanders Villalba for your continual support and dedication to ensuring a smooth and enjoyable experience at Berkeley.

I am grateful to my fellow students and friends. Your collaboration in discussing research, providing feedback on my drafts, and assisting with job market preparation has been invaluable.

I would like to acknowledge the financial support from J. Miguel Villas-Boas, the Ph.D. program office, the Fisher Center, the Sylff fellowship program, the Journal of Industrial Economics Fellowship, the Institute for Business Innovation, and the Center for Equity, Gender, and Leadership. Your contributions have made my life easier.

Above all, I owe my deepest gratitude to my parents and girlfriend. Without your unconditional love and support, none of my achievements would be possible. I love you!

Berkeley has left an indelible mark on my life, and I eagerly anticipate returning many times in the future.

Chapter 1

Dynamic Persuasion and Strategic Search

1.1 Introduction

With the rapid proliferation of digital technologies and information channels, it is increasingly common for consumers to seek detailed information before making a decision. More information can lead to less uncertainty and improve decision-making. Since consumers' search and purchase decisions depend on the information environment, firms have a strong incentive to influence it. Advertisers want to choose the advertising content to raise consumers' awareness and interests. Platforms want to design the website to attract traffic. Thus, the consumer faces endogenous information environment. For example, a consumer considering purchasing a pair of shoes may search on the internet to find out whether or not the item matches his needs.[1] The seller can influence consumers' search and purchase decisions through various methods. She can bid for search advertising spots to persuade the consumer. The content can be informative, showing features about the item, or persuasive, without detailed information. By spending more money for advertising, the seller can communicate information to potential buyers more frequently. Even if the consumer does not make a purchase right after seeing an ad, the seller can keep persuading the consumer through retargeting. At the same time, the consumer spends time and effort to search. He will only search for more information if he anticipates enough gain from it.

Such communications are also ubiquitous offline. For instance, an individual looking for a new car often visits a local dealership to talk with the dealer. The dealer will highlight various characteristics of the car. If the consumer is interested, the dealer often offers him a test drive to have a better sense of the car. This is time-consuming, in contrast to the ease of searching for information about a product on the internet, but they are often not sure whether they will like a product based solely on the information online. In the case of

[1]We refer to the information provider (seller) as "she" and the decision-maker (consumer) as "he" throughout the paper.

car-buying, test driving and then haggling with the dealer can take several hours, but people who purchase the car after visiting the dealership usually are sure that they like the car. Motivated by these examples, this paper endogenizes the consumer's information environment from the firm's perspective. By considering consumer search and firm information provision simultaneously, we wish to explain why the information environment of consumer search differs across various scenarios. We want to understand when the firm prefers to provide noisy rather than precise information and when it prefers to communicate with consumers for a longer time.

The main contribution of this paper is to endogenize the consumer's search environment from the firm's perspective. We find that the firm provides information incrementally rather than only once if the consumer is optimistic about the product fit before searching for information. If the search cost for the consumer is high, the firm designs the information such that the consumer will be certain that the product is a good match and will purchase it right after observing a positive signal. If the search cost is low, the firm provides noisy information such that the consumer will be uncertain about the product fit but will still buy the product right after observing a positive signal.

Specifically, this paper considers a dynamic model where a receiver (consumer) makes a binary decision between action G (purchase the product) and B (outside option). There are two states, good (the product is a good match) and bad (the product is a bad match). A sender (firm) always prefers G and sequentially persuades the receiver to take that action. In contrast, the receiver only wishes to take action G if the state is good. Neither the sender nor the receiver knows the state initially but have a common prior belief about it. The receiver can incur costs to search for more information about the state. The updated belief helps him make decisions. If the receiver observes a negative signal, he knows that the state is less likely to be good and will not take action G without the arrival of new information. If the receiver observes a positive signal, he knows that the state is more likely to be good, and the expected payoff of taking action G increases.

The sender designs the information structure. Given the endogenous information environment, the receiver trades off the benefit and cost of information acquisition and decides whether to search for more information. The receiver is forward-looking and forms rational expectations of the sender's strategy. The sender can incur higher costs to convince the receiver to take action G with a higher likelihood. Given the information acquisition strategy of the receiver, the sender trades off the benefit and cost of information provision and determines how much information to provide. Therefore, the receiver and the sender simultaneously trade off the benefit and cost of information acquisition/provision.

This paper characterizes the optimal information provision strategy of the sender and the optimal information acquisition strategy of the receiver. There are two periods in the main model. In each period, the sender chooses the information structure, and the receiver chooses whether to search for more information or make a decision. We later extend the two-period model to an infinite-period model and show that the main insights extend to the richer model. Instead of looking at specific parameters of the search environment, we study the design of the information environment under general signal space and characterize the optimal information structure among all feasible information policies. We develop a constrained non-linear programming method to solve the sender's information design problem, because the widely-used concavification method due to Kamenica and Gentzkow (2011) cannot solve it when the receiver's particiaption is strategic.

In equilibrium, the sender induces the receiver to take action G immediately upon observing a positive signal. This way, the sender saves the expected search time and does not need to compensate the receiver for a higher expected search cost. The sender extracts all the surplus from the receiver when she provides information in both periods, while she may leave some surpluses to the receiver when she provides information in only one period. Information smoothing can also save the persuasion cost. So, the sender has an incentive to spread information provision over multiple periods. However, the longer expected search time of the receiver will discourage him from searching if the likelihood of getting a positive ex-post payoff is low. Hence, the sender only smooths information provision if the prior belief is high. If the search cost for the receiver is high, the sender designs the information such that the receiver will be certain that the state is good and will take action G right after observing a positive signal. If the search cost is low, the sender provides noisy information such that the receiver will be uncertain about the state but still takes action G right after observing a positive signal.

We compare the profit-maximizing structure with the efficient (social welfare maximizing) information structure. When the search cost is high, the optimal strategy of the sender also maximizes social welfare because the minimal amount of information to persuade the receiver to search does not depend on the sender's objective. When the search cost is lower, the two information structures are different because the sender does not internalize the receiver's welfare.

In the main model, the sender can commit to the current-period information structure but cannot commit to the information structure in the next period. In the extension, we study the implications of dynamic commitment. We find that the ability to commit to future strategies strictly benefits the sender when the search cost is high, while the benefit vanishes as the search cost approaches zero.

We also consider the implications of discounting. When the search cost is high, it is difficult to convince the receiver to participate. Providing enough information to persuade the receiver to search dominates the force of discounting. So, the sender's optimal strategy does not depend on the discount factor. When the search cost is low and the players become less patient, the present value of the second-period profit decreases and the first-period participation constraint becomes tighter. The sender has a stronger incentive to convert the receiver early. So, she provides less information in the second period and more information in the first period. Moreover, we consider asymmetric discount factors in a model where the sender is patient while the myopic receiver only cares about the current-period payoff. We find that both players are worse off if the receiver is myopic rather than forward-looking. This implies that the sender should try to better inform the receiver of the possibility of multi-period information revelation and gradual learning.

Lastly, we extend the two-period model to an infinite-period model and show that the main insights extend to the richer model. When the search cost approaches zero, the ability to smooth the information over more extended periods is valuable for the sender. In that case, she could obtain the equilibrium payoff as if the persuasion cost were zero. We also find that the receiver will decide within a bounded number of periods.

Related Literature

There is a large stream of literature on optimal information acquisition. In particular, consumer search has raised growing interest both theoretically (Stigler 1961, Weitzman 1979, Wolinsky 1986, Moscarini and Smith 2001, Branco et al. 2012, Ke et al. 2016, Ke and Villas-Boas 2019, and Jerath and Ren 2022) and empirically (Hong and Shum 2006, Kim et al. 2010, 2017, Seiler 2013, Honka 2014, Ma 2016, Chen and Yao 2017, Honka and Chintagunta 2017, Seiler and Pinna 2017, Ursu et al. 2020, Moraga-González and Wildenbeest 2021, Morozov 2021, Morozov et al. 2021, and Yavorsky et al. 2021). In the above papers, the information environment is exogenous. Several papers study consumers' endogenous information acquisition. The consumer chooses both the search rule and the information environment. In Zhong (2022a), the decision-maker gradually gathers information about one product. Poisson learning is optimal for him. In Guo (2021), the consumer sequentially search for information about multiple products and determines how much to evaluate each product. Because the strategy and the outcome depend on the information structure, other payoff-relevant parties (e.g., firms) have a strong incentive to influence it. We take this into account by having the firm endogenously determine the information environment of the consumer.

Some papers investigate the design of the search environment from the firm's perspective. In Dukes and Liu (2016) and Kuksov and Zia (2021), the platform or the seller select the search cost to influence the consumer's search strategy. In Villas-Boas (2009), Liu and Dukes (2013), and Kuksov and Lin (2017), the product line design of the seller impacts consumer's search decision. Villas-Boas and Yao (2021) consider the optimal retargeting strategy of the firm which advertises to consumers who have a high likelihood of considering the firm's product. By advertising, the firm increases the frequency of consumers' learning information and the ability to track consumers. In Zhong (2022b), the platform recommends relevant sellers based on match values and prices to consumers. The platform designs the search algorithm by picking the match precision and the relative importance between prices and match values. Comparing the welfare outcomes among information structures emphasizing different vehicle characteristics under the counterfactuals of their structural model, Gardete and Hunter (2020) find that emphasizing the vehicle's history and obfuscating price information improves both consumer and firm welfare. Mayzlin and Shin (2011) consider a setting where the consumer can obtain an exogenously given signal by searching for information about the product quality. They find that uninformative advertising may serve as an invitation for the consumer to search. In Yao (2022), the firm can affect consumers' belief through informative advertising before they engage in costly search. In related literature on choice overload, the seller determines the amount of information provided to the consumer (Kuksov and

Villas-Boas 2010, Branco et al. 2016). The decision of the seller affects the consumer's search cost.

While the above papers are related to our paper, there are substantial differences. Instead of looking at specific parameters of the search environment, we study the design of the information environment under a general signal space. We characterize the optimal information structure among all feasible information policies. In the existing literature, the consumer either fully observes the value of a product/attribute or gets a noisy signal from an exogenously specified distribution (e.g., a normal distribution). In contrast, in our paper, the firm determines the distribution of the signal. The optimal information structure can be asymmetric and may not correspond to standard distributions.

We use a belief-based method of modeling information provision, first introduced by Aumann and Maschler (1995) and Kamenica and Gentzkow (2011) in the Bayesian persuasion and information design literature.[2] The sender picks a mean-preserving spread of the prior belief as the posterior belief, which simplifies the analysis. Some papers have studied the persuasion problem where either the receiver or the sender incurs costs. For example, Ball and Espín-Sánchez (2022) study a persuasion problem in which the sender chooses from a restricted set of feasible experiments, and the experiment can be costly. In Degan and Li (2021), the sender's persuasion cost depends on the precision of the signal. Wei (2021) considers a static persuasion problem in which a rationally inattentive receiver incurs information processing costs. Jerath and Ren (2021) consider a static model in which the consumer chooses the optimal information structures, taking into account that he needs to incur a cost to search for and process the signals. Instead of directly providing information, the firm influences the consumer's information environment by imposing constraints on the precision of the signals. Berman et al. (2022) study the information design of the recommendation algorithms under endogenous pricing and competition. Gentzkow and Kamenica (2014) extend the widely-used concavification approach of Kamenica and Gentzkow (2011) to the setting where the sender's cost is posterior-separable. We contribute to this literature by allowing the receiver to search for information voluntarily. The concavification approach cannot be used if the receiver's participation is strategic. So, we instead develop a constrained non-linear programming method to solve the sender's information design problem in the presence of consumer strategic search. Because we consider a dynamic problem, different information structures may correspond to different forms of sender's objectives and receiver's participation constraints. To reduce the dimensionality of the problem, we first show that the optimal information structure must induce the receiver to take action G immediately after receiving a positive signal. This qualitative property greatly simplifies the problem as we can limit our attention to such information structures. Nevertheless, the constraints of the non-linear program of the sender consist of multiple variables, making the optimization problem challenging. To make the problem tractable, we transform the program into a set of constrained programs, each with one constrained variable. We then select the global solution by comparing the local solutions of each program.

[2] See Bergemann and Morris (2019) for a survey.

Ke et al. (2022) study how online platforms should design the information in the presence of consumer search. In their paper, the information impact both consumer search and targeted advertising and allow them to study the trade-off between sales commission and advertising revenue. The firm designs the information to manipulate consumer's belief prior to search. The consumer always process this information but can search for more information strategically given an exogenous information structure (full revelation upon search). Our contribution is to integrate the information provision with consumer search and fully endogenize the search environment.

While some recent papers extend the static setting of Bayesian persuasion to the dynamic one, the persuasion cost is usually zero or a constant (Ely 2017, Renault et al. 2017, Ball 2019, Che et al. 2020, Ely and Szydlowski 2020, Orlov et al. 2020, Iyer and Zhong 2021, Bizzotto et al. 2021). In our model, the sender incurs a persuasion cost, and the receiver incurs a search cost. Unlike most Bayesian persuasion literature, the receiver's participation is strategic. The sender may need to convince the receiver to search or speed up the receiver's learning by incurring a higher cost. Therefore, we can investigate the sender's optimal trade-off between the benefit and cost of information provision.

The remainder of the paper is organized as follows. Section 2 presents the main model. Section 3 characterizes the optimal information provision strategy and the equilibrium outcomes. Section 4 characterizes the efficient information provision strategy and summarizes the information distortion when the sender rather than the social planner designs the information structure. Section 5 includes several model extensions. Section 6 concludes.

1.2 The Model

States, Actions, and Payoffs

There are two players, a sender and a receiver, and two states, good (g) and bad (b). The receiver ultimately makes a binary decision between G and B. The sender wishes to persuade the receiver to take action G regardless of the state, while the receiver wishes to match the decision with the state (taking action $G(B)$ when the state is $g(b)$). The payoffs of the decision for the players are the following:

(sender payoff, receiver payoff)	action G	action B
state g	(p, v_g)	$(0, 0)$
state b	(p, v_b)	$(0, 0)$

The sender earns a positive payoff, $p > 0$, if the receiver takes action G. The receiver's payoff is positive if he takes action G when the state is g, $v_g > 0$, and negative if he takes action G when the state is b, $v_b < 0$. Both players get zero payoff if the receiver takes the action B (which can be thought of as an outside option). We assume without loss of generality that $v_g = 1 + v_b$.[3] Neither the sender nor the receiver knows the state initially but have a common prior belief about it, $\mu_0 := \mathbb{P}(\text{the state is } g) \in (0, 1)$. In each period $t \in \{0, 1\}$, the sender determines and commits to the information structure of the current period but cannot commit to the information structure in the future. The receiver can search for information (action S) before deciding. The information acquisition is costly but helps the receiver make better decisions. If the receiver chooses to search, he observes the realization of a binary signal $s \in \{0, 1\}$ that reveals some information about the state. By choosing $\mathbb{P}[s = 1|g]$ and $\mathbb{P}[s = 1|b]$, the sender uniquely determines the signal. We order the value of the signal such that $\mathbb{P}[s = 1|g] > \mathbb{P}[s = 1|b]$. Hence, $s = 1$ corresponds to a positive signal and $s = 0$ corresponds to a negative signal. The players update the belief about the state according to Bayes' rule after the realization of the signal.[4] The game ends whenever the receiver makes a decision (G or B). Figure 1.1 illustrates the timing of the game.

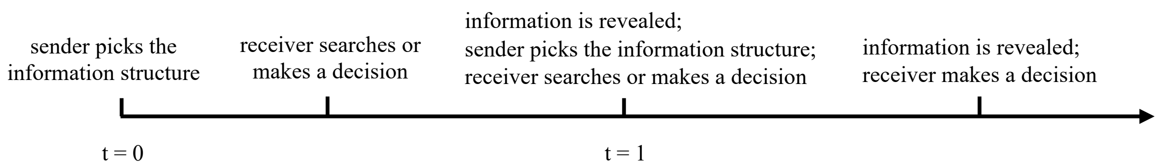


Figure 1.1: Timing of the Game

Analogous to Proposition 1 of Kamenica and Gentzkow (2011), we can work with mean-preserving posterior beliefs rather than the specific signal structure to simplify the analysis. Specifically, the existence of a binary signal is equivalent to the existence of a binary-valued posterior belief whose expectation is equal to the prior belief.[5] Denote the belief at the beginning of each period by μ_t. In each period, with probability λ_t, the receiver observes a positive signal and the belief increases to $\bar{\mu}_t$. We refer to λ_t as the probability of a positive signal and $\bar{\mu}_t$ as the belief after observing a positive signal. With probability $1 - \lambda_t$, the receiver observes a negative signal, and the belief decreases to $\underline{\mu}_t$. We refer to $\underline{\mu}_t$ as the belief after observing a negative signal.

We assume there is no discounting in the main model, and discuss the implication of discounting in the extension. Since the sender designs and provides information to the receiver and the receiver does not need to collect the information, different information should cost differently for the sender but not for the receiver. Therefore, we assume that the receiver incurs a flow cost of c per period of search. To persuade the receiver to take action G, the

[3] To see that assuming $v_g = 1 + v_b$ is without loss of generality, consider the following normalization. Let $v'_g = \frac{v_g}{v_g - v_b}, v'_b = \frac{v_b}{v_g - v_b}, p' = \frac{p}{v_g - v_b}$. Then, $v'_g - v'_b = 1$. We make this assumption to simplify the analysis and the presentation.

[4] We can assume without loss of generality that the sender also observes the signal realization. This is because the sender can perfectly infer the signal realization from the receiver's action under the optimal signal structure, according to Proposition 1.

[5] In the appendix, we state this result formally with proof.

sender wants to increase the receiver's belief about the good state. It is easy to provide information that increases the receiver's belief with a low probability but hard to do so with a high probability. It becomes impossible to always increase the receiver's belief. So, we assume that the sender's cost of information provision is increasing and convex in the probability of a positive signal, $K = K(\lambda)$. It is relatively cheap for her to provide information with a low λ. The marginal cost increases at an increasing rate as λ increases. The convex information provision cost is common in the literature (e.g., Robert and Stahl 1993).

Assumption 1. $K(\cdot) \in \mathcal{C}^2(\mathbb{R}_+), K'(\lambda) > 0, K''(\lambda) > 0,\ K(0) = 0,\ \lim_{\lambda \to 1^-} K'(\lambda) = +\infty,$ $\lim_{\lambda \to 0^+} K'(\lambda) = 0.$

Throughout this paper, we refer to the likelihood of a positive signal as the amount of information. More information means more frequent positive signals.

The total payoff for each player is the payoff of the decision net of the information provision/acquisition costs. The receiver is forward-looking and forms rational expectations about the sender's strategy in the future. To avoid the trivial case in which the sender provides no information and the receiver always takes the sender's desired action, we assume that $\mu_0 v_g + (1 - \mu_0)v_b < 0 \Leftrightarrow \mu_0 < -v_b$. So, the receiver will never take action G without searching. We also assume that the search cost is not too high, $c < v_g$. Otherwise, the receiver will never search.

Applications

Our model can be applied to many settings where a sender wishes to persuade a receiver to take a particular action while the receiver wishes to match the action with the state and can strategically gather more information before making the decision. We illustrate two applications of the model in this section.

Product Sales

A seller offers a product with unit demand. The price is p, and the marginal cost is m. The product may be a good match or a bad match with the buyer. Neither the buyer nor the seller knows the state initially but they have a common prior belief. The product is of zero value to the buyer if the state is bad and is of value v to the buyer if the state is good. The buyer decides whether or not to purchase the product. In this example, the sender is the seller and the receiver is the buyer. The payoffs of the decision for the players are the following:

(sender payoff, receiver payoff)	purchasing	not purchasing
good match	$(p - m, v - p)$	$(0, 0)$
bad match	$(p - m, -p)$	$(0, 0)$

The buyer can incur costs to search for more information about the product. The updated belief helps him make better purchasing decisions. If the buyer observes a negative signal, he knows that the product is less likely to be a good match and avoids wasting money on it without the arrival of new information. If the buyer observes positive news, he knows that the product is more likely to be a good match and the expected valuation increases. The buyer gets a positive surplus upon purchasing the good if the expected value is higher than the price.[6]

New Drug Launches

A pharmaceutical company develops a new drug and tries to get approval from the regulator (e.g., the FDA) by designing and conducting tests to convince the regulator that the drug is safe. If the regulator approves it, the firm gains an expected profit of p. The regulator gains a positive utility if the drug is safe and a negative utility if it is unsafe. In this example, the sender is the pharmaceutical company and the receiver is the regulator. The payoffs of the decision for the players are the following:

(sender payoff, receiver payoff)	approval	disapproval
safe	$(p, v_g > 0)$	$(0, 0)$
unsafe	$(p, v_b < 0)$	$(0, 0)$

Monitoring and examining the test is costly for the regulator. So, the regulator will only investigate the test if the benefit of doing so is high enough. If a single test fails to convince the regulator, the regulator may need additional information to make a decision.

Strategies and Equilibrium Concepts

Since the belief is common knowledge and there is no private information, we consider the sender-preferred subgame perfect equilibrium, as in Kamenica and Gentzkow (2011). If multiple actions (B, G, and S) give the receiver the same expected payoff, we assume that the receiver chooses an action that maximizes the sender's expected payoff. We also assume that the sender prefers to give the receiver more surplus in the first period if more than one equilibrium lead to the same expected sender's payoff. A perturbation of very little discounting justifies this assumption.

[6]One concern about this example is the possibility of product return. When it is costless to return the product, the buyer will always buy without search and return the product if it turns out to be a bad match. However, even though the buyer can get a full refund from many market places such as Amazon, he needs to incur time and effort to bring the package to the store or a shipping carrier. The qualitative properties of the main model still hold as long as a proportion of the buyers have enough return costs.

Equilibrium in the Second Period

The receiver has to make a decision between G and B at the end of the second period. Since it is costly for the sender to provide information, the sender will either give no information or provide enough information such that the receiver searches and will take action G if a positive signal arrives. We illustrate the belief evolution in the last period in Figure 1.2. Also, the distribution of the belief induced by the signal should be a mean-preserving spread of the initial belief: $\mathbb{E}[\Delta\mu] = 0$. In sum, the sender either does not provide information and obtains zero payoffs or takes into account the following constraints when designing the information structure:
(1) participation constraint:

$$\lambda_1[\bar{\mu_1} v_g + (1 - \bar{\mu_1}) v_b] = \lambda_1(\bar{\mu_1} + v_b) \geq c \tag{IR_1}$$

(2) feasibility constraint:

$$\lambda_1 \bar{\mu_1} + (1 - \lambda_1)\underline{\mu}_1 = \mu_1 \tag{F_1}$$

If the sender provides information, the constrained program of the sender is:

$$\max_{\lambda_1, \bar{\mu_1}} -K(\lambda_1) + p\lambda_1 \tag{P_1}$$
$$\text{s.t. } (IR_1), (F_1), \lambda_1 \in [0, 1], \underline{\mu}_1 \in [0, \mu_1)$$

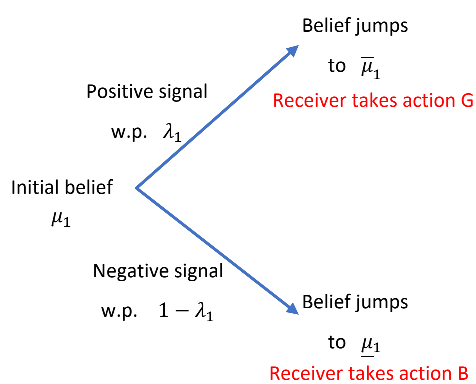


Figure 1.2: Belief Evolution in the Last Period

We analyze the solution to this problem in the next section. Though the information structure consists of $(\lambda_1, \bar{\mu_1}, \underline{\mu}_1)$, any two of them fully characterize the strategy because the third variable is then uniquely determined by (F_1). Therefore, we use $(\lambda_1, \bar{\mu_1})$, the probability of a positive signal and the belief after observing a positive signal, to represent the sender's strategy.

Equilibrium for the Entire Game

The sender has three options. Firstly, she can provide no information and obtain zero payoffs. Secondly, she can provide information in only one period. If the receiver decides to search, he will observe a one-shot signal designed by the sender. The receiver takes action G if a positive signal arrives and takes action B if a negative signal comes. The sender will not provide extra information regardless of the signal realization. Her problem is exactly (P_1). Lastly, the sender can provide information in both periods. She can give a pair of one-shot signals, which means that the receiver will take action G upon observing a positive signal in either period. If a negative signal arrives in the first period, the sender will provide another signal, hoping that a positive signal will arrive in the second period. The sender can also offer a pair of iterative signals. The receiver must search in both periods before taking action G under iterative signals. We illustrate the belief evolution of the one-shot signals and iterative signals in Figure 1.3 and 1.4.

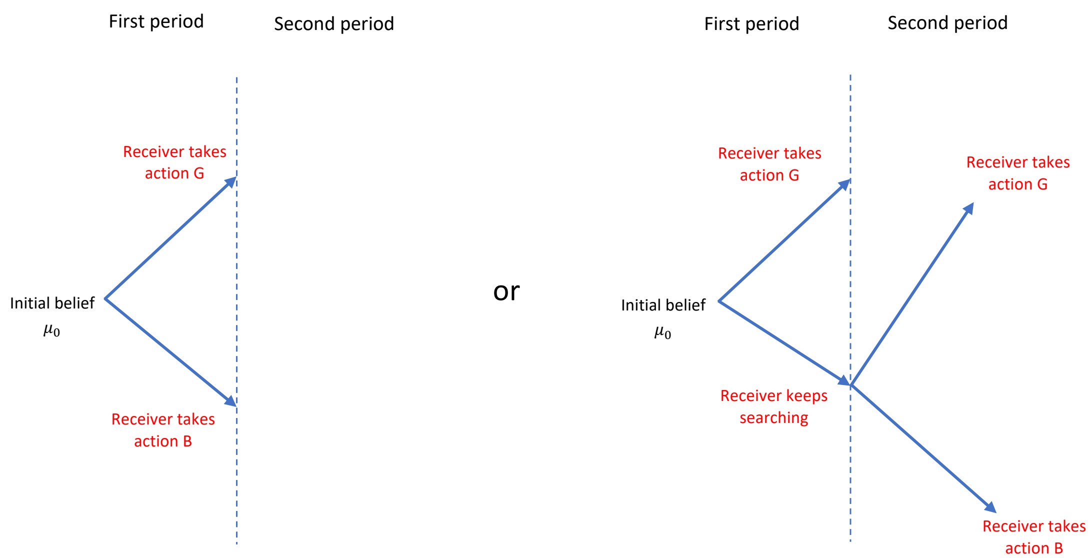


Figure 1.3: Belief Evolution of One-shot Signals
Left Figure: Sender Only Persuades in One Period; Right Figure: Sender Persuades in Both Periods

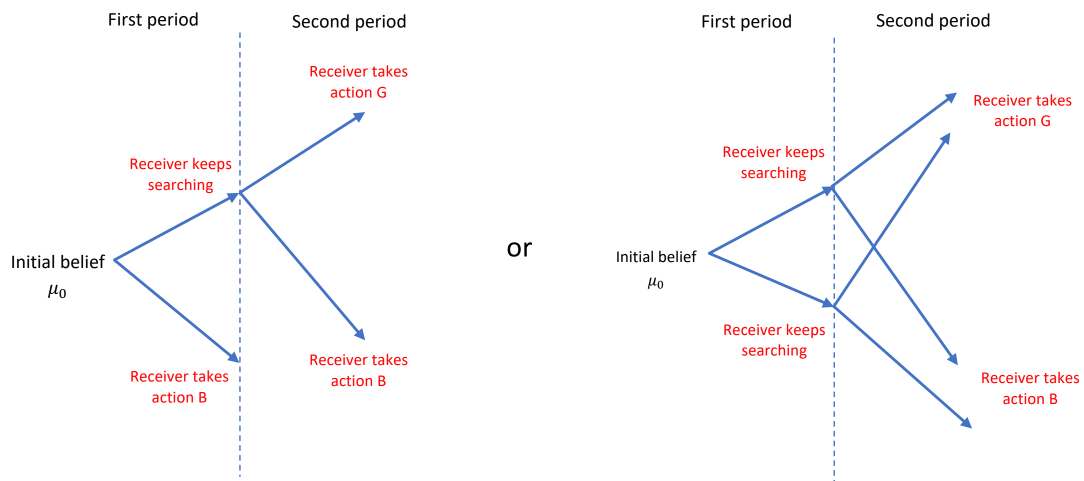


Figure 1.4: Belief Evolution of Iterative Signals
Left Figure: Receiver Keeps Searching Only after a Positive Signal; Right Figure: Receiver Searches in Both Periods

Compared to one-shot signals, iterative signals require a longer search time. To compensate the receiver for the higher expected search costs, the sender needs to provide information more favorable to the receiver, which hurts the sender's payoff. The following result shows that the sender always prefers one-shot signals in equilibrium. So, we limit our attention to the optimal one-shot signals.

Proposition 1. *For any pair of feasible iterative signals, there exists a one-shot signal that gives the sender a strictly higher payoff.*

The sender takes into account the following constraints when designing the optimal one-shot signal of the first period:
(1) participation constraint:

$$\begin{aligned}&\lambda_0[\bar{\mu_0} v_g + (1-\bar{\mu_0})v_b] + (1-\lambda_0)\mathbb{E}[\text{receiver surplus at } t=1|\text{search at } t=1]\\ =&\lambda_0(\bar{\mu_0}+v_b) + (1-\lambda_0)[\lambda_1(\bar{\mu_1}+v_b) - c] \geq c \qquad (IR_0)\end{aligned}$$

(2) feasibility constraint:

$$\lambda_0\bar{\mu_0} + (1-\lambda_0)\underline{\mu}_0 = \mu_0 \qquad (F_0)$$

If the sender provides information in both periods, her problem is:[7]

$$\max_{\lambda_0,\bar{\mu_0},\mu_1,\lambda_1,\bar{\mu_1}} -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\lambda_1) + p\lambda_1\right] \tag{P_2}$$
$$\text{s.t. } (IR_0), (F_0), (\lambda_1, \bar{\mu}_1) \text{ solves } (P_1)$$

We analyze the solution to this problem in the next section. We use $(\lambda_0, \bar{\mu_0}, \mu_1, \lambda_1, \bar{\mu_1})$, the probability of a positive signal in each period, the belief after observing a positive signal in each period, and the initial belief in the second period, to represent the sender's strategy.

1.3 Optimal Strategies

From the previous discussion, the receiver will search (take action S) whenever the sender provides information. The receiver will take action G immediately upon receiving a positive signal and action B if the sender does not provide information. Therefore, we only need to characterize the sender's strategy, which implies the receiver's strategy.

A Benchmark

We first consider a benchmark problem in which the receiver always participates and in which the sender can generate an arbitrary amount of information in each period. The sender chooses the information structure to maximize the expected payoff. We will use the solution throughout the subsequent analyses.

$$\max_{\lambda_0,\lambda_1} -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\lambda_1) + p\lambda_1\right] \tag{P_b}$$

Lemma 1. *The solution to the benchmark problem* (P_b), $(\lambda_0^{**}, \lambda_1^{**})$, *does not depend on the search cost* c *and* $\lambda_0^{**} < \lambda_1^{**}$. *The benchmark sender's payoff is strictly positive.*

The above payoff is the highest possible payoff the sender can obtain in equilibrium. When the prior belief is high enough, the sender obtains the benchmark payoff by setting the probability of a positive signal to λ_t^{**}. Since the sender can smooth the information provision at the beginning of the first period while only has one chance of providing information in the second period, she will choose $\lambda_0^{**} < \lambda_1^{**}$. When the prior is lower and a positive signal occurs with the benchmark probability, the sender needs to provide a very noisy signal (low $\bar{\mu}_t$) due to feasibility constraints. As a result, the receiver will be quite uncertain about the state even after observing a positive signal. So, he will choose not to search for information. This

[7] To simplify the notation, we omit the following common constraints in the main text in all of the programs: $\bar{\mu}_t \in [-v_b, 1], \underline{\mu_t} \in [0, \mu_t), \lambda_t \in [0,1], \underline{\mu}_0 = \mu_1$. The first constraint is required by one-shot signals. The second and third constraints ensure the information structure is well-defined. The last equality comes from the fact that the belief at the beginning of the second period, μ_1, is the belief after observing a negative signal in the first period, $\underline{\mu}_0$, under one-shot signals.

friction restricts the communication between the players and distorts the optimal strategy away from the benchmark strategy. For the problem to be non-degenerate, we concentrate on the case in which the prior is not too high throughout the paper. As a result, the optimal strategy is different from the benchmark solution.

Optimal Strategy in the Second Period

When the belief at the beginning of the second period is too low, the receiver will not search, given any feasible signals. Thus, the sender does not provide information to minimize the cost. When the belief at the beginning of the second period is higher, and the search cost is not too high, the sender provides information and obtains a positive payoff. The following proposition summarizes the optimal information structure.

Proposition 2. *In the second period, the sender does not provide information when* $\mu_1 < \mu_{0,1} := c/v_g$*. When* $\mu_1 \geq \mu_{0,1}$*, the optimal probability of a positive signal and the optimal belief after observing a positive signal,* $(\lambda_1^*, \bar{\mu}_1^*)$*, depend on the search cost c:*

1. *If* $c \geq v_g \lambda_1^{**}$*, there exists a unique* $\widehat{c} \in (v_g \lambda_1^{**}, v_g \mu_1]$ *such that the sender does not provide information if* $c > \widehat{c}$ *and* $(\lambda_1^*, \bar{\mu}_1^*) = (c/v_g, 1)$ *if* $c < \widehat{c}$*. The receiver gets zero surplus.*

2. *If* $c \in [\mu_1 + v_b \lambda_1^{**}, v_g \lambda_1^{**})$*,* $(\lambda_1^*, \bar{\mu}_1^*) = (\frac{\mu_1 - c}{-v_b}, \frac{-v_b \mu_1}{\mu_1 - c})$*. The receiver gets zero surplus.*

3. *If* $c < \mu_1 + v_b \lambda_1^{**} \wedge v_g \lambda_1^{**}$*,* $(\lambda_1^*, \bar{\mu}_1^*) = (\lambda_1^{**}, \frac{\mu_1}{\lambda_1^{**}} \wedge 1)$*. The receiver gets strictly positive surplus.*

When the search cost is too high, the sender has to provide a lot of information to persuade the receiver to search. Even if it is feasible for the sender to provide enough information that the receiver will search, it is so costly that the expected sender's payoff is negative. So, the sender chooses not to provide information, and the receiver does not search.

When the search cost is high but not too high, the sender will provide just enough information such that the receiver searches. Since the receiver's participation constraint is hard to satisfy, in equilibrium, the receiver becomes certain ($\bar{\mu}_1 = 1$) that the state is g after observing a positive signal. Suppose, instead, a positive signal does not fully reveal the state ($\bar{\mu}_1 < 1$). In that case, its arrival rate will need to be higher to persuade the receiver to search. Since the marginal cost of increasing the probability of a positive signal exceeds the marginal benefit, the sender's payoff decreases. The sender trades off the frequency of positive signal for precision.

When the search cost is intermediate, the receiver's participation constraint is easier to satisfy. Since the marginal benefit of increasing the probability of a positive signal exceeds the marginal cost, in equilibrium, the sender trades off the precision of a positive signal for frequency. The receiver is still uncertain about the state after observing a positive signal, but the belief is high enough that the receiver searches.

When the search cost is low, the information friction does not distort the information structure. The sender provides the benchmark amount of information, and the receiver gets a strictly positive surplus.

Optimal Strategy for the Entire Game

When the prior is too low, any feasible signal the sender can generate is not attractive enough for the receiver to search. Thus, it is impossible to communicate between the sender and the receiver. When the prior is higher, and the search cost is not too high, the sender provides information and obtains a positive payoff.

Proposition 3. *Suppose the search cost is not too high, $c < \widehat{c}$. There exists $\mu_{1,2} \geq c(2v_g - c)/(v_g)^2$ such that the sender does not provide information if the prior is low, $\mu_0 < \mu_{0,1}$, provides information in one period if $\mu_0 \in [\mu_{0,1}, \mu_{1,2})$, and provides information in both periods if $\mu_0 > \mu_{1,2}$. Suppose the sender provides information in both periods. A positive signal fully reveals the state being g when the search cost is high and partially reveals the state when the search cost is low. The receiver gets zero total surplus.*

The widely-used concavification approach (Kamenica and Gentzkow 2011) cannot be used to solve this kind of games because the receiver's participation is strategic. We instead develop a constrained non-linear programming method to solve the sender's information design problem. The constraints of the non-linear program of the sender consist of multiple variables, making the optimization problem challenging. To make the problem tractable, we transform the program into a set of constrained programs, each with one constrained variable. We then select the global solution by comparing the local solutions of each program.

Since the information provision cost is non-linear in the probability of a positive signal, the sender has an incentive to smooth the information provision over two periods. When the prior is low, it is not feasible for the sender to provide enough information in both periods so that the receiver will search whenever a positive signal has not arrived. As the prior increases, it becomes feasible for the sender to smooth the information provision. If the sender finds it optimal to provide information in both periods at a given prior, she also prefers to smooth the information for any higher prior.

When it is highly costly for the receiver to search, the optimal information structure fully convinces the receiver that the state is g when a positive signal arrives. In equilibrium, the receiver obtains the highest possible surplus conditional on observing a positive signal and making the purchase. Without providing this type of information, which is favorable to the receiver, the sender cannot persuade the receiver to search. In contrast, when it is less costly for the receiver to search, the optimal information structure adds some noise to a positive signal. In equilibrium, the receiver is not sure that the state is g after observing a positive signal. The state may be b after the receiver takes action G. However, the likelihood of state g after a positive signal is high enough to persuade the receiver to search. By adding some noise to a positive signal, the sender can provide more frequent positive signals and increase her payoff without violating the feasibility constraint.

When the sender provides information in both periods, she can always extract surplus from the receiver if the receiver gets a strictly positive surplus. If the sender faces information under-provision, she can increase the payoff by increasing the probability of a positive signal and decreasing the belief after observing a positive signal. If the sender faces information over-provision, she can increase the payoff by reducing the probability of a positive signal and increasing the belief after observing a positive signal. This implies that the receiver gets zero surplus in equilibrium.

The optimal strategy is consistent with real-world examples. Consider the application of product sales in section 5. A buyer may consider a pair of shoes or a car. Visiting the dealership is time-consuming, in contrast to the ease of searching for information about a pair of shoes on the internet, but consumers are often not sure whether they will like a product based solely on the information online. In the case of car-buying, test driving and then haggling with the dealer can take several hours, but people who purchase the car after visiting the dealership usually are sure that they like the car.

Comparative Statics

When the sender provides information in only one period, the optimal strategy has a closed-form solution and is easy to analyze. Here, we discuss the comparative statics when the sender provides information in both periods. The specific form of the optimal strategy depends on the relative size of the search cost. According to Proposition 2, the sender's strategy in the last period is constant when the search cost is high (but not too high). When the search cost is lower, the sender's strategy depends on her belief at the beginning of the second period, μ_1. When $\mu_1 > c - v_b\lambda_1^{**}$, the receiver expects to get a strictly positive surplus in the second period, and thus the first-period participation constraint is relaxed (denote the corresponding strategy as the S_+ strategy). When $\mu_1 \in [c/v_g, c - v_b\lambda_1^{**}]$, the receiver expects to get zero surplus in the second period (denote the corresponding strategy as the S_0 strategy). In equilibrium, the sender endogenously determines whether to use the S_0 or S_+ strategy.

Comparative Statics With Regard to the Prior Belief

When the search cost is high, the prior determines whether the sender smooths information but does not affect the information structure, conditional on the sender providing information. When the search cost is low, the prior affects the information structure monotonically. When the search cost is intermediate, the sender may switch from the S_0 strategy to the S_+ strategy as the prior increases. There can be a discrete jump in the optimal information structure. We leave the analysis of this case to the appendix.

Proposition 4. *Suppose the sender provides information in both periods. When the search cost is high, $v_g\lambda_1^{**} \le c < \widehat{c}$, positive signal fully reveals the state. Neither the probability of a positive signal, λ_t^*, nor the sender's payoff depends on the prior, $(\lambda_t^*, \bar{\mu}_t{}^*) = (c/v_g, 1)$. When* *the search cost is low, $c \leq \tilde{c} := v_g K'^{-1}\left[K(\lambda_1^{**})/\lambda_1^{**}\right]$, the probability of a positive signal, λ_t^*, is continuous and increases in the prior. The belief after observing a positive signal, $\bar{\mu}_t^*$, is continuous and decreases in the prior. The sender's payoff strictly increases in the prior.*

The optimal information is perfectly smooth when the search cost is high and the sender provides information in both periods. Since the participation constraint of the receiver is strong, a positive signal fully reveals the state is g. Because it is very costly for the receiver to acquire information, the minimal amount of information to persuade the receiver to search is high. The marginal cost of providing more information exceeds the marginal benefit. Even if the prior increases and it is feasible for the sender to provide more information, she will prefer not to do so. Hence, conditional on the sender providing information, the information structure does not depend on the prior.

When the search cost is low, and the sender provides information in both periods, she chooses between S_+ and S_0 strategies. Under the S_+ strategy, the receiver observes less frequent positive signals in the first period and more frequent positive signals in the second period. On average, he spends a longer time searching. Consequently, the sender has to provide information more favorable to the receiver to compensate for the higher expected total search cost, which reduces the sender's surplus. Therefore, the sender always chooses the S_0 strategy in equilibrium, and the optimal strategy is continuous in the prior. The sender faces information under-provision in both periods. More frequent positive signals are feasible when the prior is higher. Even if the receiver becomes less sure about the state being good after observing a positive signal, he will still search as long as the likelihood of receiving a positive signal and earning a strictly positive surplus increases. In equilibrium, the sender trades off the precision of a positive signal for frequency as the prior increases. The consumer spends less time searching for information because he is more likely to receive a positive signal and make a decision in the first period.

Comparative Statics With Regard to the Sender's Costs

Providing the same amount of information may impose different costs on the sender. To study the impact of the sender's information provision costs on the optimal strategy, we rewrite the sender's cost function as $K(\lambda) = \eta\tilde{K}(\lambda)$, where $\tilde{K}(1/2) = 1$ for identification. It is more costly for the sender to provide information when η is larger. The following proposition summarizes the comparative statics of the optimal strategy about η.

Proposition 5. *Suppose the sender provides information in both periods. Her payoff strictly decreases in the relative cost of information provision, η. When the search cost is low, $c \leq \tilde{c}$, the sender provides less information in the first period and more information in the second period, as η increases. When the search cost is high, $c \geq v_g\lambda_1^{**}$, the optimal strategy of the sender does not depend on η.*

When the search cost is high, it is very costly for the receiver to search. The sender needs to provide a lot of information to persuade the receiver to search. As a result, the marginal cost of providing more information exceeds the marginal benefit. So, the sender provides the minimum amount of information for the receiver to search, which does not depend on the sender's cost. Hence, the optimal information structure does not depend on the relative cost of information provision.

When the search cost is low, and the relative cost of information provision increases, the marginal cost of providing information increases, while the marginal benefit remains the same. Because the sender definitely incurs the information provision cost in the first period, she provides less information in the first period. This allows her to provide more information in the second period when the information provision cost is not always incurred, and when she faces information under-provision. The consumer spends more time searching for information because he is less likely to receive a positive signal and make a decision in the first period.

A Numerical Example

We illustrate the optimal strategy of the sender by a numerical example. The sender's cost function has the truncated quadratic form, $K(\lambda) = k\lambda^2/(1-\lambda)$, which satisfies Assumption 1. In each figure below, we present the optimal strategy and the sender's payoffs from the optimal one-period, S_0, and S_+ strategies[8]. The search cost is low in Figure 1.5.[9] The sender always prefers the S_0 strategy when she provides information in both periods. As illustrated, the probabilities of positive signal at both periods, λ_0^* and λ_1^*, are continuous and increase in μ_0. The beliefs after observing a positive signal in each period, $\bar{\mu}_0^*$ and $\bar{\mu}_1^*$, are continuous and decrease in μ_0. When the prior is lower than the intercept of the brown line, the sender prefers providing information in only one period to smooth the information provision. When the search cost increases to the intermediate level (Figure 1.6), the sender always provides information in both periods provided that it is feasible. So, we do not plot the sender's payoff of providing information in only one period. The sender prefers the S_0 strategy for a low prior and switches to the S_+ strategy as the prior increases. The optimal strategy is non-monotonic and discontinuous in μ_0 due to the switch. When the search cost further increases (Figure 1.7), the optimal information structure is perfectly smooth, and a positive signal always fully reveals the state.

[8]The domain of the prior is $[c(2v_g - c)/(v_g)^2, \widehat{\mu}_0 \wedge p]$. When $\mu_0 < c(2v_g - c)/(v_g)^2$, the sender provides information in at most one period. When $\mu_0 \geq \widehat{\mu}_0 := 2c - v_b\lambda_1^{**} - [c + (1-\lambda_1^{**})v_b]\lambda_0^{**}$, the sender always chooses the benchmark solution.

[9]The choice of the specific parameter values do not affect the qualitative property of the optimal strategy (i.e., the shape of the figure). What matters is the relative value. The search cost is low if $c \leq v_g\lambda_0^{**}$, intermediate if $v_g\lambda_0^{**} < c < v_g\lambda_1^{**}$, and high if $v_g\lambda_1^{**} \leq c < \widehat{c}$.

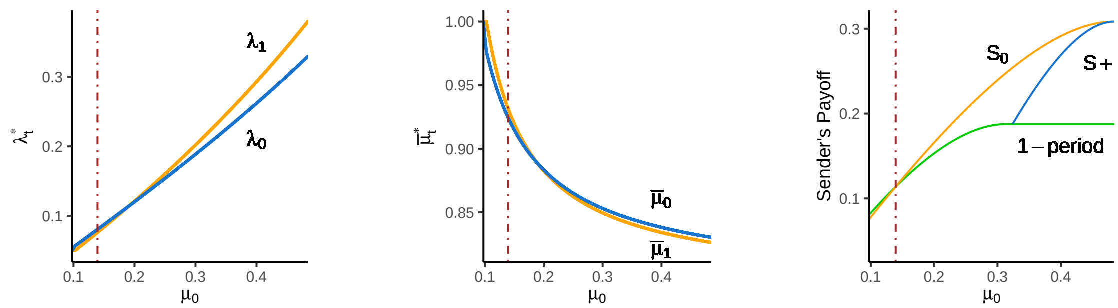


Figure 1.5: The optimal strategy when $c = 0.01, p = 0.8, v_g = 0.2, v_b = -0.8, k = 0.5$

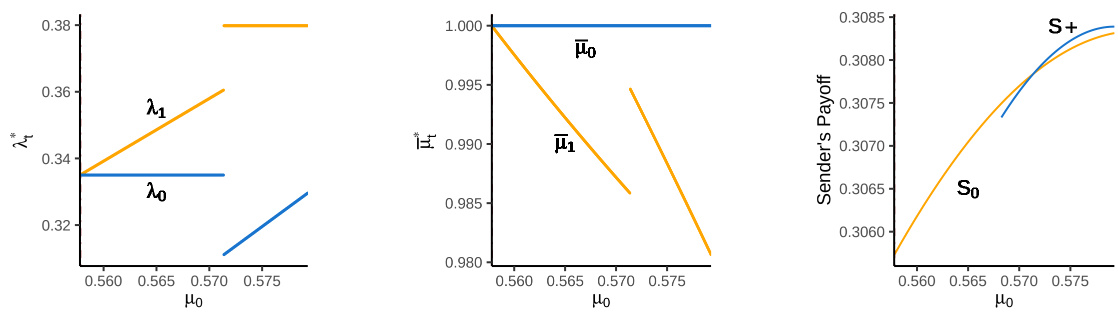


Figure 1.6: The optimal strategy when $c = 0.067, p = 0.8, v_g = 0.2, v_b = -0.8, k = 0.5$

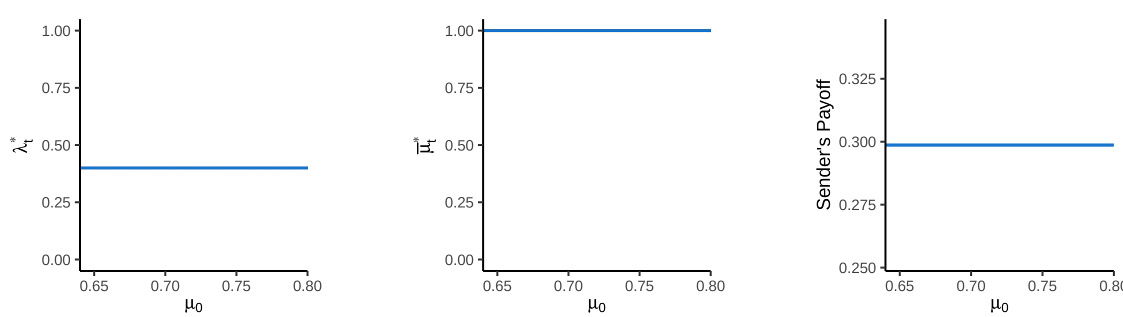


Figure 1.7: The optimal strategy when $c = 0.08, p = 0.8, v_g = 0.2, v_b = -0.8, k = 0.5$

1.4 The Efficient Information Structure

In the previous sections, the sender designs the information structure to maximize the expected payoff. This section characterizes the efficient strategy when a social planner designs the information structure to maximize total welfare. We then investigate the information distortion caused by not taking into account receiver surplus. For tractability reasons, we use a special form of the payoff function, specified in section 5. So, in this section, $v_g = 1 - p$ and $v_b = -p$.

Efficient Strategy in the Last Period

As discussed in the previous section, the sender does not provide information if the belief at the beginning of the second period is too low, $\mu_1 < \mu_{0,1}$. So, we concentrate on the case in which $\mu_1 \geq \mu_{0,1}$. In the second period, the social planner's problem is:

$$\max_{\lambda_1, \bar{\mu_1}} -K(\lambda_1) + p\lambda_1 + \lambda_1(\bar{\mu_1} - p) - c \quad (E_1)$$
$$\text{s.t. } (IR_1), (F_1)$$

Below, we discuss the efficient information structure in the last period intuitively. The formal characterization of the efficient strategy in the last period is in the appendix. When the search cost is high, it is very costly for the receiver to search. The sender needs to provide a lot of information to persuade the receiver to search. As a result, the marginal cost of providing more information exceeds the marginal benefit. So, the sender provides the minimum amount of information necessary to induce the receiver to search, which does not depend on whether the sender maximizes the sender's surplus or total welfare. Hence, there is no information distortion. When the search cost is lower, it is easier to persuade the receiver to search. The sender provides more than the minimum amount of information necessary to induce the receiver to search. The marginal costs of providing more information are the same for both the sender and the social planner, while the marginal benefit of providing more information is smaller for the sender. Therefore, the sender provides less information than the social planner does when the search cost is high or the initial belief is high. One exception is that, when both the search cost and the initial belief are low, the sender provides more information than the social planner does because the social planner can only generate infrequent signals.

Efficient Strategy for the Entire Game

As in the previous section, the social planner does not provide information if the prior is too low or the search cost is too high. When she provides information in only one period, the previous subsection characterizes the efficient strategy. When she provides information in both periods, her problem is:

$$\max_{\lambda_0,\bar{\mu_0},\mu_1,\lambda_1,\bar{\mu_1}} -K(\lambda_0) + \lambda_0\bar{\mu_0} - c + (1-\lambda_0)\left[-K(\lambda_1) + \lambda_1 - c\right] \qquad (E_2)$$
$$\text{s.t. } (IR_0), (F_0), (\lambda_1, \bar{\mu}_1) \text{ solves } (E_1)$$

The following proposition compares the payoff-maximizing strategy and the efficient strategy when the sender provides information in both periods.

Proposition 6. *Suppose the sender provides information in both periods. When $c \geq v_g\lambda_1^{**}$, the payoff-maximizing strategy is efficient. When $c < v_g\lambda_1^{**}$, the sender, who maximizes her own payoff, provides less information in the first period and more information in the second period than does the social planner, who maximizes total welfare.*

When the search cost is high, similar to the argument in the previous subsection, the sender provides the minimum amount of information for the receiver to search, which does not depend on the sender's objective (maximizing sender surplus or total surplus). Hence, there is no information distortion. When the search cost is lower, it is easier to persuade the receiver to search. The sender provides more than the minimum amount of information for the receiver to search. The sender's information structure is no longer efficient because she does not internalize the receiver's welfare. Since the social planner benefits directly from reducing the receiver's search cost, she designs the information to speed up the search process. The receiver takes action G with a higher probability in the first period under the efficient information structure.0

1.5 Model Extensions

Dynamic Commitment

Under many circumstances, the assumption that the sender can generate credible signals within each period but does not have dynamic commitment power is reasonable. It is hard for the sender to commit to the entire information structure across all periods ex-ante and to convince the receiver that the sender will stick to the information structure when it is profitable to deviate to a different information structure during intermediate periods. However, factors such as reputation can give the sender stronger commitment power. Here, we study the implications of the case in which the sender has dynamic commitment power. The sender chooses and commits to the entire information structure to maximize the ex-ante expected surplus.

$$\max_{\lambda_0,\bar{\mu_0},\mu_1,\lambda_1,\bar{\mu_1}} -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\lambda_1) + p\lambda_1\right] \qquad (P_{dc})$$
$$\text{s.t. } (IR_0), (F_0), (IR_1), (F_1)$$

Proposition 7. *Suppose the search cost is high, $v_g\lambda_1^{**} < c < \widehat{c}$, and $\mu_0 > c(2v_g - c)/(v_g)^2$. The sender provides information in both periods regardless of the dynamic commitment power. If the sender has dynamic commitment power, her payoff is strictly higher, and the receiver gets a strictly positive surplus in the second period. The benefit of dynamic commitment power for the sender vanishes as the search cost approaches zero.*

We have shown that, when the search cost is high, the sender perfectly smooths the information if she doesn't have dynamic commitment power. If the sender instead has dynamic commitment power, she will commit to providing information more favorable to the receiver in the second period. As a result, the receiver will search even if the sender provides less information in the first period, relaxing the information over-provision issue. Though it hurts the sender's payoff in the second period, it increases the sender's payoff in the first period by reducing the information provision cost. The overall effect is strictly positive. So, the optimal information provision will not be perfectly smooth.

The above finding is related to results on durable good pricing (e.g., Coase 1972). Without dynamic commitment power, the monopolist tends to reduce the price as time goes on, which reduces profit, because a rational receiver will strategically wait. Here, dynamic commitment power also benefits the sender, but the underlying mechanisms differ. This paper focuses on persuasion (information provision) rather than incentive (pricing). In addition, in the durable good pricing example, the ability to commit not to provide a more favorable price in the future benefits the sender. By contrast, in this paper, the ability to commit to more favorable information in the future benefits the sender.

However, when the search cost approaches zero, the difference between the sender surplus with and without dynamic commitment power approaches zero. The intuition is that the benefit of dynamic commitment power comes from relaxing the participation constraint in the first period by committing to providing more favorable information in the second period. However, the participation constraint is already very loose when the search cost is low. Thus, the benefit of dynamic commitment power approaches zero.

Discounting

In the main model, we assume that there is no discounting. In reality, information acquisition and provision usually happen in a short period, and that assumption is reasonable. However, some communications between the sender and the receiver can take longer. Here, we study the information provision strategy when the sender and the receiver have the same discount factor, $\delta \in (0, 1)$. With discounting, the sender's problem becomes:

$$\max_{\lambda_0,\bar{\mu_0},\mu_1,\lambda_1,\bar{\mu_1}} -K(\lambda_0) + p\lambda_0 + \delta(1-\lambda_0)\left[-K(\lambda_1) + p\lambda_1\right] \qquad (P_{2,\delta})$$

$$\text{s.t. } \lambda_0(\bar{\mu_0} + v_b) + \delta(1-\lambda_0)[\lambda_1(\bar{\mu_1} + v_b) - c] \geq c \qquad (IR_{0,\delta})$$

$$(F_0), (\lambda_1, \bar{\mu_1}) \text{ solves } (P_1)$$

One can see that both the objective function and the first-period participation constraint change. When the players become less patient (the discount factor δ decreases), the present value of the second-period sender surplus decreases, and the first-period participation constraint becomes tighter. Thus, it is less attractive for the sender to sell the goods in the second period.

Proposition 8. *When the search cost is high, $c \geq v_g \lambda_1^{**}$, the optimal strategy does not depend on the discount factor, δ. When the search cost is low, $c \leq \tilde{c}$, the sender provides more information in the first period and less information in the second period, as players become less patient.*

When the search cost is high, it is hard to satisfy the participation constraints of the receiver. Providing enough information to persuade the receiver to search dominates the force of discounting. Therefore, the sender's strategy remains the same as the no-discounting case, and the sender perfectly smooths information provision. When the search cost is low, the sender provides less information in the second period when the players are less patient because of discounting. The sender provides more information in the first period as she becomes more tempted to convert the receiver early.

We assume in the main model that the receiver is forward-looking and takes into account the potential payoff of the second period when he chooses his action in the first period. This assumption is reasonable if the information environment is transparent and the receiver knows that gradual learning is possible. We now consider the possibility of asymmetric discount factors. In particular, the sender is perfectly patient while the receiver is perfectly impatient (myopic). If the receiver is myopic, then he trades off only the current-period benefit and cost in deciding whether or not to search. The information in the second period cannot relax the first-period participation constraint. Therefore, when the receiver is myopic, the feasible information structure is a subset of when the receiver is forward-looking. If the receiver's surplus in the second period is strictly positive when the receiver is forward-looking, the optimal information structure may not be feasible when the receiver is myopic. Hence, the sender is (weakly) worse off if the receiver is myopic rather than forward-looking. This result has managerial implications, as it suggests that the sender should try to better inform the receiver of the possibility of gradual learning. Common knowledge of gradual information revelation improves the sender's surplus.

Infinite Number of Periods

The two-period model can capture information smoothing and gradual learning. We extend the main model to a model with an infinite number of periods and show that the main insights extend to this richer model. This also provides some additional insights.

Time is discrete, $t = 0, 1, 2, ...$ In each period, the sender determines and commits to the information structure of the current period but cannot commit to the information structure in the future. The receiver can search for information before deciding. Unlike the two-period model, there is no deadline. The receiver can search for as long as he wants. Since the sender's payoff is bounded by p, the payoff function is well-defined even without the discount factor. So, we do not consider discounting for consistency with the main model. We can analyze the problem similarly if we include discounting.

Proposition 9. *When the search cost is high, $v_g \lambda_1^{**} \leq c < \widehat{c}$, the sender provides perfectly smooth information for $k := \lfloor \frac{ln(1-\mu_0)}{ln(1-c/v_g)} \rfloor$ periods, and a positive signal fully reveals the state, $(\lambda_t^*, \bar{\mu_t}^*) = (c/v_g, 1)$, for $t = 0, 1, ..., k-1$. When the search cost is low, the sender adds noise to positive signals. As the search cost approaches zero, the sender could obtain the equilibrium payoff as if the persuasion cost were zero.*

This proposition shows that the main insights are robust to the specification of the length of time. The two-period model corresponds to the case when there is a deadline in the information acquisition. When time is infinite, there is no limit on how long the receiver can search. Under high search costs, the optimal information structure fully convinces the receiver that the state is g when a positive signal arrives. Under low search costs, the optimal information structure adds some noise to the positive signal. The intuition is the same as the two-period case. Because the receiver can keep searching for a longer period, the sender can better smooth the information. When the search cost is high, the sender may provide information for more than two periods if she believes that the state is likely to be g and the receiver is willing to spend more time searching. The proposition shows that the sender smooths information provision over more periods for a higher prior belief. The ability to smooth the information is valuable for the sender, especially when the search cost approaches zero. In that case, she could obtain the equilibrium payoff as if the persuasion cost were zero. Because of the low search cost, the sender can convince the receiver to search with very little information in each period. The sender's cost becomes very low by smoothing the information over many periods.

1.6 Concluding Remarks

Consumers frequently search for information before making decisions. Since their search and purchase decisions depend on the information environment, firms have a strong incentive to influence it. This paper endogenizes consumers' information environment from the firm's perspective under a general signal space.

We examine the optimal information provision strategy of a sender and the optimal information acquisition strategy of a receiver when the sender sequentially persuades a receiver to take a particular action (e.g., to purchase a good). The sender prefers that action regardless of the unknown state, while the receiver only wishes to take that action if the state is good. In our model, the sender incurs a cost to provide information, and the receiver incurs a cost to search. The receiver trades off the cost of searching and the benefit of obtaining more accurate information to make better decisions. The sender trades off the cost of information provision and the benefit of persuading the receiver to search and then take the sender's preferred action. We allow for gradual communication between the sender and the receiver. Consequently, the sender also makes the intertemporal trade-off of smoothing the information to reduce the persuasion cost. In equilibrium, she uses one-shot signals that induce the receiver to immediately take the sender's preferred action upon observing a positive signal. The sender smooths information over multiple periods if and only if there is a high prior that the state is good. The sender extracts all the surplus from the receiver when she provides information in both periods, while she may leave some surplus for the receiver when she provides information in only one period. When the search cost for the receiver is high, the receiver is sure that the state is good when he takes the desired action. When the search cost is low, the optimal information structure does not fully reveal the state, which may be bad even though the receiver takes the desired action. We compare the payoff-maximizing information structure with the efficient information structure and find no information distortion when the search cost is high. There can be upward or downward information distortion when the search cost is lower.

There are some limitations to the current work. The implementation of a given signal depends on the institutional details of the specific problem. Further empirical work can complement the current paper by putting the theoretical results into practice. In addition, it will be interesting to study the optimal persuasion strategy when there is more than one sender. Such competition may lead the sender to provide more information and improve equilibrium efficiency. Moreover, the sender has complete control of the information structure in this paper. It will be interesting to consider the case where the sender can only partially control the information environment.

Chapter 2

Multi-attribute Search and Informative Advertising

2.1 Introduction

When considering whether or not to buy a product, the consumer can often evaluate different attributes of it. An incoming college student finding a laptop can learn about the operating system, weight, exterior design, warranty, and many other attributes before making the final decision. Learning costs both time and effort, while consumers often have limited attention. So, she needs to decide which attribute to learn (first). Sometimes, she makes the decision based on exogenous reasons, such as an attribute is more prominent (Bordolo et al. 2013, Zhu and Dukes 2017), or her options in an attribute generate a greater range of consumption utility (Kőszegi and Szeidl 2013).

Consider a consumer deciding whether to buy a used car. She can gather information about many different attributes. Figure 2.1 summarizes some factors of the used car value. For example, the consumer can check details about the car's add-on packages by some review articles. She can also purchase a car report to find out the car's accident history. Both options help the consumer learn more about the car and improve the decision. However, it takes time and effort to search for such information. The consumer needs to decide to which attribute to pay attention. If one attribute is much more important than the other, she will search for information about the most important attribute. However, the consumer may not know whether the add-on features or the condition of the car matter more to her. What attribute should she search for if she is unclear about which one is more important ex-ante?

Even if the consumer decides which attribute to search for, she will not learn everything about it immediately. Instead, she gradually gathers information about the attribute. For instance, Even if the consumer spends half an hour searching for information about the car's safety features and finds out that the car has airbags in each of the seats, she still does not know everything about the car's safety. She can continue searching for information about whether the car has an automatic braking system. But, the consumer may not want to stick to one attribute. The relative importance of attributes may change as she learns more. After obtaining enough positive information about the car's safety, she may find it a better use of her time to switch to other attributes. She may feel confident that the car is safe but uncertain whether she will enjoy driving in it. At some point, the consumer may switch to learning more about the car's design. When will the consumer switch to search for another attribute because the relative importance of attributes changes as she gathers more information?

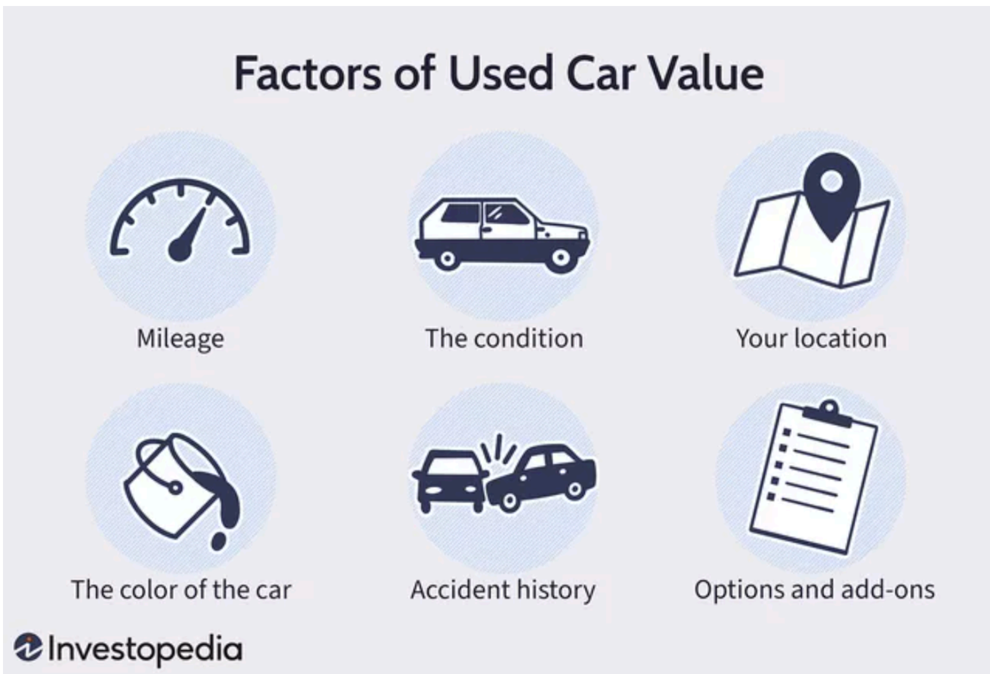


Figure 2.1: Attributes of the Used Car[1]

This paper considers a consumer deciding whether to purchase a good or not. The good has two attributes, whose values are independent. The payoff of purchasing the good is the total value of the attributes net of the price. The consumer does not know the value of either attribute. She has a prior belief about the value of each attribute, and can incur a cost to search for information about the attributes before making a decision. By receiving a noisy signal about an attribute from searching, she can update her belief about the value of that attribute and thus about the value of the product. By assuming that the search cost and the informativeness of the signal are the same for each attribute, we ensure that the attributes are symmetric. So, the consumer will not prefer searching for information about one attribute to the other for exogenous reasons. Which attribute to search at any given time is determined endogenously by the expected gain from an extra piece of information about each attribute.

[1]Source of the figure: https://www.investopedia.com/articles/investing/090314/just-what-factors-value-your-used-car.asp

The consumer will stop searching and buy the good if she becomes optimistic enough about its value (the total value of both attributes), and will stop searching without purchasing if she becomes pessimistic enough about its value. When the consumer's belief about the value of the good is in between, she will search for more information. We characterize the search region by a set of ordinary differential equations for intermediate beliefs and by a system of equations for extreme beliefs. We find that it is always optimal for the consumer to search the attribute about which the consumer has the higher uncertainty due to the faster speed of learning. The consumer only searches for the more uncertain attribute if she holds a strong prior belief about one of the attributes and may search for both attributes otherwise. In the car purchasing example, a consumer may not bother to search for the safety features of a Volvo car because Volvo has a good reputation for safety. So, she may instead focus on other aspects of the car. In contrast, Faraday Future has not produced any cars yet. If a consumer considers pre-ordering a car, she probably has a lot of uncertainty about everything. So, she may search for information about every attribute.

We study the comparative statics of the optimal search strategy. An increase in the price shifts the entire search region upwards because the consumer needs to gain a higher value from the good to compensate for the higher price. An increase in either the search cost or the noise of the signal makes searching less attractive for the consumer and shrinks the search region.

We also investigate how the consumer's purchasing likelihood depends on the prior belief. When the consumer is optimistic enough about both attributes, she will purchase the product for sure. When she is pessimistic enough about both attributes, she will never purchase the product. When her belief is in between, she will purchase the product with some probability.

In reality, firms can intervene the consumer search and purchase processes by changing consumers' prior beliefs through marketing activities such as advertising. We study the firm's optimal pre-search intervention by assuming that it can disclose the value of one attribute by informative advertising. We find that the firm will not advertise if the consumer's prior beliefs about both attributes are extreme. If the consumer is very optimistic about both attributes, she will purchase the product for sure or with a very high likelihood. So, the firm does not have an incentive to advertise. If the consumer is very pessimistic about both attributes, she will never purchase the product even if she knows that one attribute is good. So, the firm does not advertise either. If the consumer's prior belief is milder, the firm can increase the purchasing probability by advertising. The firm will advertise the better attribute if the consumer is optimistic enough about the worse attribute, and will advertise the worse attribute if the consumer is less optimistic about it.

This paper makes two main contributions. First, we endogenize the search order of different attributes of a product based on the consumer's optimal Bayesian learning. Second, we connect informative advertising with consumer search by comparing the one-dimensional search with advertising and two-dimensional search without advertising.

Related Literature

This paper is related to the literature on how consumers with limited attention allocate their attention to different attributes or options. Existing literature mainly looks at the case in which the attributes or options are asymmetric (Arbatskaya 2007, Armstrong et al. 2009, Xu et al. 2010, Armstrong and Zhou 2011, Bordolo et al. 2013, Kőszegi and Szeidl 2013, Branco et al. 2016, Zhu and Dukes 2017, Jeziorski and Moorthy 2018). In those papers, consumers know that they face attributes with different prominence/importance ex-ante. For example, the search order is exogenous in Arbatskaya (2007). Armstrong et al. (2009) extend the symmetric search model of Wolinsky (1986) by assuming that there is a prominent firm for which all the consumers will search first. In their model, the prominent firm is exogenous. They do not model why consumers want to search for that firm first. In Bordolo et al. (2013), the salient attribute of a good is the attribute furthest away from the average value of the same attribute in the choice set. In Zhu and Dukes (2017), each competing firm can promote one or both attributes of a product. Though the prominence of the product is endogenously determined by competition, it is exogenously given from the consumer's perspective. Jeziorski and Moorthy (2018) examine the effect of prominence in search advertising. There are two types of prominence in their setting, the position of the ads and the prominence of the advertiser. They find that the ad position prominence and the advertiser prominence are substitutes in consumers' clicking behavior. One of the main contributions of our paper is to endogenize the optimal attribute to search from the consumer's perspective. Instead of assuming that the consumer knows the value of each attribute or learns it at once, as is common in this literature, the Bayesian decision-maker in our model gradually learns the value from noisy signals. So, the relative importance of the attributes may change as the consumer gathers more information. In contrast, the prominence attribute/option in the existing literature does not change over time because they impose exogenous differences on the attributes.

This paper also fits into the literature on optimal information acquisition, particularly consumer search. Stigler 1961 and Weitzman 1979 are among the first papers to derive the optimal search rules under simultaneous and sequential search, respectively. In both papers, the relative importance of different alternatives is exogenous. Consumers observe the distribution of the rewards before making the search decision. Later papers incorporate gradual learning (Moscarini and Smith 2001, Branco et al. 2012, Ke et al. 2016). Like our paper, the attributes are symmetric in those papers. However, the consumer randomly searches for an attribute in those papers. In our model, the consumer decides when to search and which attribute to search. Ke and Villas-Boas (2019) are closely related to our paper. They study the gradual learning of information about multiple alternatives. The decision-maker endogenously determines which alternative to search. There are two main differences between their paper and this one. First, the expected payoff of choosing one of the alternatives depends only on the information gathered from that alternative. So, the objective of searching is to differentiate different alternatives. In our paper, the expected payoff of adopting the product jointly depends on the information gathered from all the alternatives. So, the objective of searching is to learn about the overall distribution of all the attributes. Second, they focus on the decision maker's optimal search strategy. In contrast, we also study the firm's response. We show how the firm can change the consumer's search behavior and increase its profits by informative advertising, given the optimal search strategy of the consumer.

Lastly, this paper is related to the literature on informative advertising. People have begun to consider the informational role of advertising since Nelson (1974). Subsequent papers study the disclosure of price (Anderson and Renault 2006) and quality (Lewis and Sappington 1994, Anderson and Renault 2009) by informative advertising. Sun (2011) is the closest paper that studies a seller's disclosure incentive for a product with multiple attributes. It shows that the unraveling result by Grossman (1981) and Milgrom (1981) will not hold if the product has a vertical attribute and a horizontal one. If the product has a high vertical quality, the seller may not disclose the product's horizontal attribute.

Consumers' only source of information about the product comes from the firm in most of the existing advertising literature. In reality, consumers can search for more information after they see the ads. We take it into account by building a micro-founded consumer search model. After the firm advertises, the consumer can still search for information about any attributes. The firm anticipates it when choosing the advertising strategy. Mayzlin and Shin (2011) consider a setting where the consumer can obtain an exogenously given signal by searching for information about the product quality after the firm advertises. Our paper differs from their paper in two ways. On one hand, the quality is vertical in their paper, and the advertising strategy is driven by signaling the firm's private information. We focus on horizontal quality, and the advertising strategy is driven by the difference in one-dimensional search with advertising and two-dimensional search without advertising. On the other hand, the consumer can only search once and observe an aggregate signal about the firm's quality in Mayzlin and Shin (2011). We model the search process in detail so that the consumer chooses what attribute and how long to search. This allows us to endogenize the search order and understand more about the consumer's search behavior and the firm's best response to it.

The remainder of the paper is organized as follows. Section 2 presents the main model. Section 3 studies the comparative statics of the search region. Section 4 characterizes the purchasing likelihood given a prior belief and the consumer's optimal search strategy. Section 5 discusses firms' advertising strategy. Section 6 concludes.

2.2 Model

A consumer considers whether to purchase a product or not. The product has two attributes whose values are independent. The product's value for the consumer is the sum of the values of the attributes, $U = U_1 + U_2$. The value of each attribute is one if it is good and zero if it is bad. The consumer's prior belief that attribute i is good is $\mu_i(0)$. We assume that the firm does not have private information about the value of the attribute.[2] The price p is exogenously given. We assume that the marginal cost of producing the product is high enough, and thus the price is high enough ($p \geq 3/2$) such that the consumer will quit without purchasing the product for any $\vec{\mu} = (\mu_1, \mu_2)$, if $\mu_1 + \mu_2 \leq 1$. Hence, we restrict our attention to the case in which $\mu_1 + \mu_2 > 1$. The consumer can learn more about the attributes via costly learning before making a decision. At time t, the consumer can make a purchasing decision or search for information. Because of limited attention, she can only search for information about one attribute at a time. So, if the consumer chooses to search for information, she also needs to decide which attribute to search for information about. The game ends when the consumer makes a decision. If the consumer decides to search for information, she will obtain noisy signals about an attribute by incurring a flow cost of c. Define $T_i(t)$ as the cumulative time that attribute i has been searched until time t. We model the signal, S_i, by a Brownian motion (W_i are independent Wiener processes):

$$dS_i(t) = U_i dT_i(t) + \sigma dW_i(T_i(t))$$

The consumer will be more likely to observe a larger signal realization if the attribute is good. Given the received signal, the consumer continuously updates her belief on the value of each attribute according to Bayes' rule.[3] The belief evolution can be characterized by the following ODE:

$$d\mu_i(t) = \frac{1}{\sigma^2}\mu_i(t)[1 - \mu_i(t)]\{dS_i(t) - \mathbb{E}[U_i|\mathcal{F}_t]dT_i(t)\} \tag{2.1}$$

, where $\{\mathcal{F}_t\}_{t=0}^{+\infty}$ is a filtration with all the observed information up to time t.

The consumer's expected payoff for a given belief $\vec{\mu}$, learning rule α, and stopping time τ is:

$$J(\vec{\mu}, \alpha, \tau) = \mathbb{E}\left\{\max\left[\mu_1(\tau) + \mu_2(\tau) - p, 0\right] - \tau c|\vec{\mu}(0) = \vec{\mu}\right\}$$

The value function of the consumer's problem is:

$$V(\vec{\mu}) := \sup_{\alpha,\tau} J(\vec{\mu}, \alpha, \tau)$$

Since the learning rule and stopping time should not depend on any future information, the decision at time t should only be based on the observed information up to time $t, \mathcal{F}_t$. It is well known that the current belief $\vec{\mu} = (\mu_1, \mu_2)$ is a sufficient statistic for $\mathcal{F}_t$. So, the learning rule and stopping time will depend only on $\vec{\mu}$. If a learning rule α^* and a stopping time τ^* achieve that value for any given belief, they will be the optimal learning rule and the optimal stopping time.

---

[2]This will be more realistic if we consider the horizontal preference rather than the vertical preference.

[3]Notice that the consumer's belief about an attribute will remain the same when she searches for information about the other attribute.

$$V(\vec{\mu}) = J(\vec{\mu}, \alpha^*, \tau^*)$$

The next section characterizes the consumer's value function and optimal search strategy, including the optimal learning rule and the optimal stopping time.

Optimal Strategy

When the consumer searches for information about attribute one, the value function satisfies (ignoring the time index t for simplicity):

$$V(\mu_1, \mu_2) = -cdt + \mathbb{E}[V(\mu_1 + d\mu_1, \mu_2)]$$

By Taylor's expansion and Ito's lemma, we get:

$$\frac{\mu_1^2(1-\mu_1)^2}{2\sigma^2} V_{\mu_1\mu_1}(\mu_1, \mu_2) - c = 0 \tag{2.2}$$

Similarly, when the consumer searches for information about attribute two, we have:

$$\frac{\mu_2^2(1-\mu_2)^2}{2\sigma^2} V_{\mu_2\mu_2}(\mu_1, \mu_2) - c = 0 \tag{2.3}$$

The HJB equation of the entire problem is:

$$\max\left\{\max_{i=1,2}\left[\frac{\mu_i^2(1-\mu_i)^2}{2\sigma^2} V_{\mu_i\mu_i}(\mu_1, \mu_2) - c\right], \max\left[\mu_1 + \mu_2 - p, 0\right] - V(\mu_1, \mu_2)\right\} = 0 \qquad (\star)$$

A standard method of solving this kind of stochastic control problem is the"guess and verify" approach. We first conjecture an optimal search rule and use it to characterize the search region and the value function. We then verify that the conjectured search rule is indeed optimal. Because of symmetry, we only need to consider the case in which $\mu_1 \geq \mu_2$. Analytically, we can fully characterize the optimal search strategy when the search cost is low. We do not think the result for the low search cost case is a strong restriction, as we are interested in the consumer's search behavior and how the firm can influence it by informative advertising. Naturally, the more interesting case is when the consumer searches more given the low search cost. When the search cost or the price is very high, the consumer searches little and the problem is less interesting and relevant.

Intuitively, the consumer will stop searching, not buy the product if the belief becomes too low, and will purchase the product if the belief becomes high enough. When the belief is in between, she keeps searching for information. We also conjecture that it is optimal for the consumer to search attribute two, conditional on searching, if $\mu_1 + \mu_2 > 1$ and $\mu_1 \geq \mu_2$.[4]

[4]By symmetry, if $\mu_1 + \mu_2 > 1$ and $\mu_1 < \mu_2$, it is optimal for the consumer to search attribute one, conditional on searching.

The intuition for this learning rule to be optimal is that the consumer prefers to search the attribute with a higher rate of learning, as the learning costs are identical. From equation (2.1), one can see that the more uncertain the belief is, the faster the consumer learns about an attribute. Therefore, she always learns the attribute with a belief closer to 1/2.

Figure 2.2 illustrates the optimal search strategy. The dashed orange line is the quitting boundary, and the solid blue line is the purchasing boundary. The grey arrow represents which attribute the consumer searches for information about, given the current belief. When the overall beliefs of the attributes are low enough, the likelihood of obtaining lots of positive signals and purchasing the good is too low. The consumer stops searching and quits to save the search cost. When the overall beliefs of the attributes are high enough, purchasing the good gives the consumer a higher enough expected surplus. So, she makes the purchase. In other cases, the consumer searches for more information to make a better decision. Denote the intersection of the quitting boundary and the main diagonal by (μ^*, μ^*), the intersection of the purchasing boundary and the main diagonal by (μ^{**}, μ^{**}). Represent the quitting boundary when $\mu_1 \geq \mu_2$ by $\underline{\mu}(\cdot)$, whose domain is $[\mu^*, 1]$ (the other half of the quitting boundary is determined by symmetry). Represent the purchasing boundary when $\mu_1 \geq \mu_2$ by $\bar{\mu}(\cdot)$, whose domain is $[\mu^{**}, 1]$ (the other half of the purchasing boundary is determined by symmetry).

The PDE when the consumer searches attribute two, equation (2.3), has the following general solution:

$$V(\mu_1, \mu_2) = 2\sigma^2 c(1 - 2\mu_2) \ln \frac{1 - \mu_2}{\mu_2} + B_1(\mu_1)\mu_2 + B_2(\mu_1), \mu_1 \in [\mu^*, 1]$$

We also have $V(\mu_1, \mu_2) = 0$ at the quitting boundary $\mu_2 = \underline{\mu}(\mu_1)$. For the value function in the search region, value matching and smooth pasting (wrt μ_2) at the quitting boundary $(\mu_1, \underline{\mu}(\mu_1))$ imply:[5]

$$\frac{V(\mu_1, \mu_2)}{2\sigma^2 c} = (1 - 2\mu_2) \ln \frac{1 - \mu_2}{\mu_2} + \phi(\underline{\mu}(\mu_1))\mu_2 - \psi(\underline{\mu}(\mu_1)) \tag{2.4}$$

, where $\phi(x) = 2 \ln \frac{1-x}{x} + \frac{1}{x} - \frac{1}{1-x}$ and $\psi(x) = \ln \frac{1-x}{x} + \frac{1-2x}{1-x}$.

By symmetry, for $\mu_1 < \mu_2$, the value function in the search region satisfies:

$$\frac{V(\mu_1, \mu_2)}{2\sigma^2 c} = (1 - 2\mu_1) \ln \frac{1 - \mu_1}{\mu_1} + \phi(\underline{\mu}(\mu_2))\mu_1 - \psi(\underline{\mu}(\mu_2)) \tag{2.5}$$

Equation (2.4) characterzies the value function for beliefs $\mu_1 \geq \mu_2$. Equation (2.5) characterzies the value function for beliefs $\mu_1 < \mu_2$. The two regions are separated by the main diagonal $\{(\mu_1, \mu_2) : \mu_1 = \mu_2\}$. Continuity of $V_{\mu_1}(\mu_1, \mu_2)$ at this boundary implies that:

$$\underline{\mu}'(\mu) = \frac{\phi(\underline{\mu}(\mu)) - \phi(\mu)}{\phi'(\underline{\mu}(\mu))[\mu - \underline{\mu}(\mu)]}, \text{ for } \mu \in (\mu^*, \mu^{**}] \tag{D_1}$$

[5]For technical details, please refer to Dixit (1993).

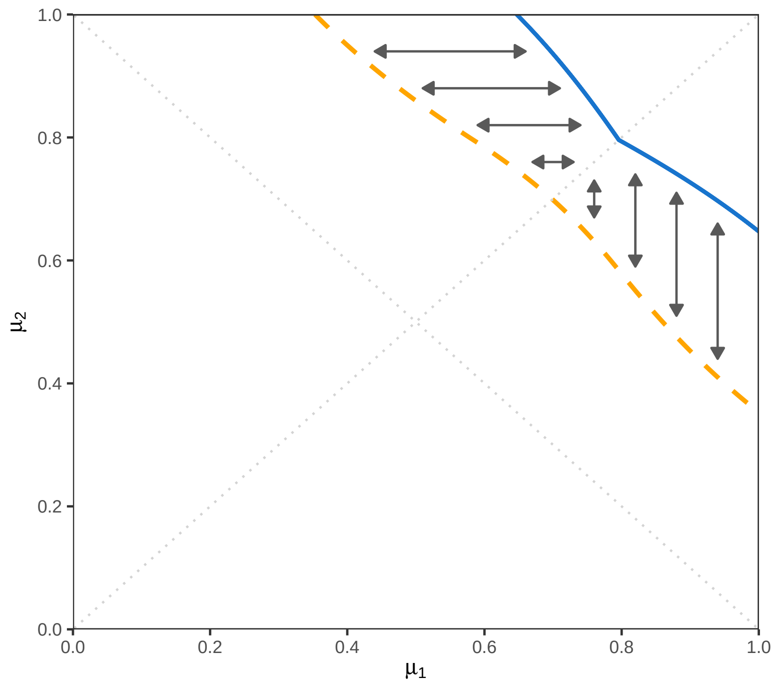


Figure 2.2: Optimal Search Strategy

For $\mu_1 \in [\mu^{**}, 1]$, $V(\mu_1, \mu_2) = \mu_1 + \mu_2 - p$ at the purchasing boundary $\mu_2 = \bar{\mu}(\mu_1)$. Value matching and smooth pasting (wrt μ_2) at the purchasing boundary $(\mu_1, \underline{\mu}(\mu_1))$ imply (in the search region):

$$\frac{V(\mu_1, \mu_2)}{2\sigma^2 c} = (1 - 2\mu_2) \ln \frac{1 - \mu_2}{\mu_2} + \phi(\bar{\mu}(\mu_1))\mu_2 - \psi(\bar{\mu}(\mu_1)) + \frac{\mu_1 - \mu_2 - p}{2\sigma^2 c} \tag{2.6}$$

Equation (2.4) and (2.6) use the quitting boundary and the purchasing boundary to pin down the value function, respectively. The resulting expression should be equivalent in the common domain $\mu_1 \in [\mu^{**}, 1]$. By equalizing V and V_{μ_2} of equation (2.4) and (2.6), we obtain the following system of equations:

$$\begin{cases} \phi(\underline{\mu}(\mu)) - \phi(\bar{\mu}(\mu)) = \frac{1}{2\sigma^2 c} \\ \psi(\underline{\mu}(\mu)) - \psi(\bar{\mu}(\mu)) = \frac{p-\mu}{2\sigma^2 c} \end{cases}, \text{ for } \mu \in [\mu^{**}, 1] \tag{2.7}$$

For each belief, μ, the system of equations above consists of two unknowns ($\bar{\mu}(\mu)$ and $\underline{\mu}(\mu)$) and two equations. They uniquely determine the function for the purchasing boundary

$\bar{\mu}(\mu)$ and the function for the quitting boundary $\underline{\mu}(\mu)$, for $\mu \in [\mu^{**}, 1]$, given a cutoff belief μ^{**}.

Instead of determining $\bar{\mu}(\mu)$ and $\underline{\mu}(\mu)$ by a system of equations (2.7), we can also implicitly determine $\bar{\mu}(\mu)$ and $\underline{\mu}(\mu)$ in two separate equations. Representing $\bar{\mu}(\mu)$ by $\underline{\mu}(\mu)$ from the first equation of (2.7), we have:

$$\bar{\mu}(\mu) = \phi^{-1}\left[\phi(\underline{\mu}(\mu)) - \frac{1}{2\sigma^2 c}\right]$$

Plugging it into the second equation of (2.7), we have:

$$\underline{\mu}(\mu) = \psi^{-1}\left\{\psi\left(\phi^{-1}\left[\phi(\underline{\mu}(\mu)) - \frac{1}{2\sigma^2 c}\right]\right) + \frac{p-\mu}{2\sigma^2 c}\right\}$$

The equation above implicitly determines $\underline{\mu}(\mu)$, for $\mu \in [\mu^{**}, 1]$. Similarly, we can implicitly determine $\bar{\mu}(\mu)$ by the following equation:

$$\bar{\mu}(\mu) = \psi^{-1}\left\{\psi\left(\phi^{-1}\left[\phi(\bar{\mu}(\mu)) + \frac{1}{2\sigma^2 c}\right]\right) - \frac{p-\mu}{2\sigma^2 c}\right\}$$

We now solve for the cutoff belief at the intersection of the purchasing boundary and the main diagonal, μ^{**}. Since (μ^{**}, μ^{**}) is on the purchasing boundary, we have $\mu^{**} = \bar{\mu}(\mu^{**})$, μ^{**} is determined by:

$$\begin{cases} \phi(\underline{\mu}(\mu^{**})) - \phi(\mu^{**}) = \frac{1}{2\sigma^2 c} \\ \psi(\underline{\mu}(\mu^{**})) - \psi(\mu^{**}) = \frac{p-\mu^{**}}{2\sigma^2 c} \end{cases} \tag{2.8}$$

The system of equations above consists of two unknowns (μ^{**} and $\underline{\mu}(\mu^{**})$ and two equations. They uniquely determine the cutoff belief μ^{**} via the following equations:

$$\phi^{-1}\left[\phi(\mu^{**}) + \frac{1}{2\sigma^2 c}\right] = \psi^{-1}\left[\psi(\mu^{**}) + \frac{p-\mu^{**}}{2\sigma^2 c}\right] \tag{I^{**}}$$

We have pinned down the cutoff belief μ^{**}. Given this cutoff beliefs, we have determined the purchasing boundary $(\mu, \bar{\mu}(\mu))$ and the quitting boundary $(\mu, \underline{\mu}(\mu))$, for $\mu \in [\mu^{**}, 1]$.

The ODE (D_1) and the initial condition (I^{**}) implicitly determine the function for the quitting boundary $\underline{\mu}(\mu)$, for $\mu \in (\mu^*, \mu^{**}]$, given a cutoff belief μ^*.

We now solve for the cutoff belief at the intersection of the quitting boundary and the main diagonal, μ^*. Since (μ^*, μ^*) is on the quitting boundary, we have $\mu^* = \underline{\mu}(\mu^*)$. This initial condition determines μ^*.

In sum, we have pinned down the cutoff belief μ^* and the quitting boundary $(\mu, \underline{\mu}(\mu))$, for $\mu \in [\mu^*, \mu^{**}]$.

We have fully characterized the purchasing boundary $(\mu, \bar{\mu}(\mu))$ and the quitting boundary $(\mu, \underline{\mu}(\mu))$, for $\mu_1 \geq \mu_2$. The other case in which $\mu_1 < \mu_2$ is readily determined by symmetry. The following proposition characterizes the slope of the purchasing and the quitting boundary and the shape of the search region.

Proposition 10. *For $\mu \in (\mu^*, \mu^{**}]$, we have:*

$$\underline{\mu}'(\mu) = \frac{\phi(\underline{\mu}(\mu)) - \phi(\mu)}{\phi'(\underline{\mu}(\mu))[\mu - \underline{\mu}(\mu)]} \tag{D_1}$$

For $\mu \in [\mu^{**}, 1]$, we have:

$$\bar{\mu}'(\mu) = \frac{\phi(\underline{\mu}(\mu)) - \phi(\bar{\mu}(\mu))}{\phi'(\bar{\mu}(\mu))[\bar{\mu}(\mu) - \underline{\mu}(\mu)]} \tag{$\overline{D_2}$}$$

$$\underline{\mu}'(\mu) = \frac{\phi(\underline{\mu}(\mu)) - \phi(\bar{\mu}(\mu))}{\phi'(\underline{\mu}(\mu))[\bar{\mu}(\mu) - \underline{\mu}(\mu)]} \tag{$\underline{D_2}$}$$

Both $\underline{\mu}(\mu)$ and $\bar{\mu}(\mu)$ strictly decrease in μ, while the width of the search region, $\bar{\mu}(\mu) - \underline{\mu}(\mu)$, strictly increases in μ. In addition, if $\underline{\mu}(\mu) \geq 1/2$, then the slope of the quitting boundary is less than -1 and the slope of the purchasing boundary is greater than -1.

We find that the optimal search region has a butterfly shape - the consumer searches for information in a broader region when the consumer is more certain that the more favorable attribute is good. The intuition is the following. The product has a higher expected value if the consumer is more confident about one attribute being good. So, the consumer will search for information about the other attribute even if she has more uncertainty about it. Because the speed of learning is higher when searching a more uncertainty attribute, the benefit of search increases while the search cost remains the same. Therefore, the consumer will search more.

If the consumer likes an attribute more, she will purchase the product even if she has a higher uncertainty about the other attribute. She will also be less likely to stop searching and quit. Therefore, the search region shifts downwards as the belief about one attribute, μ, increases. The value of the slope of the search region is also interesting. It is the marginal rate of substitution between the values of attribute one and two. If the slope equals -1, then the two attributes are perfect substitutes. One may expect this to be the case in general because the product's value is the sum of the values of two attributes. However, both the slope of the quitting boundary and the slope of the purchasing boundary are not -1 in general because of the asymmetry of learning. If the quitting boundary is above 1/2, a unit increase of the belief about attribute one can substitute for more than a unit of the belief about attribute two near the quitting boundary, $\underline{\mu}'(\mu^*) < -1$. The consumer will keep searching for information about attribute two instead of quitting even if μ_2 decreases by slightly more than a unit. This is because the consumer has more uncertainty about attribute 2. The speed of learning is higher when the consumer searches a more uncertainty attribute. So, the benefit of search increases while the search cost remains the same. Similarly, a unit increase of the belief about attribute one can substitute for less than a unit of the belief about attribute two near the purchasing boundary, $\bar{\mu}'(\mu^*) > -1$. The consumer will keep searching for information about attribute two instead of purchasing the product even if μ_2 only decreases by slightly less than a unit.

Given the value function and the optimal strategy derived under the conjectured search strategy, we now verify that the conjectured search strategy is indeed optimal (satisfying the HJB equation $(\star)$).

Theorem 1. *Suppose* $\mu_1 \geq \mu_2$ $(\mu_1 < \mu_2)$ *and* $\mu_1 + \mu_2 > 1$.[6] *If the search cost is low,* $c \leq \frac{1}{2\sigma^2[\phi(1/2)-\phi(\frac{2}{3}p-\frac{1}{6})]}$, *then it is optimal for the consumer to search for information about attribute two (one), conditional on searching.*

We have characterized the search region by a set of ordinary differential equations for moderate beliefs and by a system of equations for extreme beliefs. The optimal search strategy implies that the decision-maker only searches the more uncertain attribute if she holds a strong prior belief on one of the attributes and may search both attributes otherwise. This result is the main testable implication of the paper. Future empirical studies on multi-attribute consumer search can test whether this prediction holds, especially with the aid of the eye-tracking data.

2.3 Comparative Statics

If the firm wants to use the above results, it needs to understand how the model primitives affect the consumer's search behavior. The following proposition summarizes the comparative statics of the search region with regard to the price, search cost, and noise of the signal.

Proposition 11. *Suppose* $\mu_1 \geq \mu_2$. *The purchasing threshold* $\bar{\mu}(\mu)$ *increases in the price* p*, and decreases in the search cost* c *and the noise of the signal* σ^2. *The quitting threshold* $\underline{\mu}(\mu)$ *increases in the price* p*, the search cost* c*, and the noise of the signal* σ^2.

An increase in the price shifts the entire search region upwards because the consumer needs to gain a higher value from the good to compensate for the higher price. For example, as Figure 2.3 illustrates, the consumer may be willing to pay 1.5 for a good when she believes that each attribute has an 80% probability of being good. She will obtain a positive expected surplus from purchasing the product. However, if the price of the good increases to 1.75, she will not buy the good given the same belief because of the negative expected utility. She may not even keep searching for information because the likelihood that the belief becomes high enough to compensate for the high price is low. She will be better off stopping searching,

[6] Note that $\mu_1 + \mu_2 > 1$ always holds in the search region. So, this condition can be omitted. We leave it in the statement to emphasize that the consumer searches for information about the attribute with more uncertainty.

saving the search cost. Similarly, the consumer may be willing to search for more information when she believes that each attribute has a 70% probability of being good if the price is 1.5. Though she will obtain a negative utility from purchasing the product right away, she may like the product more after some search and gain a positive surplus by purchasing it. In contrast, if the price of the good increases to 1.75, she will stop searching because the likelihood of receiving a lot of positive information and raising the valuation for the product above the high price is very low.

Given a prior belief (μ_1, μ_2), increasing the price has two opposite effects on the firm. A higher price raises the profit conditional on purchasing but reduces the purchasing likelihood. The next section discusses in detail how the consumer's purchasing likelihood depends on the prior belief.

The change in the search cost or the signal's noise has the same effect on the consumer's search behavior because they always appear together in the value function as $c\sigma^2$. An increase in either the search cost or the signal noise makes searching less attractive for the consumer and shrinks the search region. The consumer will only search for information in a narrower range of beliefs. Figure 2.4 illustrates how the seach region depends on the search cost and the signal noise. For example, for a product whose price is 1.5, the consumer may want to keep searching if she believes that each attribute has a 78% probability of being good and $c\sigma^2 = 0.1$. She can obtain a positive surplus by purchasing the good immediately. However, she may receive some negative information about the product and aviod purchasing a bad product by mistake. So, she may prefer to make a deicsion when she becomes more certain about the value of the product. However, if it takes more time or effort to search for information or the information is not very accurate, $c\sigma^2 = 0.2$, the benefit from search will be lower and the consumer may instead purchase the good immediately.

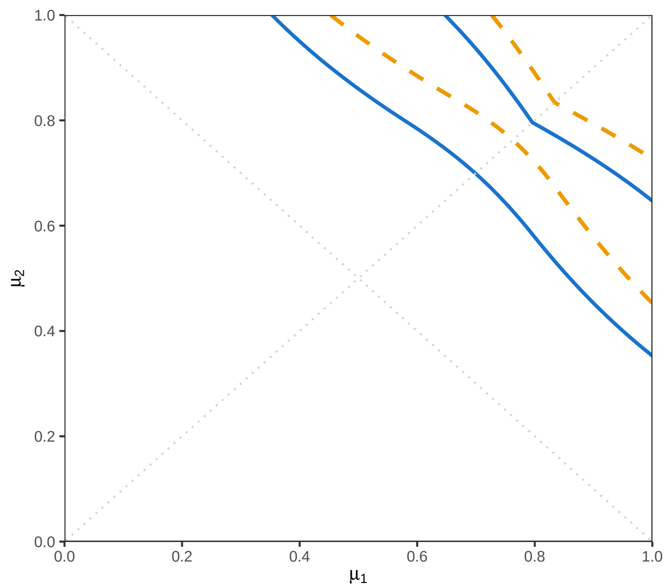


Figure 2.3: Optimal Search Region for $p = 1.5$ (solid blue) or 1.6 (dashed orange), $c = 0.1, \sigma^2 = 1$.

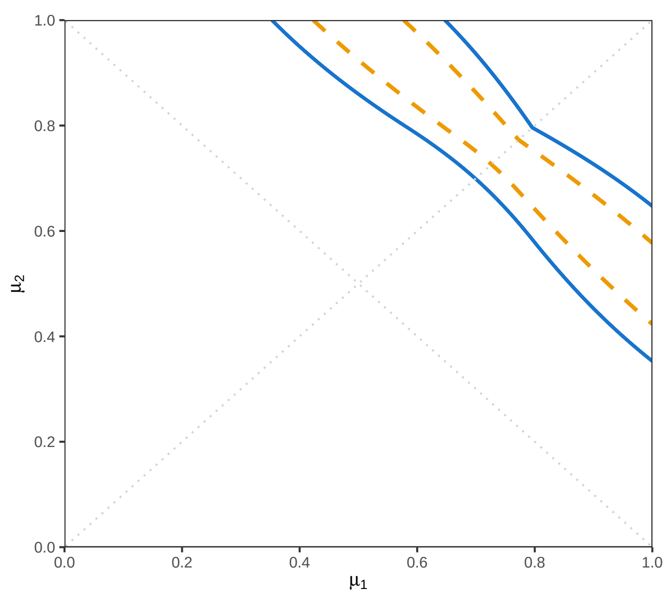


Figure 2.4: Optimal Search Region for $p = 1.5$, $c\sigma^2 = 0.1$ (solid blue) or 0.2 (dashed orange).

2.4 Purchasing Likelihood

We now look at the consumer's belief path to purchase. If the consumer strongly believes that one of the attributes is good, she will never search for information on that attribute. The consumer will keep searching for information about the other attribute. She will purchase the product if she obtains enough positive information and the belief reaches the purchasing boundary $\bar{\mu}$. If she receives enough negative information and the belief reaches the quitting boundary $\underline{\mu}$, she will quit searching without buying the good. For example, when deciding whether to buy a Volvo, a consumer may not bother to search for its safety features because Volvo has a good reputation for safety. She gains more from searching for other attributes of the car.

In contrast, the consumer must search for information on both attributes before purchasing the good if she has mild beliefs about both attributes. Moreover, she will be equally certain about the value of each attribute if she decides to buy the good. For example, Faraday Future has not produced any cars yet. If a consumer considers pre-ordering a car, she probably has a lot of uncertainty about everything. So, she may search for information about every attribute. Given the consumer's optimal search strategy, we can calculate the purchasing likelihood given a prior belief (μ_1, μ_2).

Proposition 12. *Suppose $\mu_1 \geq \mu_2$. The probability that the consumer purchases the product is:*

$$\begin{aligned} P(\mu_1,\mu_2) :=\ & \mathbb{P}[purchasing | starting\ at\ (\mu_1,\mu_2)] \\ =\ & \begin{cases} 1, \ \text{if } \mu_1 \in [\mu^{**},1] \text{ and } \mu_2 \in [\bar{\mu}(\mu_1),\mu_1] \\ \frac{\mu_2-\underline{\mu}(\mu_1)}{\bar{\mu}(\mu_1)-\underline{\mu}(\mu_1)}, \ \text{if } \mu_1 \in [\mu^{**},1] \text{ and } \mu_2 \in [\underline{\mu}(\mu_1),\bar{\mu}(\mu_1)] \\ h(\mu_1,\mu_2)\tilde{P}(\mu_1), \ \text{if } \mu_1 \in [\mu^{*},\mu^{**}] \text{ and } \mu_2 \in [\underline{\mu}(\mu_1),\mu_1] \\ 0, \ \text{if } \mu_1 \leq \mu^{*} \text{ or } \mu_2 \leq \bar{\mu}(\mu_1) \end{cases} \end{aligned}$$

, where $h(\mu_1,\mu_2) = \frac{\mu_2-\underline{\mu}(\mu_1)}{\mu_1-\underline{\mu}(\mu_1)}$ and $\tilde{P}(\mu) = e^{-\int_{\mu}^{\mu^{**}} \frac{2}{x-\underline{\mu}(x)}dx}$. By symmetry, $P(\mu_1,\mu_2) = P(\mu_2,\mu_1)$ if $\mu_1 < \mu_2$.

We can see that there are four regions, as Figure 2.5 illustrates. The consumer makes the purchase immediately if the belief lies in the region $S1$ and quits without purchasing immediately if the belief lies in the region $S4$. For beliefs in between, the value of information is the highest. The consumer will search for more information before making a decision. If the belief lies in the region $S3$ on the right-hand side of the figure, the consumer strongly believes that the first attribute is good. So, instead of spending more time confirming it, she searches for information about the more uncertain attribute, attribute two. If she receives enough positive information about the second attribute, she will be very optimistic about the product's value and will make the purchase. If she receives enough negative information about the second attribute, she will be pessimistic about the product's value and will stop searching. Because the conumer has had a pretty good sense of the first atribute's value, she will not switch back to searching for information about it regardless of what she learns about the second attribute. Therefore, the second attribute is the pivotal attribute in this case.

If the belief lies on the right-hand side of the region $S2$, the consumer is quite uncertain about the value of both attributes. She will search for information about attribute two because she is more uncertain about attribute two than attribute one. However, the consumer also does not have a strong belief about the value of attribute one. So, the consumer will switch to search for information about attribute one if she receives enough positive signals about attribute two. She may switch back to attribute two if she gets enough positive signals about attribute one and may switch back and forth before being confident about both attributes and purchasing the product. As shown in Figure 2.5, the belief must reach (μ^{**}, μ^{**}) for the consumer to make the purchase decision. So, she will be equally confident about the value of both attributes when she stops searching and buying the good. She will stop searching and quit if she receives enough negative signals about either attribute.

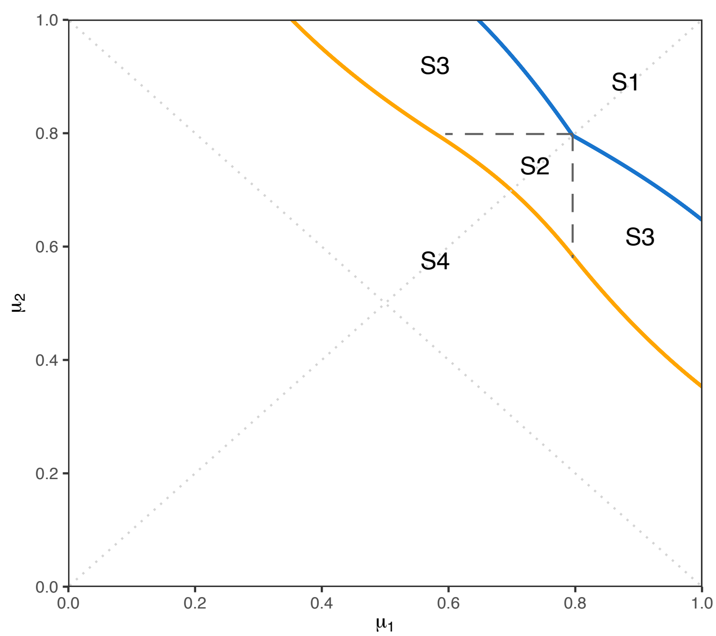


Figure 2.5: Four Regions for Purchase

2.5 Pre-search Intervention

The previous section determines the purchasing probability given the prior belief. In reality, firms can intervene the consumer search and purchase processes through marketing activities such as advertising. The firm can reveal the value of the attribute by *informative* *advertising*. The consumer does not need to incur costs to search for information about the attribute. Due to the limited bandwidth of ads, we assume that the firm can only reveal the value of one attribute.

Informative Advertising

By conducting informative advertising, the firm can disclose the value of one of the attributes. Given the updated information, the consumer can search for more information before making a decision. If the firm advertises one attribute, it reveals its value. So, the consumer only has uncertainty about the other attribute. Her search problem becomes a single-attribute problem. Suppose the firm advertises attribute $i \in \{1, 2\}$. The value of attribute i, U_i, becomes 1 with probability μ_i and 0 with probability $1 - \mu_i$.[7] The consumer can make a decision right away or search for information about attribute $j := 3 - i$. The real price of the product is $p' := p - U_i$. One can see that the consumer will quit if $U_i = 0$. So, we consider the case in which $U_i = 1$ now (p' becomes $p - 1$). The optimal search strategy has been shown in Branco et al. (2012) and Ke and Villas-Boas (2019). There exists $0 < \underline{\mu}_j < \bar{\mu}_j < 1$ such that the consumer searches for more information if $\mu_j \in (\underline{\mu}_j, \bar{\mu}_j)$, purchases the product if $\mu_j \geq \bar{\mu}_j$, and quits if $\mu_j \leq \underline{\mu}_j$. In the search region, the value function is determined by:

$$\begin{aligned} &\frac{\mu_j^2(1-\mu_j)^2}{2\sigma^2} W''(\mu_j) - c = 0 \\ \Rightarrow \ &W(\mu_j) = 2\sigma^2 c(1 - 2\mu_j) \ln \frac{1-\mu_j}{\mu_j} + K_1\mu_j + K_2, \ \mu_j \in (\underline{\mu}_j, \bar{\mu}_j) \end{aligned}$$

Since $W(\underline{\mu}_j) = W'(\underline{\mu}_j) = 0$, $W(\bar{\mu}_j) = \bar{\mu}_j - p'$, and $W'(\bar{\mu}_j) = 1$, value matching and smooth pasting at $\underline{\mu}_j$ and $\bar{\mu}_j$ determine the cutoff belief:

$$\begin{cases} \phi(\underline{\mu}_j) - \phi(\bar{\mu}_j) = \frac{1}{2\sigma^2 c} \\ \psi(\underline{\mu}_j) - \psi(\bar{\mu}_j) = \frac{p-1}{2\sigma^2 c} \end{cases} \tag{2.9}$$

By symmetry, we only need to consider the firm's advertising strategy when $\mu_1 \geq \mu_2$, which is summarized by the following proposition.

Proposition 13. *Suppose $\mu_1 \geq \mu_2$. There exists $\tilde{\mu}(\mu_1)$ and $\widehat{\mu}(\mu_1)$ such that $\underline{\mu}(1) < \tilde{\mu}(\mu_1) \leq \widehat{\mu}(\mu_1) < \bar{\mu}(\mu_1)$ and $\tilde{\mu}(\mu_1)$ decreases in μ_1. The firm does not advertise if $\mu_1 \leq \underline{\mu}(1)$ or $\mu_2 \geq \widehat{\mu}(\mu_1)$, advertises attribute two if $\mu_1 \in (\underline{\mu}(1), \bar{\mu}(1)]$, or $\mu_1 > \bar{\mu}(1)$ and $\mu_2 \leq \tilde{\mu}(\mu_1)$, advertises attribute one if $\mu_1 > \bar{\mu}(1)$ and $\mu_2 \in (\tilde{\mu}(\mu_1), \widehat{\mu}(\mu_1))$.*

[7]We denote $\mu_i(0)$ by μ_i to simplify the notation in this section.

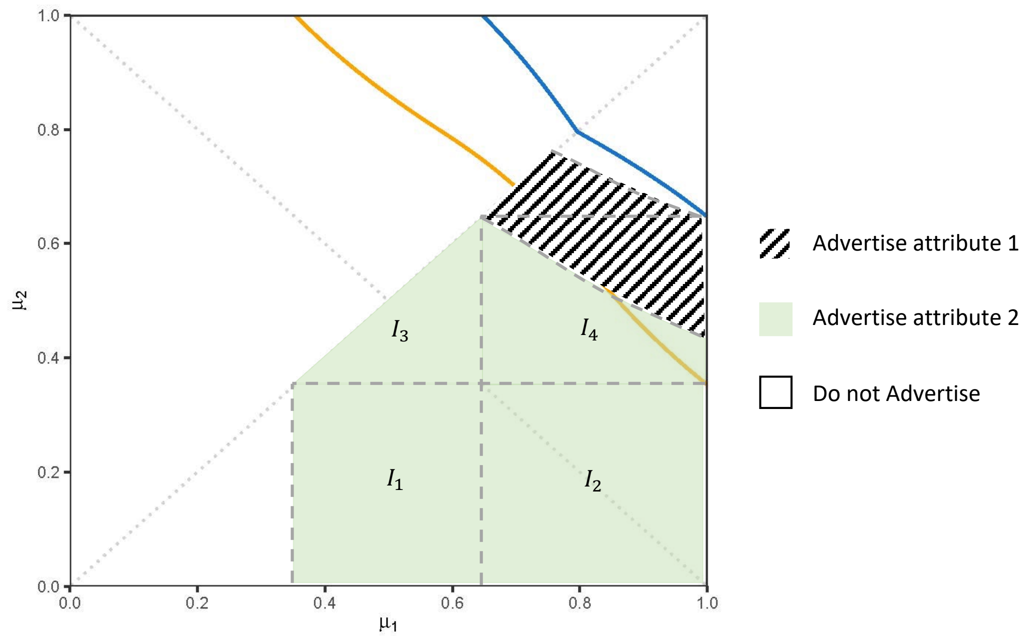


Figure 2.6: Advertising Strategy

Figure 2.6 illustrates the advertising strategy. The firm advertises attribute one in the diagonal striped black region, attribute two in the solid green region, and does not advertise in the white region. If the consumer's prior beliefs about both attributes are too low, the product will not be attractive to the consumer even if she knows that one attribute is good. The consumer will neither search for information nor purchase the product even if the firm advertises. So, the firm does not advertise. If the consumer has high enough prior beliefs about both attributes, she will purchase the product without searching. The firm also has no incentive to advertise. Even if the consumer's belief is within the search region, she will purchase the product after receiving a little positive information as long as her belief is close to the purchasing boundary. The purchasing probability is close to 1. In contrast, if the firm advertises, the consumer will quit for sure if she finds out that one attribute is bad. So, the purchasing probability is lower. The firm is better off by not advertising. The intuition is the following. If the consumer finds out that one attribute is good from advertising, her belief about the product value will be higher than what is needed for her to purchase the product immediately. Such excessive belief is wasteful from the firm's standpoint. If the firm does not advertise, the consumer will be just indifferent between searching for more information and purchasing the product after receiving a little positive information. The firm does not waste any belief. Therefore, the consumer will be more likely to purchase the product without advertising. Therefore, the firm does not advertise in the white region.

Now let's consider the solid green region and the diagonal striped black region. We divide the solid green region into four sub-regions. If the belief lies in the region I_1 or I_2, the consumer is very pessimistic about the second attribute. Even if she knows for sure that the first attribute is good, she needs to receive a lot of positive signals about attribute two to purchase the product. The search cost outweighs the benefit of the search. So, she will not search for information. The only way of inducing the consumer to search is to advertise attribute two. With a high probability, the consumer will find out that attribute two is bad and quit. However, if the consumer find out that attribute two is good, she needs fewer positive signals to purchase the product by searching for attribute one. The benefit of search outweighs the search cost. So, the consumer will search for information about attribute one and purchase the product with a positive probability. Therefore, the firm advertises attribute two.

If the belief lies in the region I_3, the consumer will never purchase the product without advertising but may purchase the product if the firm advertises either attribute. So, the firm advertises. Since the consumer is more optimistic about attribute one, she will be more likely to search for information if the firm advertises attribute one than two. However, she needs more positive signals to purchase the product. So, the conversion rate conditional on searching is lower. It turns out that the second effect is stronger than the first effect. So, the firm advertises attribute two.

Lastly, we consider the case where the belief lies in the region I_4 or the diagonal striped black region. If the consumer's belief about attribute two is high enough, she will purchase the product immediately if she knows attribute one is good. One can see that the firm always prefers to advertise attribute one to attribute two. If the consumer's belief about attribute two is lower, the comparison between advertising attribute one and two is non-trivial. If the firm advertises attribute one and the consumer finds out it is good, the consumer will always search for information about attribute two before making a decision. In contrast, the consumer will be very positive about the product value if the firm advertises attribute two and the consumer knows that attribute two is good. In that case, she will purchase the product immediately. So, some beliefs are "wasted" - the consumer will purchase the product immediately even if her belief is lower. The more optimistic she is about the first attribute, the more beliefs are wasted. So, the firm will be more likely to advertise attribute one.

In sum, the firm will not advertise if the consumer's prior beliefs about both attributes are extreme and will advertise if the consumer's prior belief is milder. In that case, the firm will advertise the better attribute if the consumer is optimistic enough about the worse attribute, and will advertise the worse attribute if the consumer is less optimistic about it.

Advertising Costs

In the previous discussion, we did not consider the advertising costs. In reality, the firm needs to incur a cost to advertise. Our framework can incorporate this cost, but the analysis will be more tedious. So, we abstract away the advertising costs in the previous analysis.

We briefly discuss what happens if we take into account the advertising costs. Suppose the firm needs to incur a cost c_A to advertise attribute i. The comparison between advertising attribute one and two will not change because both require an extra cost, c_A. However, whether the firm prefers to advertise or not may change. If the prior belief of the consumer without advertising is close to the purchasing boundary, then the firm will not advertise. Even without advertising, the consumer will purchase the product with a high probability. By not advertising, the firm saves advertising costs. The firm will also not advertise if the belief about one of the attributes is too low. Even if the firm can raise the purchasing probability above zero by advertising, the purchasing likelihood is very low. The profit will be negative because of the advertising costs. So, the firm will not advertise, and the consumer will neither search nor purchase. For all other beliefs, the firm's advertising strategy is the same as the case without advertising costs.

2.6 Conclusion

Understanding how consumers decide which attribute to pay more attention to has important managerial implications. It helps the firm decide how to design the product and which attributes to emphasize. In this paper, we study the optimal search strategy of a Bayesian decision-maker by endogenizing the optimal attribute to search for, when to keep searching, and when to stop and make a decision. We characterize the search region by a set of ordinary differential equations for moderate beliefs and by a system of equations for extreme beliefs. We find that it is always optimal to search the attribute the consumer has the highest uncertainty due to the fastest learning speed. The decision-maker only searches the more uncertain attribute if she holds a strong prior belief on one of the attributes, and may search both attributes otherwise. We also study the firm's optimal pre-search intervention by assuming that it can disclose the value of one attribute by informative advertising. We find that the firm will not advertise if the consumer's prior beliefs about both attributes are extreme. If the consumer is very optimistic about both attributes, she will purchase the product for sure or with a very high likelihood. So, the firm does not have an incentive to advertise. If the consumer is very pessimistic about both attributes, she will never purchase the product even if she knows that one attribute is good. So, the firm does not advertise either. If the consumer's prior belief is milder, the firm can increase the purchasing probability by advertising. The firm will advertise the better attribute if the consumer is optimistic enough about the worse attribute, and will advertise the worse attribute if the consumer is less optimistic about it.

There are some limitations to this paper. The consumer only considers one product in our model. If there are multiple products, the consumer needs to make two decisions - which product to search for and which attribute of the product to search for. Studying this richer problem can lead to interesting findings. It will also be interesting to extend the number of attributes beyond two and see whether the consumer still searches for the attribute with the highest uncertainty due to the fastest learning speed. Lastly, we consider an exogenous price throughout the paper to focus on the role of information. Future research can study the optimal pricing of the product given the consumer's optimal search strategy.

Chapter 3

Failure of Reputation for Privacy

3.1 Introduction

The information market emerges in the digital era. The business of collecting and selling consumer data is estimated to be worth around \$200 billion.[1] Firms use detailed information about individuals to offer a personalized product, price discriminate, show targeted ads, etc. Aware of the costs of revealing information, consumers are becoming increasingly concerned about their privacy. People started to raise concerns about their privacy even in the 1990s. About 0.01% of the US population opted out of the database of Lotus MarketPlace.[2] But most people at that time were either not aware of the privacy issues or did not care much about it. A recent survey by KPMG in 2021[3] among the US general population found that 86% of consumers viewed data privacy as a growing concern. One of the reasons people worry about the firm collecting their data is that they do not know how the firm will use it. According to the same survey, 40% of the consumers do not trust the firm to use their data ethically. Taylor (2004) shows that the firm can be better off by not protecting consumer privacy (selling customer data) if consumers are naive and unaware of it. However, selling data can backfire if consumers are sophisticated and expect the firm to sell their data. A large body of literature has documented that commitment benefits the firm. As a result, companies pay increasing attention to privacy. Apple, for instance, invested heavily in the operating systems to protect consumer privacy and spent lots of resources advertising their progress in privacy protection. In this particular setting, the firm desires the ability to commit to protecting consumer privacy by not selling consumer data. However, the non-verifiable nature of digital data makes it hard for the firm to commit. This paper looks at one possible solution - building trust by reputation.

The main contribution of the paper is to characterize some sufficient conditions such that

[1]https://www.latimes.com/business/story/2019-11-05/column-data-brokers

[2]https://www.forbes.com/sites/forbestechcouncil/2020/12/14/the-rising-concern-around-consumer-data-and-privacy/?sh=73c76330487e

[3]https://advisory.kpmg.us/content/dam/advisory/en/pdfs/2021/corporate-data-responsibility-bridging-the-consumer-trust-gap.pdf

reputation considerations cannot serve as a commitment device for privacy, even if firms are arbitrarily patient. When the firm is a monopoly and is patient enough, reputation enables it to commit never to sell the data. It achieves the Stackelberg payoff in all but a finite number of periods. However, when there are multiple firms, reputation may fail to enable any firms to commit. The intuition is that one firm's reputation depends on other firms' actions. Selling data by one firm has a negative externality on other firms. Firms do not take it into account in equilibrium. So, the benefit of not selling data is lower, as other firms' behavior may still hurt the firm's reputation. Anticipating such externality, consumers penalize each firm less when observing data sales. In addition, the likelihood of the deviation being pivotal reduces in the number of firms. Therefore, the cost of selling data is lower. So, the firm has more incentive to deviate. When the number of firms is large, or the monitoring technology is good, the incentive for the firm is so strong that no firm could commit never to sell the data. This reputation failure result hurts all the firms.

We consider long-lived firms interacting with short-lived consumers repeatedly in two markets. In the product market, the consumer decides how much information to reveal. Each firm infers consumer preferences based on the revealed information and offers a personalized product and price. The consumer then makes the purchase decision. By revealing more information, she[4] gets a better recommendation. However, the firm will charge a higher price when it collects more information from the consumer, knowing that she has a higher expected valuation for the product. So, the consumer faces a tradeoff between better product fit and lower price. In the information market, the firm could sell consumer data to third parties (e.g., data intermediaries). Consumers may suffer disutility from the sale of their data. For example, they may experience scam emails/calls or account hacking. If consumers reveal more information, they will be more vulnerable to data sales. Therefore, the consumer's decision of how much information to reveal in the product market depends on her belief about the firm's behavior in the information market. If the consumer thinks the firm will sell her data, she will reveal no information to minimize the cost of privacy loss. If she trusts the firm not to sell her data, she will reveal some information to get a better product recommendation. The Stackelberg action of the firm is not to sell data. But the decision of whether to sell data or not is made after the consumer reveals the information. Hence, the holdup problem prevents the firm from doing so in a static setting.

This paper studies whether the reputation consideration of the firm can serve as a commitment device in a long-run game when consumers have imperfect monitoring technology. Reputation can be a commitment device for a patient enough monopoly but may fail to be one when there are multiple firms, even when firms are arbitrarily patient. In particular, we have a reputation failure result when the number of firms is large, or the difference between the payoff with and without commitment is small, the likelihood of selling data being pivotal is low, and the privacy loss of the consumer is high. The intuition is that the monopoly will never restore its reputation by deviating from selling the data and being caught. The high and permanent reputation cost provides a strong incentive for the monopoly to commit

[4]We refer to the consumer as "she" throughout the paper.

to privacy. In contrast, when there are multiple firms, consumers do not know which firm exactly sold the data, even if they observe data sales. Therefore, the penalty for selling data is lower, and a firm's reputation may be hurt even if it did not sell data. The low and temporary reputation cost provides a strong incentive for the firm to deviate.

We consider several extensions to the main model. Consumers can voluntarily incur efforts to monitor firms better. We find that endogenous monitoring helps a monopoly build up a reputation faster, benefiting both the rational firm and consumers. However, it does not provide enough incentives for multiple firms to commit not to sell data. Also, we consider asymmetric monitoring. The monopoly case implies that rational firms can commit without noise. In contrast, any noise from other firms will break down the commitment power. This fragility result shows that the possibility rather than the level of interaction of firms' behavior in the reputation-building process is critical to the reputation failure.

Literature Review

This paper contributes to the literature on the economics of privacy (see Acquisti et al. for a survey). Goldfarb and Tucker (2012) and Lin (2019) document the existence of substantial privacy concerns of the consumer. In a static framework, Ichihashi (2020) shows that sellers prefer to commit to the price of the good for the buyers to reveal more information. We investigate when such commitment is feasible without an external commitment device. Recent papers have paid much attention to the economic impact of regulations such as GDPR, CCPA, and AdChoices, which seek to protect consumer's privacy and give them more control over their data (Athey et al. 2017, Ke and Sudhir 2020, Goldberg et al. 2019, Goldfarb and Tucker 2011, Johnson et al. 2020, Johnson et al. 2021). There are two reasons why reputation is essential despite various regulations. First, the main focus of those regulations is to give consumers more control over the usage of their data rather than to provide the firm with commitment power. Second, the transparency and verifiability nature of data transactions raises concerns about the credibility of such policy. Even if firms do not sell data in the presence of such regulation, consumers may still not reveal enough information to the firm. Protecting consumer privacy will not benefit the firm if it fails to obtain consumers' trust about how firms handle their data. Absent the information market and the possibility of selling data, Chen and Iyer (2002) study competing firms' incentives to collect data. They find that firms may voluntarily collect less information about consumers to mitigate the price competition. Closely related to our paper, Jullien et al. (2020) study a website's incentive to sell consumer information in a two-period model. Unlike in our paper, the website in their paper does not try to change consumers' beliefs about its type. Instead, the website wants to affect consumer's behavior on the vulnerability of bad experiences due to data sales.

This paper is also related to the literature on reputation. The idea of modeling reputation by incomplete information comes from Kreps et al. 1982, Kreps and Wilson (1982), and Milgrom and Roberts (1982). Fudenberg and Levine (1989) show that a patient long-run player will commit to the Stackelberg action in the presence of a behavioral type and perfect monitoring. Reputation serves as a commitment device and selects away bad equilibria for the long-lived player. In contrast, Ely and Välimäki (2003) and Morris (2001) show that reputation concerns may hurt the firm under imperfect monitoring. Substantively, the paper most closely related to us is Phelan (2006), which studies a problem where the government builds a reputation for trust. The reputation shock is non-permanent despite perfect monitoring because the government's type can change over time. Tirole (1996) studies the economics of collective reputation. Similar to our paper, individual reputation and incentive depend not only on one's past behavior but also on other players' because of the noisy signal. The inability to build a reputation relies on the different arrival times of the players. In our paper, reputation failure is driven by the externality of one firm's behavior on the other one's reputation rather than the arrival time. Despite the long development of this literature, people have not paid much attention to the reputation for privacy. This paper shows that reputation may fail to help the firm commit when their reputations depend on each other's behavior, and there is a bad type who does not care about consumer privacy. It connects the bad reputation and collective reputation literature.

The remainder of the paper is organized as follows. Section 2 presents the main model. Section 3 shows the ability of the monopoly to commit. Section 4 characterizes some sufficient conditions under which reputation fails to serve as a commitment device when there are multiple firms. The next two sections consider several extensions to the main model. Section 5 studies endogenous monitoring. Section 6 studies asymmetric monitoring. Section 7 concludes.

3.2 Model

Time is infinite, $t = 0, 1, 2, ...$ and the discount factor is δ. There are N long-lived firms and a short-lived consumer at each period. The consumer interact with all the firms. The firm's payoff is $(1-\delta)\sum_{t=0}^{+\infty}\delta^t u_t$, where u_t is the stage payoff at time t. There is a product market and an information market.

Product Market

Consumers have different horizontal preferences and are located uniformly on a circle with a circumference of 1. When a consumer visits the firm in the product market, she chooses how much information to reveal. If the consumer locating at $x \sim [0,1)$ reveals $\eta \in [0,1]$ proportion of information, the firm gets a noisy signal $l \sim U[x-(1-\eta)/2 \mod 1, x+(1-\eta)/2 \mod 1]$ about the consumer's location,[5] as illustrated by Figure 3.1. Based on the signal, the firm offers a personalized product and sets the price p. The consumer then makes the purchase decision. Denote the distance between the product's location, y, and the consumer's location, x, by $d = |x-y|$. We have that $d \sim U[0, (1-\eta)/2]$. The baseline valuation of the product is v, and the disutility from the mismatch of the recommended product and the consumer's

[5] This implies that the firm will offer a product located at l given the signal.

horizontal taste is td. Therefore, the consumer gets $v - td - p$ if she buys and 0 if she does not buy. We assume that there is enough horizontal differentiation, $t > v$.

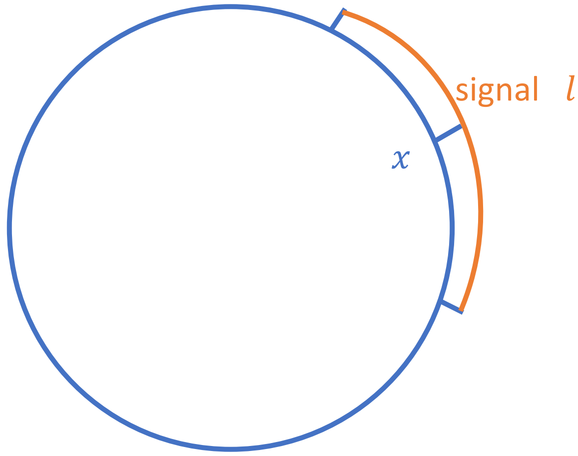


Figure 3.1: Consumer's location x and the signal l

Given the product recommendation and price, the consumer purchases if and only if the expected payoff is positive, $v - td - p \geq 0$. Thus, the firm's problem is:

$$\max_p p \cdot \mathbb{P}[v - td - p \geq 0] = p\left[\frac{2(v-p)}{(1-\eta)t} \wedge 1\right]$$

Therefore, the optimal price is:

$$p^*(\eta) = \begin{cases} \frac{v}{2}, \ if \ \eta \leq 1 - \frac{v}{t} \\ v - \frac{(1-\eta)t}{2}, \ if \ \eta > 1 - \frac{v}{t} \end{cases}$$

When the consumer reveals a lot of information, the firm accurately knows her preference. The optimal price makes the consumer located farthest away from the recommended product indifferent between purchasing or not. Therefore, she always makes the purchase. When the consumer reveals less information, the firm gets a noisier signal about her preference. The profit from each purchase will be too low if the firm wants the consumer always to buy the product. Therefore, only consumers with a high enough valuation for the recommended product purchases it under the optimal price. If the recommended product locates far away from the consumer, the consumer will not buy it.

Information Market

The firm could sell consumer data in the information market to third parties (e.g., data intermediaries). For each consumer, the firm has the data directly revealed by her and the behavioral data of whether she makes the purchase given the product and price offered.[6] The firm gets $D(\eta)$ by selling the data. We assume that $D(\eta)$ increases in η to reflect that more accurate information is more valuable. The consumer might experience a scam or account hack if the firm sells data. Consumers are more vulnerable to such undesired activities when they reveal more information. So, we assume that the expected privacy cost of the consumer is ηu_b.[7] Consumers could imperfectly monitor the behavior of the firm in the information market. If a firm sold data in the previous period, the consumer detects it with probability q. The consumer will receive a signal $s = y$ if they caught any of the sales and $s = n$ if they did not detect any sales.[8]

Denote the probability of the firm selling the data by μ_s, by picking the privacy level η, the consumer's expected ex-ante payoff is:

$$U_0(\eta) = \begin{cases} -\mu_s \eta u_b + \frac{v^2}{4(1-\eta)t}, \; if \; \eta \leq 1 - \frac{v}{t} \\ -\mu_s \eta u_b + \frac{(1-\eta)t}{4}, \; if \; \eta > 1 - \frac{v}{t} \end{cases}$$

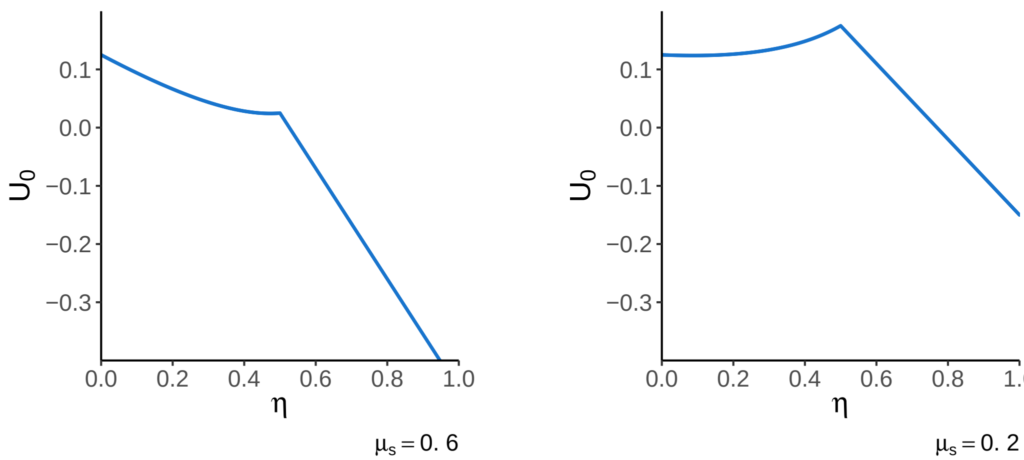


Figure 3.2: Ex-ante consumer payoff as a function of η for $v = 1,\ t = 2,\ u_b = 0.75$, and $\mu_s = 0.6$ (left) or 0.2 (right).

In the product market, revealing more information has two opposite consequences. On the one hand, the firm could offer a better-matched product, which benefits the consumer.

[6]Firms can infer consumers' willingness to pay from the data they reveal directly (Bergemann et al. 2020), or from the behavioral data of the consumer (Shen and Villas-Boas 2018, Taylor 2004, Villas-Boas 1999, 2004).

[7]Changing it to $\eta u_b + k$ won't change the result qualitatively, and we choose this form for simplicity.

[8]The assumption that consumers cannot distinguish which firm sold the data gives the sharpest illustration of the main idea. We extend it later to give consumers a better sense of which firm sold the data.

On the other hand, the firm will charge a higher price, knowing that the consumer has a higher expected valuation.[9] Price discrimination hurts the consumer. In the extreme case, if the firm perfectly knows the consumer's preference, it will extract all the consumer surplus. Therefore, the consumer never reveals all the information. The firm can only recommend a random product if the consumer does not reveal anything. The poor match also hurts the consumer. So, it is optimal for her to reveal the information partially. However, the consumer also needs to consider the effect of information revelation in the information market. Disclosing more information to the firm always hurts the consumer there, as the consumer is more vulnerable when the firm sells her data. As Figure 3.2 illustrates, revealing too much information is never optimal for the consumer. The firm can charge a high price because the product recommendation is very accurate with lots of information about the consumer. In addition, the privacy loss from data sales in the information market is also high. Consumers may, however, prefer revealing a moderate amount of information to revealing nothing. By revealing some information, consumers benefit from a better recommendation in the product market but suffer a privacy cost if the firm sells it in the information market. If the firm's likelihood of selling the data is high, the high expected privacy loss in the information market outweighs the gain from the better match in the product market. Thus the consumer reveals no information. On the contrary, If the firm's likelihood of selling data is low, the consumer partially reveals her preference for a better recommendation. We have the following result.

Proposition 14. *The optimal amount of information to reveal is* $\eta^* = \begin{cases} 1 - v/t, \ if \ \mu_s \leq \widehat{\mu} \\ 0, \ if \ \mu_s > \widehat{\mu} \end{cases}$ *where* $\widehat{\mu} = \frac{v}{4u_b}$.[10]

Corollary 1. *The firm's profit in the product market is* $\Pi^* = \begin{cases} v/2, \ if \ \mu_s \leq \widehat{\mu} \\ v^2/2t, \ if \ \mu_s > \widehat{\mu} \end{cases}$.

If the firm could commit not to sell consumer data, the consumer will pick $\eta = 1 - v/t$, which gives the firm a stage payoff of $v/2$. If, instead, the firm always sells consumer data, the consumer will not reveal any information by picking $\eta = 0$. The firm obtains a stage payoff of $v^2/2t + D(0)$. When the following assumption holds, the firm will prefer to commit to protecting consumer privacy and never sell the data. The benefit from the increased profit from the product market outweighs the cost of not selling consumer data.

Reputation

There are two types of firms. A behavioral type (type B) always sells the consumer data. A rational type (type R) maximizes the expected sum of discounted utilities. The firm's reputation is consumers' belief about the probability that the firm is type B. The common

[9]For more discussions about this kind of holdup problem, see Villas-Boas (2009) and Wernerfelt (1994).

[10]For the problem to be interesting, we assume that the threshold $\widehat{\mu} \in (0, 1)$. Also, consumers are indifferent between $\eta = 1 - v/t$ and 0. We assume they choose $\eta = 1 - v/t$, which does not affect any analyses.

priors on each firm being type B are $\mu_0 \in (0, 1)$. Consumers update the belief of the firm's type by Baye's rule. Denote the belief about firm i's type at time t by $\mu_{i,t}$. Reputation for privacy in this paper refers to the reputation for protecting consumer privacy in the information market. Figure 3.3 illustrates the timing of the game.

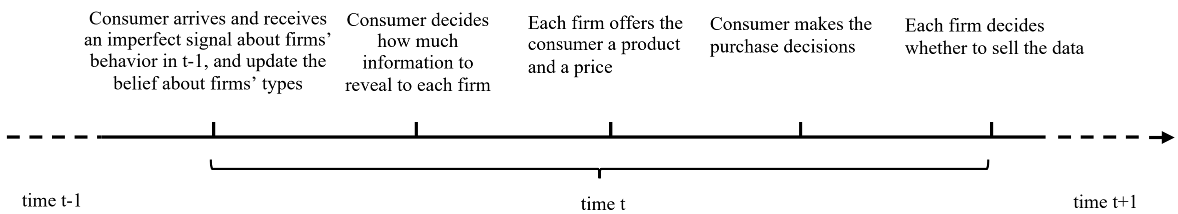


Figure 3.3: Timing of the Game

Solution Concept

We focus on whether a Markov Perfect Equilibrium (MPE)[11] exists where rational firms could commit to never selling the data. To make the problem interesting, we assume that rational firms prefer commitment, $v/2 > v^2/2t + D(0)$. MPE requires that firms' and consumers' strategies are measurable with respect to some payoff relevant states. It is widely used in the reputation literature as the belief $\vec{\mu}_t = (\mu_{1,t}, \mu_{2,t}, ..., \mu_{N,t})$ is a natural state variable.[12]

3.3 Reputation as a Commitment Device for Monopoly

Stage Game

We first analyze the property of the stage game of a single firm. If the firm can commit to any action in the information market (e.g., by moving first), the firm takes the *Stackelberg action* and obtains the *Stackelberg payoff.*

Definition 1. *Suppose player 1 chooses action $a \in A$ and player 2 chooses action $b \in B$. Player $i \in \{1, 2\}$'s stage-game payoff is $u_i(a, b)$. $BR_2(a) \subset B$ is player 2's best response*

[11]Technically, the solution concept we use is the Markov Perfect Bayesian Equilibrium since there is incomplete information about the firm's type. However, the reputation literature usually uses the notion MPE.

[12]In the symmetric equilibrium where every firm always has the same reputation, we can use $\mu_t = \mu_{i,t}$ as the state variable.

correspondence to a. Then, player 1's Stackelberg action is $\arg\max_{a\in A}[\min_{b\in BR_2(a)} u_1(a,b)]$*, and player 1's Stackelberg payoff is* $\max_{a\in A}[\min_{b\in BR_2(a)} u_1(a,b)]$.

One can see that the Stackelberg action for the firm is not to sell the data. The consumer will reveal $\eta = 1 - v/t$ proportion of information, and the firm gets the Stackelberg payoff of $v/2$. If the consumer acts first and minimizes the firm's payoff, one can see that the consumer reveals no information, and the firm sells data. The firm gets the *minmax payoff* of $v^2/2t + D(0)$, which is the payoff the firm can guarantee regardless of the consumer's action.

Definition 2. *Suppose player 1 chooses action* $a \in A$ *and player 2 chooses action* $b \in B$*. Player* $i \in \{1,2\}$*'s stage-game payoff is* $u_i(a,b)$*. Then, player 1's minmax payoff is* $\min_{\beta\in\Delta(B)}[\max_{a\in A} u_1(a,b)]$.

However, since the firm decides whether to sell data after the consumer reveals the information, it always sells the data in a static game. We will see that reputation considerations enable the monopoly to commit to the Stackelberg action.

Belief Updating

We first derive the belief updating of the consumers about a monopoly's type, assuming that the rational type never sells the data. We could derive the belief updating when the rational firm uses other strategies by similar methods. Consider the consumer's belief about the monopoly's type after observing a signal s. If $s = y$, the consumer knows that the firm sold the data in the previous period. So, the belief that the firm is a bad type will be one forever. The firm suffers a permanent reputation shock. If $s = n$, either the firm is the rational type and did not sell the data, or the firm is the bad type, but the consumer did not observe the data sales. The firm is more likely to be the rational type, but the consumer does not know it for sure. So, the belief of firm 1 being bad type decreases from μ_t to $\mu_{t+1} > 0$. Formally, the belief updating is as follows.

Proposition 15. *Suppose the rational type never sells the data in equilibrium.* $\mu_{t+1} = \begin{cases} \frac{1-q}{1-q\mu_t}\mu_t, \ if \ s = n \\ 1, \ if \ s = y \end{cases}$*. After receiving signal* n *for* k *consecutive periods, the belief becomes* $\mu_{t+k} = \frac{(1-q)^k}{(1-q)^k\mu_t + 1 - \mu_t}\mu_t$*, which approaches 0 as* $k \to +\infty$.

So, if the monopoly keeps not selling consumer data, the consumer's belief will keep decreasing. After enough time, the consumer is almost certain that the firm is not the bad type.

Equilibrium

Suppose consumers expect the rational firm never to sell the data in equilibrium. In that case, a signal n will destroy the firm's reputation by making the consumer believe that the firm is the bad type in all of the current and future periods. Then, the firm is stuck with the minmax payoff. By deviating, the firm risks being detected by the consumer with a positive probability. The persistent punishment gives the firm a strong incentive not to sell the data for short-term benefit. As a result, regardless of the monitoring technology or the price of data in the information market, the monopoly can always achieve commitment by reputation, as long as it is patient enough.

Proposition 16. *There exists a $\widehat{\delta} < 1$ such that for any $\delta > \widehat{\delta}$, there exists a MPE where the rational firm never sells consmer data and the consumer always reveals $\eta = 1 - v/t$ proportion of information after a finite period of time.*

By protecting consumer privacy, the rational firm keeps reducing the consumer's belief that it is a bad type. When the belief is below a threshold, the consumer is willing to reveal some information that benefits the firm. A patient firm does not want to deviate, as the consumer may observe the deviation and believe that the firm is the bad type. She will never reveal any information. So, the firm suffers from lower revenue in the product market forever. This severe punishment provides a strong incentive for the firm to trade the short-term benefit of selling data in the information market for the long-term benefit of earning a higher profit in the product market.

3.4 Reputation Failure with Multiple Firms

When there is more than one firm, reputation may fail to serve as a commitment device for privacy. The difference comes from the interaction of firms' behavior in the reputation-building process.

Belief Updaing

When there are multiple firms ($N \geq 2$), the belief updating is qualitatively different from the monopoly case. Consider the consumer's belief about firm 1's type after observing a signal s, assuming that a rational firm never sells the data. If $s = y$, the consumer knows that at least one firm sold the data in the previous period but is not sure whether firm 1 sold it. So, the belief of firm 1 being bad type increases but is still lower than 1, unlike the monopoly case. The reputation shock is temporary, and firm 1 can rebuild the reputation. Conditional on other firms' behavior, the likelihood that $s = n$ if firm 1 is the rational type and did not sell the data is higher than if firm 1 is the bad type, but consumers did not observe the sale of data. Consequently, firm 1 is more likely to be the rational type, but the consumer is uncertain. So, the belief of firm 1 being a bad type decreases but is still positive. Formally, the belief updating is as follows.

Proposition 17. *Suppose the rational type never sells the data in equilibrium.* $\mu_{t+1} = \begin{cases} \frac{1-q}{1-q\mu_t}\mu_t, \; if \;\; s = n \\ \frac{1-(1-q)(1-q\mu_t)^{N-1}}{1-(1-q\mu_t)^N}\mu_t, \; if \;\; s = y \end{cases}$. μ_{t+1} *does not depend on the number of firms* N *if* $s = n$ *and decreases in* N *if* $s = y$.

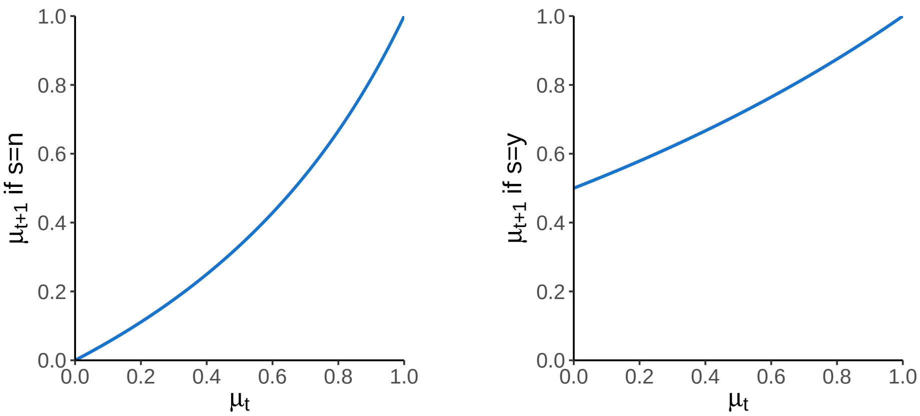


Figure 3.4: Belief Updating as a function of μ_t for $q = 0.5$ and $N = 2$.

Even if firm 1 sold the data and the consumer observes a signal y, she knows at least one firm sold the data but did not know the exact firm. Therefore, she penalizes firm 1 less than she will do in the monopoly case. Even if the consumer obtains a signal y, she does not know for sure that firm 1 is the bad type and the updated belief is lower than 1. Therefore, the reputation cost is temporary and the consumer's belief will decrease if she receives signal n in the future. When there are more firms, the signal's noise is larger, and the consumer has less idea about which firm sold the data. Therefore, the belief increase upon getting signal y will be smaller. If firm 1 did not sell the data and the consumer observes a signal n, the belief reduction does not depend on the number of firms. So, the firm is penalized less for selling the data but not rewarded more for not doing so. In addition, the realization of the signal depends little on a single firm's action when there are many firms. So, the likelihood that firm 1's action is pivotal decreases in the number of firms. Figure 17 illustrates the belief updating when there are two firms.

The above forces imply that the firm has more incentive to sell the data when the number of firms increases.

Fixed Discount Factor

We first fix the discount factor and look at the effect of the number of firms on reputation building.

Proposition 18. *For any $\delta \in (0,1), \exists N_\delta$ s.t. $\forall N \geq N_\delta$, firms always sell data and consumers reveal nothing in the unique MPE. Consumer's belief about each firm's type is always μ_0.*

No matter how patient firms are, they cannot build a reputation for privacy. The intuition for the failure of reputation as a commitment device is the following. On the one hand, the consumer has a noisy signal about which firm sold the data. Even if the firm deviates and the consumer observes it, the penalty for that particular firm is less than that for the monopoly firm. Moreover, it decreases in the number of firms. On the other hand, even if all the firms do not deviate, other firms may be the bad type and sell the data. So, it is less likely that the sale of the data is pivotal as the number of firms increases. Both forces give the firm more incentive to deviate and sell the data. So, it becomes harder to commit when the number of firms increases. Eventually, the firm loses all the commitment power and sells data every period.

Anticipating the firm always to sell data, consumers do not reveal anything. The belief of each firm's type remains the same over time, and there is no reputation building. The monopoly can get Stackelberg payoffs in all but a finite number of periods under substantial punishment for selling data. In contrast, each firm can only get the minmax payoff under weaker punishment when there are multiple firms, even if there is no competition.

Fixed Number of Firms

In this section, we study whether it is possible to achieve commitment by reputation when the number of firms is fixed. The previous section shows that it is harder to commit by reputation when there are more firms. So, we look at the case of two firms. If reputation can not help duopoly commit, we will also have reputation failure when there are more firms.

Proposition 19. *Suppose there are two firms. There does not exist any MPE in which any rational firms could commit to never sell the data, even when $\delta \to 1$, if the following conditions hold:*

$$(1-q)(v/2 - v^2/2t) < D(0) \tag{3.1}$$

$$q(1-q)v/2 < D(0) \tag{3.2}$$

$$v < 2u_b \tag{3.3}$$

Even if firms are very patient, reputation may not suffice to be a commitment device for privacy. This proposition identifies sufficient conditions under which firms cannot commit even if they are (almost) perfectly patient. The role of each condition is the following. Condition (3.1) requires that the commitment payoff is not much higher than the one without commitment. Condition (3.2) requires that the likelihood of selling data being pivotal is low. Because of the imperfect monitoring technology, consumers may get a signal y if a firm did not sell data and n if a firm sold data. If the signal's noise is very high, the consumer is likely to get a signal y even if a firm did not sell data because the other firm sold the data. If it is very low, the consumer is likely to get a signal n even if a firm sold data because of the poor monitoring. Condition (3.3) requires that the privacy loss of the consumer is high enough such that consumers will not reveal information if the firm is equally likely to sell data or not. According to the belief update formula in proposition 17, a single signal y will increase the belief above 1/2. Thus, consumers will reveal nothing to the firm when they get a signal y. Even if the firm does not sell data and reduces the belief, a single signal y in the future periods will make the consumer reveal no information. The fast depreciation of reputation makes building it less attractive. Equivalently, rational firms have a stronger incentive to sell data. When all these three conditions hold,[13] selling data does not hurt the reputation much, a better reputation increases firm's stage payoff slightly, and good reputation is highly non-persistent. Firms have little incentive to build a reputation. As a result, reputation considerations provide no commitment power to rational firms.

Managerial Implications

Even though commitment may be desirable for the firm, it may not be possible without strict external regulations. A monopoly can always build a reputation for caring about consumer privacy by not selling data. After a finite period, consumers will reward it by sharing more information. The monopoly can enjoy a high profit by recommending better-fit products and charging a premium. However, when there are multiple firms in the market, it may not be in the firm's best interest to protect consumer privacy. Even if a firm never sells consumer data, it may not be able to build a reputation for privacy. So, it loses the revenue from selling consumer information while does not have any (or enough) gain. Since firms benefit from committing never to sell consumer data, they need to think about other ways of achieving the commitment. Our model shows that the key to the commitment power is the tradeoff between the short-term benefit of increased revenue in the information market and the long-term benefit of the profit in the product market. A potential solution is to improve the recommendation algorithm so that the firm has a higher marginal benefit from the information collected from consumers. It will have a stronger incentive to maintain a good reputation and profit from the product market. The other solution is to invest in better monitoring technology to make it easier for consumers to identify what specific firm sells the data. Lastly, the firm can offer some compensation to consumers if the signal is y. It will face an additional penalty for selling consumer data. Therefore, the "free lunch" in the information market is more costly for the firm.

3.5 Endogenous Monitoring

The monitoring technology is exogenous in the main model. In reality, consumers observe some data sales without any effort. They know that the phone number has been sold if they get a scam call. If they get a pre-approved credit card with their name in the mail, they know

[13]For example, these conditions will hold if the privacy cost for consumers is high and the monitoring technology has low noise.

that some firms have sold their address and credit history. However, as consumers become more concerned about privacy issues, they may endogenously invest in better monitoring firms' use of their data by incurring more effort or purchasing security apps. We consider this possibility and allow for endogenous monitoring in this section.

The setup is the same as before, except that the consumer can incur costs to obtain an extra signal s' about the data sale after observing the costless signal s. By incurring an effort $h \in [0, \bar{h}]$ ($\bar{h} < 1$), the consumer obtains a signal s_h. If a firm sold data in the previous period, the consumer detects it with probability h. The consumer would receive a signal $s_h = y$ if they caught any sales and $s_h = n$ if they did not detect any sales. We make the following assumption on the cost $c(h)$.

Assumption 2. $c(\cdot) \in \mathcal{C}^2(\mathbb{R}_+), c(0) = 0, c'(h) > 0, c''(h) > 0, \lim_{h \to \bar{h}} c'(h) = +\infty.$

We assume that it is costless not to incur any effort, the marginal cost of the consumer increases at an increasing rate when the precision of monitoring improves, and it is very costly to monitor the data sales very precisely.

Monopoly

Let μ be the consumer's belief after observing signal s. The consumer can incur effort h to gain an additional signal s_h. We consider the MPE in which the rational firm never sells data. Suppose there exists such an equilibrium. There are two cases.

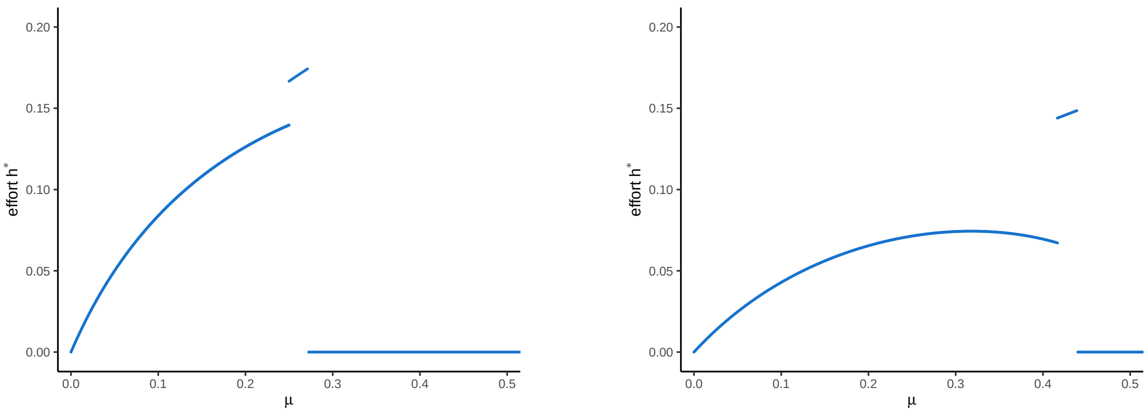
$\mu > \widehat{\mu}$

Without an extra signal, the consumer will reveal nothing according to Proposition 14. If the consumer obtains a costly signal, she must take a different action under some circumstances. Otherwise, she will be better off by not incurring any costs. Therefore, the belief must be below $\widehat{\mu}$ if the consumer incurs effort h and receives a signal $s_h = n$. When the belief μ is too high, the updated belief will be above $\widehat{\mu}$ regardless of the effort. So, the consumer will not incur an effort to get an extra signal. When the belief μ is close to $\widehat{\mu}$, the belief will be below $\widehat{\mu}$ if the consumer incurs enough efforts and receive signal n. The consumer benefits from costly monitoring by being more likely to identify the rational type.

$\mu \leq \widehat{\mu}$

Without an extra signal, the consumer will reveal $1 - \eta/t$ amount of information according to Proposition 14. If the consumer seeks an extra signal and receives $s_h = y$, she knows that the firm is a bad type and does not reveal information. If $s_h = n$, the belief decreases, and the consumer reveals some information. The consumer benefits from costly monitoring by being more likely to identify the bad type.

Examining both cases, we have the following result.



(a) $v = 1, t = 3, u_b = 1, \bar{h} = 0.5, k = 0.1$ (b) $v = 1, t = 3, u_b = 0.6, \bar{h} = 0.5, k = 1$

Figure 3.5: The optimal monitoring effort, where $c(h) = \frac{kh^2}{\bar{h}-h}$.

Proposition 20. *There exists a $\widehat{\delta} < 1$ such that for any $\delta > \widehat{\delta}$, there exists a MPE where the rational firm never sells consmer data. In such equilibrium, there exists a $\widehat{\widehat{\mu}} > \widehat{\mu}$ such that the consumer incurs efforts in monitoring if and only if $\mu \leq \widehat{\widehat{\mu}}$. The monitoring effort strictly increases in μ for $\mu > \widehat{\mu}$. It vanishes as μ approaches zero.*

Figure 3.5 illustrates the monitoring effort as a function of the belief. As we can see, the consumer does not incur any monitoring costs when the belief is far above the threshold belief of revealing information, $\widehat{\mu}$. When the consumer strongly believes that the firm is rational, she also incurs little costs because the likelihood of detecting the data sale is very low. She will reveal the same amount of information without an extra signal. Hence, costly monitoring provides little benefits to her. In contrast, additional monitoring will be valuable for the consumer when the belief is slightly above $\widehat{\mu}$. Since the consumer is quite uncertain about the firm's type, her expected payoff is low. If she reveals nothing and the firm rational, she gives up the opportunity of receiving better product recommendations. If she reveals some information and the firm is bad, she suffers a high privacy loss. By getting another signal, the consumer becomes more certain about the firm's type. A y signal convinces her that the firm will sell her data. So, she reveals nothing. A n signal makes her more confident that the firm will not sell her data. So, she reveals some information. Consequently, the consumer incurs a relatively high effort in this case.

The consumer starts revealing information at a higher belief when they can voluntarily monitor the firm's behavior. So, the rational firm builds up the reputation and achieves the Stackelberg payoff faster under endogenous monitoring. The consumer makes better decisions with an extra signal. Both players are better off.

Multiple Firms

From the monopoly case, we can see that the ability of consumers to gain additional signals makes it easier to build a reputation. In the presence of multiple firms, we have the reputation failure results when the monitoring is exogenous. One natural question is whether endogenous monitoring suffices to reverse the negative results. The following result shows that it is not enough to restore reputation building.

Proposition 21. *For any $\alpha > 0$ and $\delta \in (0,1), \exists N_\delta$ s.t. $\forall N \geq N_\delta$, firms always sell data and consumers reveal nothing in the unique MPE. Consumer's belief about each firm's type is always μ_0.*

Even though endogenous monitoring can help firms build up reputation faster when they do not sell data, it also hurts their reputation more frequently when some firms sell data. With the possibility of a bad type who always sells data, rational firms are tempted to sell data as well because their reputation is affected by other firms. When the number of firms increases, it becomes harder for rational types to commit. Eventually, the firm loses all the commitment power and sells data every period.

3.6 Asymmetric Monitoring

In the main model, the monitoring technology of the consumer is symmetric. The consumer observes whether some firms sold the data without further information about which firm is more likely to sell it. We now consider an asymmetric monitoring technology.

If firm 1 sells the data in the previous period, the consumer detects it with probability q. If firm $i \neq 1$ sells the data in the previous period, the consumer detects it with probability αq $(0 < \alpha < 1)$. The consumer will receive a signal $s = y$ if they caught any sales and $s = n$ if they did not detect any. The consumer has less noise about whether firm 1 sold the data. To get some intuition, notice that $\alpha = 1$ corresponds to the monitoring technology in the main model. If $\alpha = 0$ instead, the consumer knows that firm 1 sold the data in the previous period. The result about the monopoly, Proposition 16, implies that a sufficiently patient firm 1 could commit privacy by reputation, no matter how large the total number of firms is. When α is close to zero, the monitoring technology is close to the monopoly case. The following result shows that any noise from other firms leads to reputation failure.

Proposition 22. *For any $\alpha > 0$ and $\delta \in (0,1), \exists N_\delta$ s.t. $\forall N \geq N_\delta$, firms always sell data and consumers reveal nothing in the unique MPE. Consumer's belief about each firm's type is always μ_0.*

When $\alpha = 0$, the result in the monopoly case implies that rational firms can commit to never selling data. However, any noise from other firms will break down the commitment power. This fragility result shows that the possibility rather than the level of interaction of firms' behavior in the reputation-building process is critical to the reputation failure.

3.7 Conclusion

This paper studies whether reputation consideration can serve as a commitment device for privacy. We show that it depends on the market structure. For a patient enough monopoly, reputation enables it to commit to the Stackelberg action of not selling consumers' data. However, reputation may fail to help the firm commit when there are multiple firms. We characterize some sufficient conditions in which firms cannot commit not to sell the data even if they are very patient. Consumers know the monopoly sold data when they observe it. So, the firm will never restore its reputation by selling data and being caught. The high and permanent reputation cost provides a strong incentive for the monopoly to commit to privacy. In contrast, consumers can never know the specific firm selling the data when there are multiple firms. Therefore, the penalty for data sale is lower, and a firm's reputation may be hurt even if it does not sell data. The minor and temporary reputation cost provides a strong incentive for the firm to deviate.

Reputation failure in the presence of multiple firms persists when we consider several extensions. Endogenous monitoring helps a monopoly build up a reputation faster, benefiting both the rational firm and consumers. However, it does not provide enough incentives for multiple firms to commit not to sell data. Also, we consider asymmetric monitoring. The monopoly case implies that rational firms can commit without noise. In contrast, any noise from other firms will break down the commitment power. This fragility result shows that the possibility rather than the level of interaction of firms' behavior in the reputation-building process is critical to the reputation failure.

There are a couple of limitations to the current work. Consumers can reveal an arbitrary amount of information in the product market. However, the firm sometimes restricts the communication space. So, the consumer can only choose from a menu of the amounts of information to disclose. It will be interesting to study the optimal design of the menu and how much advantage the firm could gain by offering such a contract. Also, the consumer's privacy loss from data sales is exogenous in this paper. Engonenizing the privacy cost in a game theoretic model can provide further insights. We leave them for future research.

Bibliography

[1] S. P. Anderson and R. Renault. “Advertising content”. In: *American Economic Review* 96.1 (2006), pp. 93–113.

[2] S. P. Anderson and R. Renault. “Comparative advertising: disclosing horizontal match information”. In: *The RAND Journal of Economics* 40.3 (2009), pp. 558–581.

[3] M. Arbatskaya. “Ordered search”. In: *The RAND Journal of Economics* 38.1 (2007), pp. 119–126.

[4] M. Armstrong, J. Vickers, and J. Zhou. “Prominence and consumer search”. In: *The RAND Journal of Economics* 40.2 (2009), pp. 209–233.

[5] M. Armstrong and J. Zhou. “Paying for prominence”. In: *The Economic Journal* 121.556 (2011), F368–F395.

[6] S. Athey, C. Catalini, and C. Tucker. *The digital privacy paradox: Small money, small costs, small talk*. Tech. rep. w23488. National Bureau of Economic Research, 2017.

[7] R. J. Aumann and M. Maschler. *Repeated games with incomplete information*. MIT press, 1995.

[8] I. Ball. “Dynamic information provision: Rewarding the past and guiding the future”. Available at SSRN 3103127. 2019.

[9] I. Ball and J. A. Espín-Sánchez. “Experimental Persuasion”. 2022.

[10] D. Bergemann, A. Bonatti, and T. Gan. “The economics of social data”. 2020.

[11] D. Bergemann and S. Morris. “Information design: A unified perspective”. In: *Journal of Economic Literature* 57.1 (2019), pp. 44–95.

[12] R. Berman, H. Zhao, and Y. Zhu. “Strategic Recommendation Algorithms: Overselling and Demarketing Information Designs”. Available at SSRN 4301489. 2022.

[13] J. Bizzotto, J. Rüdiger, and A. Vigier. “Dynamic persuasion with outside information”. In: *American Economic Journal: Microeconomics* 13.1 (2021), pp. 179–194.

[14] P. Bordalo, N. Gennaioli, and A. Shleifer. “Salience and consumer choice”. In: *Journal of Political Economy* 121.5 (2013), pp. 803–843.

[15] F. Branco, M. Sun, and J. M. Villas-Boas. “Optimal search for product information”. In: *Management Science* 58.11 (2012), pp. 2037–2056.

[16] F. Branco, M. Sun, and J. M. Villas-Boas. “Too much information? Information provision and search costs”. In: *Marketing Science* 35.4 (2016), pp. 605–618.

[17] Y. K. Che, K. Kim, and K. Mierendorff. “Keeping the listener engaged: a dynamic model of Bayesian persuasion”. In: *arXiv preprint arXiv:2003.07338* (2020).

[18] Y. Chen and G. Iyer. “Research note consumer addressability and customized pricing”. In: *Marketing Science* 21.2 (2002), pp. 197–208.

[19] Y. Chen and S. Yao. “Sequential search with refinement: Model and application with click-stream data”. In: *Management Science* 63.12 (2017), pp. 4345–4365.

[20] R. H. Coase. “Durability and monopoly”. In: *The Journal of Law and Economics* 15.1 (1972), pp. 143–149.

[21] M. G. Crandall, H. Ishii, and P. L. Lions. *User’s guide to viscosity solutions of second order partial differential equations*. Vol. 27. 1. 1992, pp. 1–67.

[22] A. Degan and M. Li. “Persuasion with costly precision”. In: *Economic Theory* 72.3 (2021), pp. 869–908.

[23] A. Dixit. *The art of smooth pasting*. Harwood Academic Publishers: Chur, Switzerland, 1993.

[24] A. Dukes and L. Liu. “Online shopping intermediaries: The strategic design of search environments”. In: *Management Science* 62.4 (2016), pp. 1064–1077.

[25] E. Durbin and G. Iyer. “Corruptible advice”. In: *American Economic Journal: Microeconomics* 1.2 (2009), pp. 220–42.

[26] J. C. Ely. “Beeps”. In: *American Economic Review* 107.1 (2017), pp. 31–53.

[27] J. C. Ely and M. Szydlowski. “Moving the goalposts”. In: *Journal of Political Economy* 128.2 (2020), pp. 468–506.

[28] J. C. Ely and J. Välimäki. “Bad reputation”. In: *The Quarterly Journal of Economics* 118.3 (2003), pp. 785–814.

[29] D. Fudenberg and D. K. Levine. “Reputation and equilibrium selection in games with a patient player”. In: *Econometrica* 57.4 (1989), pp. 759–778.

[30] P. Gardete and M. Hunter. “Guiding Consumers through Lemons and Peaches: An Analysis of the Effects of Search Design Activities”. In: (2020).

[31] M. Gentzkow and E. Kamenica. “Costly persuasion”. In: *American Economic Review* 104.5 (2014), pp. 457–62.

[32] S. Goldberg, G. Johnson, and S. Shriver. “Regulating privacy online: the early impact of the GDPR on European web traffic & e-commerce outcomes”. In: (2019). Available at SSRN 3421731.

[33] A. Goldfarb and C. Tucker. “Shifts in privacy concerns”. In: *American Economic Review* 102.3 (2012), pp. 349–53.

[34] A. Goldfarb and C. E. Tucker. “Privacy regulation and online advertising”. In: *Management science* 57.1 (2011), pp. 57–71.

[35] S. J. Grossman. “The informational role of warranties and private disclosure about product quality”. In: *The Journal of Law and Economics* 24.3 (1981), pp. 461–483.

[36] L. Guo. “Endogenous Evaluation and Sequential Search”. In: *Marketing Science* 40.3 (2021), pp. 413–427.

[37] H. Hong and M. Shum. “Using price distributions to estimate search costs”. In: *The RAND Journal of Economics* 37.2 (2006), pp. 257–275.

[38] E. Honka. “Quantifying search and switching costs in the US auto insurance industry”. In: *The RAND Journal of Economics* 45.4 (2014), pp. 847–884.

[39] E. Honka and P. Chintagunta. “Simultaneous or sequential? search strategies in the us auto insurance industry”. In: *Marketing Science* 36.1 (2017), pp. 21–42.

[40] S. Ichihashi. “Online privacy and information disclosure by consumers”. In: *American Economic Review* 110.2 (2020), pp. 569–95.

[41] G. Iyer and Z. Z. Zhong. “Pushing Notifications as Dynamic Information Design”. In: *Marketing Science* (2021).

[42] K. Jerath and Q. Ren. “Consumer Attention to Favorable and Unfavorable Product Information, and Firm Information Design”. In: *Journal of Marketing Research* (2021).

[43] K. Jerath and Q. Ren. “Consumer Search and Product Returns”. In: (2022).

[44] G. A. Johnson, S. K. Shriver, and S. Du. “Consumer privacy choice in online advertising: Who opts out and at what cost to industry?” In: *Marketing Science* 39.1 (2020), pp. 33–51.

[45] E. Kamenica and M. Gentzkow. “Bayesian persuasion”. In: *American Economic Review* 101.6 (2011), pp. 2590–2615.

[46] T. T. Ke, S. Lin, and M. Y. Lu. “Information design of online platforms”. In: (2022). Available at SSRN.

[47] T. T. Ke, Z. J. M. Shen, and J. M. Villas-Boas. “Search for information on multiple products”. In: *Management Science* 62.12 (2016), pp. 3576–3603.

[48] T. T. Ke and K. Sudhir. “Privacy rights and data security: Gdpr and personal data driven markets”. In: (2020). Available at SSRN 3643979.

[49] T. T. Ke and J. M. Villas-Boas. “Optimal learning before choice”. In: *Journal of Economic Theory* 180 (2019), pp. 383–437.

[50] J. B. Kim, P. Albuquerque, and B. J. Bronnenberg. “Online demand under limited consumer search”. In: *Marketing Science* 29.6 (2010), pp. 1001–1023.

[51] J. B. Kim, P. Albuquerque, and B. J. Bronnenberg. “The probit choice model under sequential search with an application to online retailing”. In: *Management Science* 63.11 (2017), pp. 3911–3929.

[52] B. Kőszegi and A. Szeidl. “A model of focusing in economic choice”. In: *The Quarterly Journal of Economics* 128.1 (2013), pp. 53–104.

[53] D. M. Kreps and R. Wilson. “Reputation and imperfect information”. In: *Journal of Economic Theory* 27.2 (1982), pp. 253–279.

[54] D. M. Kreps et al. “Rational cooperation in the finitely repeated prisoners’ dilemma”. In: *Journal of Economic Theory* 27.2 (1982), pp. 245–252.

[55] D. Kuksov and Y. Lin. “Signaling low margin through assortment”. In: *Management Science* 63.4 (2017), pp. 1166–1183.

[56] D. Kuksov and J. M. Villas-Boas. “When more alternatives lead to less choice”. In: *Marketing Science* 29.3 (2010), pp. 507–524.

[57] D. Kuksov and M. Zia. “Benefits of customer loyalty in markets with endogenous search costs”. In: *Management Science* 67.4 (2021), pp. 2171–2190.

[58] J. Levin. “The dynamics of collective reputation”. In: *The BE Journal of Theoretical Economics* 9.1 (2009), pp. 1–25.

[59] T. R. Lewis and D. E. Sappington. “Supplying information to facilitate price discrimination”. In: *International Economic Review* (1994), pp. 309–327.

[60] T. Lin. “Valuing intrinsic and instrumental preferences for privacy”. In: *Available at SSRN 3406412* (2019).

[61] L. Liu and A. Dukes. “Consideration set formation with multiproduct firms: The case of within-firm and across-firm evaluation costs”. In: *Management Science* 59.8 (2013), pp. 1871–1886.

[62] L. Ma. “Only the interested learn: A model of proactive learning of product reviews”. In: *Working Paper* (2016).

[63] D. Mayzlin and J. Shin. “Uninformative advertising as an invitation to search”. In: *Marketing Science* 30.4 (2011), pp. 666–685.

[64] P. Milgrom and J. Roberts. “Predation, reputation, and entry deterrence”. In: *Journal of Economic Theory* 27.2 (1982), pp. 280–312.

[65] P. R. Milgrom. “Good news and bad news: Representation theorems and applications”. In: *The Bell Journal of Economics* (1981), pp. 380–391.

[66] J. L. Moraga-González, Z. Sándor, and M. R. Wildenbeest. “Consumer search and prices in the automobile market”. In: *Available at SSRN* (2021).

[67] I. Morozov. “Measuring benefits from new products in markets with information frictions”. In: *Available at SSRN 3446398* (2021).

[68] I. Morozov et al. “Estimation of preference heterogeneity in markets with costly search”. In: *Marketing Science* 40.5 (2021), pp. 871–899.

[69] S. Morris. “Political correctness”. In: *Journal of Political Economy* 109.2 (2001), pp. 231–265.

[70] G. Moscarini and L. Smith. “The optimal level of experimentation”. In: *Econometrica* 69.6 (2001), pp. 1629–1644.

[71] P. Nelson. “Advertising as information”. In: *Journal of Political Economy* 82.4 (1974), pp. 729–754.

[72] D. Orlov, A. Skrzypacz, and P. Zryumov. “Persuading the principal to wait”. In: *Journal of Political Economy* 128.7 (2020), pp. 2542–2578.

[73] C. Phelan. “Public trust and government betrayal”. In: *Journal of Economic Theory* 130.1 (2006), pp. 27–43.

[74] J. Renault, E. Solan, and N. Vieille. “Optimal dynamic information provision”. In: *Games and Economic Behavior* 104 (2017), pp. 329–349.

[75] J. Robert and D. O. Stahl. “Informative price advertising in a sequential search model”. In: *Econometrica: Journal of the Econometric Society* (1993), pp. 657–686.

[76] S. Seiler. “The impact of search costs on consumer behavior: A dynamic approach”. In: *Quantitative Marketing and Economics* 11.2 (2013), pp. 155–203.

[77] S. Seiler and F. Pinna. “Estimating search benefits from path-tracking data: measurement and determinants”. In: *Marketing Science* 36.4 (2017), pp. 565–589.

[78] Q. Shen and J. Miguel Villas-Boas. “Behavior-based advertising”. In: *Management Science* 64.5 (2018), pp. 2047–2064.

[79] G. J. Stigler. “The economics of information”. In: *Journal of Political Economy* 69.3 (1961), pp. 213–225.

[80] M. Sun. “Disclosing multiple product attributes”. In: *Journal of Economics & Management Strategy* 20.1 (2011), pp. 195–224.

[81] C. R. Taylor. “Consumer privacy and the market for customer information”. In: *RAND Journal of Economics* (2004), pp. 631–650.

[82] J. Tirole. “A theory of collective reputations (with applications to the persistence of corruption and to firm quality)”. In: *The Review of Economic Studies* 63.1 (1996), pp. 1–22.

[83] R. M. Ursu, Q. Wang, and P. K. Chintagunta. “Search duration”. In: *Marketing Science* 39.5 (2020), pp. 849–871.

[84] J. M. Villas-Boas. “Dynamic competition with customer recognition”. In: *The Rand Journal of Economics* (1999), pp. 604–631.

[85] J. M. Villas-Boas. “Price cycles in markets with customer recognition”. In: *RAND Journal of Economics* (2004), pp. 486–501.

[86] J. M. Villas-Boas. “Product variety and endogenous pricing with evaluation costs”. In: *Management Science* 55.8 (2009), pp. 1338–1346.

[87] J. M. Villas-Boas and Y. Yao. “A dynamic model of optimal retargeting”. In: *Marketing Science* 40.3 (2021), pp. 428–458.

[88] D. Wei. “Persuasion under costly learning”. In: *Journal of Mathematical Economics* 94 (2021), p. 102451.

[89] M. L. Weitzman. “Optimal search for the best alternative”. In: *Econometrica: Journal of the Econometric Society* (1979), pp. 641–654.

[90] B. Wernerfelt. “Selling formats for search goods”. In: *Marketing Science* 13.3 (1994), pp. 298–309.

[91] A. Wolinsky. “True monopolistic competition as a result of imperfect information”. In: *The Quarterly Journal of Economics* 101.3 (1986), pp. 493–511.

[92] L. Xu, J. Chen, and A. Whinston. “Oligopolistic pricing with online search”. In: *Journal of Management information systems* 27.3 (2010), pp. 111–142.

[93] Y. J. Yao. “Multi-attribute Search and Informative Advertising”. In: (2022).

[94] D. Yavorsky, E. Honka, and K. Chen. “Consumer search in the US auto industry: The role of dealership visits”. In: *Quantitative Marketing and Economics* 19.1 (2021), pp. 1–52.

[95] W. Zhong. “Optimal dynamic information acquisition”. In: *Econometrica* 90.4 (2022a), pp. 1537–1582.

[96] Z. Zhong. “Platform search design: The roles of precision and price”. In: *Marketing Science* (2022b).

[97] Y. Zhu and A. Dukes. “Prominent attributes under limited attention”. In: *Marketing Science* 36.5 (2017), pp. 683–698.

Appendix A

Appendix

A.1 Appendix to Chapter 1

The following proposition formalizes the claim that we can work with mean-preserving posterior beliefs rather than the specific signal structure in section 1.2.

Proposition 23. ***(Equivalent Representations of the Signal)*** *The following are equivalent:*

1. *There exists a binary signal $s \in \Delta(\{0,1\})$ such that $\mathbb{P}[s=1|g] > \mathbb{P}[s=1|b]$.*

2. *There exists a binary-valued posterior belief whose expectation is equal to the prior.*

Proof of Proposition 23. $1 \Rightarrow 2$: Given a binary signal such that $\mathbb{P}[s=1|g] > \mathbb{P}[s=1|b]$, law of iterated expectation implies that $\mathbb{E}[\mathbb{P}[g|s]] = \mathbb{E}[\mathbb{E}[\mathbf{1}_{[g]}|s]] = \mathbb{E}[\mathbf{1}_{[g]}] = \mathbb{P}[g]$. So, the expectation of the posterior belief is equal to the prior. Note that $\mathbb{P}[s=0|g] < \mathbb{P}[s=01|b]$. By Bayes' rule, $\mathbb{P}[g|s=1] = \frac{\mathbb{P}[s=1|g]\mathbb{P}[g]}{\mathbb{P}[s=1|g]\mathbb{P}[g]+\mathbb{P}[s=1|b]\mathbb{P}[b]} > \frac{\mathbb{P}[s=1|g]\mathbb{P}[g]}{\mathbb{P}[s=1|g]\mathbb{P}[g]+\mathbb{P}[s=1|g]\mathbb{P}[b]} = \mathbb{P}[g] = \frac{\mathbb{P}[s=0|g]\mathbb{P}[g]}{\mathbb{P}[s=0|g]\mathbb{P}[g]+\mathbb{P}[s=0|g]\mathbb{P}[b]} > \frac{\mathbb{P}[s=0|g]\mathbb{P}[g]}{\mathbb{P}[s=0|g]\mathbb{P}[g]+\mathbb{P}[s=0|b]\mathbb{P}[b]} = \mathbb{P}[g|s=0]$. So, the posterior belief is binary-valued.

$2 \Rightarrow 1$: Given a binary-valued posterior belief whose expectation is equal to the prior, μ_0. Denote the distribution of the belief by $\mu = \begin{cases} \bar{\mu}_1 > \mu_0 & w.p.\ \lambda_1 \\ \underline{\mu}_1 < \mu_0 & w.p.\ 1-\lambda_1 \end{cases}$. We now construct a binary signal $s \in \Delta(\{0,1\})$. Define $\mathbb{P}[s=1|g] = \frac{\bar{\mu}_1\lambda_1}{\mu_0}$ and $\mathbb{P}[s=1|b] = \frac{(1-\bar{\mu}_1)\lambda_1}{(1-\mu_0)}$. One can verify by Bayes' rule that this signal s induces exactly the same posterior belief, using the assumption that $\mu_0 = \lambda_1\bar{\mu}_1 + (1-\lambda_1)\underline{\mu}_1$. We just need to show that $\mathbb{P}[s=1|g] > \mathbb{P}[s=1|b]$, which follows from the fact that $\bar{\mu}_1 > \mu_0$. □

Proof of Proposition 1. We first characterize the optimal one-period strategy (providing an one-shot signal) of the sender. Analagous to section 6, the sender's problem is:

$$\max_{\lambda_0,\bar{\mu_0}} -K(\lambda_0)+p\lambda_0 \tag{P_0}$$

$$\text{s.t. } \lambda_0(\bar{\mu_0}+v_b)\geq c \tag{IR_0'}$$

$$(F_0), \lambda_0\in[0,1], \underline{\mu}_0\in[0,\mu_0)$$

We transform (P_0) into an equivalent program that is easier to analyze.

Lemma 2. *If $\mu_0 < c/v_g$, the sender does not provide information in the second period. If $\mu_0 \geq c/v_g$, (P_0) is equivalent to:*

$$\Pi_1(\mu_0) := \max -K(\lambda_0)+p\lambda_0 \tag{P_0'}$$

$$s.t.\ \lambda_0\in\left[\frac{c}{v_g}, \frac{\mu_0-c}{-v_b}\right]$$

Proof. We first show that any (λ_0,μ_0) satisfying the constraints in (P_0) also satisfy the constraints in (P_0'): $(IR_0') \Rightarrow \lambda_0 \geq \frac{c}{\bar{\mu_0}+v_b} \geq \frac{c}{v_g}$. (IR_0') & $(F_0) \Rightarrow \lambda_0 \leq \frac{\mu_0-c-\underline{\mu}_0}{-v_b-\underline{\mu}_0} \leq \frac{\mu_0-c}{-v_b}$. Thus, $\lambda_0\in\left[\frac{c}{v_g}, \frac{\mu_0-c}{-v_b}\right]$. It is feasible for the sender to provide information in the second period iff $\left[\frac{c}{v_g}, \frac{\mu_0-c}{-v_b}\right]$ is non-empty: $\frac{c}{v_g}\leq\frac{\mu_0-c}{-v_b} \Leftrightarrow \mu_0\geq\frac{c}{v_g}$. So, If $\mu_0<\frac{c}{v_g}$, the sender will not provide information in the second period.

We then show that for any (λ_0,μ_0) satisfying the constraints in (P_0') and $\mu_0\geq\frac{c}{v_g}$, we can find $\bar{\mu_0},\underline{\mu}_0$ such that $(\lambda_0,\mu_0,\bar{\mu_0},\underline{\mu}_0)$ satisfies the constraints in (P_0). The conclusion then follows. Suppose (λ_0,μ_0) satisfies the constraints in (P_0'): $\lambda_0\in\left[\frac{c}{v_g}, \frac{\mu_0-c}{-v_b}\right], \mu_0\geq\frac{c}{v_g}$. Consider $\bar{\mu_0}=\frac{c}{\lambda_0}-v_b$ and $\underline{\mu}_0=\frac{\mu_0-c+v_b\lambda_0}{1-\lambda_0}$. One can verify that $(\lambda_0,\mu_0,\bar{\mu_0},\underline{\mu}_0)$ satisfies (IR_0') & (F_0). So, we just need to show that $-v_b\leq\bar{\mu_0}\leq 1$ and $\underline{\mu}_0\geq 0$. $\bar{\mu_0}=\frac{c}{\lambda_0}-v_b\geq -v_b$. $\lambda_0\geq\frac{c}{v_g}\Rightarrow\bar{\mu_0}=\frac{c}{\lambda_0}-v_b\leq 1$. $\lambda_0\leq\frac{\mu_0-c}{-v_b}\Rightarrow\mu_0\geq c-v_b\lambda_0\Rightarrow\underline{\mu}_0=\frac{\mu_0-c+v_b\lambda_0}{1-\lambda_0}\geq 0$. □

Now consider the transformed program (P_0') when $\mu_0\geq\frac{c}{v_g}$.

1. If $c\geq v_g\lambda_1^{**}$ (*i.e.* $\lambda_1^{**}\leq\frac{c}{v_g}$) and the sender provides information, then $\lambda_0^*=\frac{c}{v_g}$ due to strict concavity of the objective function. One can show that $(\lambda_0,\bar{\mu_0},\underline{\mu}_0)=(\frac{c}{v_g},1,\frac{\mu_0 v_g-c}{v_g-c})$ is the only feasible information structure that satisfies (IR_0') and (F_0). Thus, the sender will provide information with $(\lambda_0,\bar{\mu_0},\underline{\mu}_0)=(\frac{c}{v_g},1,\frac{\mu_0 v_g-c}{v_g-c})$ iff the sender surplus, $-K(\frac{c}{v_g})+p\cdot\frac{c}{v_g}$, is positive (when it is 0, the sender is indifferent between providing information or not). Let $f(\tilde{c})=-K(\frac{\tilde{c}}{v_g})+p\cdot\frac{\tilde{c}}{v_g}$. We have $f(0)=0$, f is strictly concave and obtains the maximum at $\tilde{c}^*=v_g\lambda_1^{**}<c<1$. In addition, $f(v_g)<0$ because $\lim_{\lambda\to 1}K'(\lambda)=+\infty$. Therefore, there exists a unique $\widehat{c}\in(v_g\lambda_1^{**},v_g)$ $s.t.$ $f(c)\begin{cases}\geq 0, & if\ 0\leq c\leq\widehat{c}\\ <0, & if\ c>\widehat{c}\end{cases}$.
Moreover, when the sender provides information, $\mu_0\geq\frac{c}{v_g}\Rightarrow\widehat{c}\leq\mu_0 v_g$. So, the sender

does not provide information if $c > \widehat{c}$ and provides information with $(\lambda_0^*, \bar{\mu}_0^*) = (\frac{c}{v_g}, 1)$ if $c < \widehat{c}$. The receiver surplus is 0.

2. If $c \in [\mu_0 + v_b\lambda_1^{**}, v_g\lambda_1^{**})$ $(i.e.\ \lambda_1^{**} \geq \frac{\mu_0 - c}{-v_b} > \frac{c}{v_g})$ and the sender provides information, then $\lambda_0^* = \frac{\mu_0 - c}{-v_b}$ due to strict concavity of the objective function. One can show that $(\lambda_0, \bar{\mu}_0, \underline{\mu}_0) = (\frac{\mu_0 - c}{-v_b}, \frac{-\mu_0 v_b}{\mu_0 - c}, 0)$ is the only feasible information structure that satisfies (IR_0') and (F_0). Thus, the sender will provide information with $(\lambda_0, \bar{\mu}_0, \underline{\mu}_0) = (\frac{\mu_0 - c}{-v_b}, \frac{-\mu_0 v_b}{\mu_0 - c}, 0)$ iff the sender surplus, $-K(\frac{\mu_0 - c}{-v_b}) + p \cdot \frac{\mu_0 - c}{-v_b}$, is positive. Since $-K(0) + p \cdot 0 = 0$, $\frac{\mu_0 - c}{-v_b} < \lambda_1^{**}$, and the objective function is strictly concave, the sender surplus is always strictly positive. So, the sender will always provide information. The receiver surplus is zero.

3. If $c < \mu_0 + v_b\lambda_1^{**} \wedge v_g\lambda_1^{**}$ $(i.e.\ \lambda_1^{**} \in \left(\frac{c}{v_g}, \frac{\mu_0 - c}{-v_b}\right))$, then the sender can obtain the maximum possible payoff by setting $(\lambda_0, \bar{\mu}_0) = (\lambda_1^{**}, \frac{\mu_0}{\lambda_1^{**}} \wedge 1)$. Let $\underline{\mu}_0 = \begin{cases} 0, \ if\ \mu_0 \leq \lambda_1^{**} \\ \frac{\mu_0 - \lambda_1^{**}}{1 - \lambda_1^{**}}, \ if\ \mu_0 > \lambda_1^{**} \end{cases}$. One can verify that $(\lambda_0, \bar{\mu}_0, \underline{\mu}_0)$ is feasible and satisfies (IR_0') and (F_0). We have shown in the proof of Lemma 1 that the sender surplus is strictly positive. So, the sender will provide information and $(\lambda_0^*, \bar{\mu}_0^*) = (\lambda_1^{**}, \frac{\mu_0}{\lambda_1^{**}} \wedge 1)$. The receiver surplus is $\begin{cases} \mu_0 + v_b\lambda_1^{**} - c, \ if\ \mu_0 \leq \lambda_1^{**} \\ \lambda_1^{**} v_g - c, \ if\ \mu_0 > \lambda_1^{**} \end{cases} > 0.$

There are two types of iterative signals.

(a) The receiver searches regardless of the signal realization in the first period, and takes action G (B) after observing a positive (negative) signal in the second period.

Denote the information structure in the first period by $(\lambda_0, \bar{\mu}_0, \underline{\mu}_0)$. Denote the information structure in the second period by $(\lambda_1^p, \bar{\mu}_1^p, \underline{\mu}_1^p)$ if the receiver observes a positive signal in the first period, and by $(\lambda_1^n, \bar{\mu}_1^n, \underline{\mu}_1^n)$ if the receiver observes a negative signal in the first period. Now consder a one-period strategy $(\lambda_0', \bar{\mu}_0', \underline{\mu}_0') = (\lambda_0\lambda_1^p + (1-\lambda_0)\lambda_1^n, \frac{\lambda_0\lambda_1^p}{\lambda_0\lambda_1^p + (1-\lambda_0)\lambda_1^n}\bar{\mu}_1^p + \frac{(1-\lambda_0)\lambda_1^n}{\lambda_0\lambda_1^p + (1-\lambda_0)\lambda_1^n}\bar{\mu}_1^n, \frac{\mu_0 - \lambda_0\lambda_1^p\bar{\mu}_1^p - (1-\lambda_0)\lambda_1^n\bar{\mu}_1^n}{1 - \lambda_0\lambda_1^p - (1-\lambda_0)\lambda_1^n})$. One can check that the variables are well-defined and the beliefs are feasible. We now check the participation constraint. $\lambda_0'(\bar{\mu}_0' + v_b) = [\lambda_0\lambda_1^p + (1-\lambda_0)\lambda_1^n](\bar{\mu}_0' + v_b) = \lambda_0\lambda_1^p(\bar{\mu}_1^p + v_b) + (1-\lambda_0)\lambda_1^n(\bar{\mu}_1^n + v_b) \geq 2c \geq c$, where the first inequality comes from the first-period participation constraint for the iterative signals.

$$
\begin{aligned}
\Pi_1(\mu_0) \geq & -K(\lambda_0') + p\lambda_0' \\
> & -\lambda_0 K(\lambda_1^p) - (1-\lambda_0)K(\lambda_1^n) + \lambda_0 p\lambda_1^p + (1-\lambda_0)p\lambda_1^n \text{ (convexity of } K) \\
> & -K(\lambda_0) + \lambda_0(-K(\lambda_1^p) + p\lambda_1^p) + (1-\lambda_0)(-K(\lambda_1^n) + p\lambda_1^n) \\
= & \text{ sender's payoff using the iterative signals}
\end{aligned}
$$

(b) The receiver searches (takes action B) if the signal is positive (negative) in the first period, and takes action G (B) if the signal is positive (negative) in the second period.

If $c \geq v_g\lambda_1^{**}$, or $c < v_g\lambda_1^{**}$ and $\mu_1 \leq c - v_b\lambda_1^{**}$, the expected receiver surplus in the second period is 0. The receiver incurs search cost without any immediate benefit in the first period under iterative signals. The expected receiver surplus in the first period is strictly negative if he searches. Therefore, he will not search, and iterative signals are not feasible. Now we consider the case in which $c < v_g\lambda_1^{**}$ and $\mu_1 > c - v_b\lambda_1^{**}$. The sender's problem when she uses such iterative signals is:

$$\Pi_{iter}(\mu_0) := \max -K(\lambda_0) + \lambda_0\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] \qquad (P_{iter})$$

$$\text{s.t. } \lambda_1(\bar{\mu_1} + v_b) \geq \frac{1+\lambda_0}{\lambda_0}c \qquad (IR_{0,iter})$$

$$\lambda_1(\bar{\mu_1} + v_b) \geq c \qquad (IR_{1,iter})$$

$$(F_0), (F_1), \mu_1 = \bar{\mu_0}, \lambda_1 = \lambda_1^{**}$$

Note that $(IR_{0,iter})$ implies (IR_1^{iter}) and $(\lambda_1^*, \bar{\mu}_1^*) = (\lambda_1^{**}, \frac{\mu_1}{\lambda_1^{**}})$ satisfies (F_1). $\forall \mu_1 \geq \lambda_1^{**}$, the optimal second-period strategy is always $(\lambda_1^*, \bar{\mu}_1^*) = (\lambda_1^{**}, 1)$. Therefore, choosing μ_1 above λ_1^{**} does not increase the second-period sender's payoff or relax the first-period constraints. So, we can restrict μ_1 to be less than or equal to λ_1^{**}.

i) $\mu_0 \geq c - v_b\lambda_1^{**}$
$\Pi_1(\mu_0) = -K(\lambda_1^{**}) + p\lambda_1^{**} > \Pi_{iter}(\mu_0)$.

ii) $\mu_0 < c - v_b\lambda_1^{**}$
$\Pi_1(\mu_0) = -K(\frac{\mu_0 - c}{-v_b}) + \frac{(\mu_0 - c)p}{-v_b}$

$$(F_1) \ \& \ (IR_{0,iter}) \Rightarrow \lambda_0 \geq \frac{c}{\mu_1 - c + v_b\lambda_1^{**}} \left(\Rightarrow \mu_1 \geq \frac{1+\lambda_0}{\lambda_0}c - v_b\lambda_1^{**}\right) \qquad (A.1)$$

$$\overset{\mu_1 \leq \lambda_1^{**}}{\geq} \frac{c}{v_g\lambda_1^{**} - c}$$

$$(F_0) \Rightarrow \lambda_0 = \frac{\mu_0 - \underline{\mu}_0}{\mu_1 - \underline{\mu}_0} \leq \frac{\mu_0}{\mu_1} \overset{(A.1)}{\leq} \frac{\mu_0}{\frac{1+\lambda_0}{\lambda_0}c - v_b\lambda_1^{**}} \Rightarrow \lambda_0 \leq \frac{\mu_0 - c}{c - v_b\lambda_1^{**}}$$

A necessary condition for λ_0 to be well-defined is:

$$\frac{c}{v_g\lambda_1^{**} - c} \leq \frac{\mu_0 - c}{c - v_b\lambda_1^{**}} \Leftrightarrow \mu_0 \geq \frac{\lambda_1^{**}c}{v_g\lambda_1^{**} - c} \left(> \frac{c}{v_g}\right)$$

Therefore, it is feasible for the sender to provide a one-period signal whenever it is feasible to provide iterative signals.

Define $\bar{\Pi}_{iter}(\mu_0) := \max_{0 \leq \lambda_0 \leq \frac{\mu_0 - c}{c - v_b\lambda_1^{**}}} -K(\lambda_0) + \lambda_0\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right]$. One can see $\bar{\Pi}_{iter}(\mu_0) \geq \Pi_{iter}(\mu_0)$. Let $\lambda_0^*(\mu_0) := \arg\max_{0 \leq \lambda_0 \leq \frac{\mu_0 - c}{c - v_b\lambda_1^{**}}} -K(\lambda_0) + \lambda_0\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right]$. We want to show:

$$\bar{\Pi}_{iter}(\mu_0) < \Pi_1(\mu_0)$$
$$\Leftrightarrow -K(\lambda_0^*(\mu_0)) + \lambda_0^*(\mu_0)\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] < -K(\frac{\mu_0 - c}{-v_b}) + \frac{\mu_0 - c}{-v_b}p \tag{A.2}$$

Notice that

$$\frac{d}{d\lambda_0}\left\{-K(\lambda_0) + \lambda_0\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right]\right\} = -K'(\lambda_0) - K(\lambda_1^{**}) + p\lambda_1^{**}$$
$$\Rightarrow \lambda_0^*(\mu_0) = \begin{cases} \frac{\mu_0 - c}{c - v_b\lambda_1^{**}} & if\ \mu_0 \le \mu_0^t \\ \frac{\mu_0^t - c}{c - v_b\lambda_1^{**}} & if\ \mu_0 > \mu_0^t \end{cases}$$

, where $\lambda_0^t > 0$ is defined by $-K'(\lambda_0^t) - K(\lambda_1^{**}) + p\lambda_1^{**} = 0$, $\mu_0^t = \lambda_0^t(c - v_b\lambda_1^{**}) + c$.

When $\mu_0 \le \mu_0^t$, we have $\begin{cases} \lambda_0^*(\mu_0) = \frac{\mu_0 - c}{c - v_b\lambda_1^{**}} > \frac{\mu_0 - c}{-v_b} \Rightarrow -K(\lambda_0^*(\mu_0)) < -K(\frac{\mu_0 - c}{-v_b}) \\ \lambda_0^*(\mu_0)p\lambda_1^{**} = \frac{\mu_0 - c}{c - v_b\lambda_1^{**}}p\lambda_1^{**} < \frac{\mu_0 - c}{-v_b}p \\ -\lambda_0^*(\mu_0)K(\lambda_1^{**}) < 0 \end{cases}$

, where the first inequality holds because $-v_b > \mu_1 > c - v_b\lambda_1^{**}$. Hence, (A.2) holds.

When $\mu_0 > \mu_0^t$, $\bar{\Pi}_{iter}(\mu_0) = \bar{\Pi}_{iter}(\mu_0^t) < \Pi_1(\mu_0^t) < \Pi_1(\mu_0)$. So, (A.2) holds.

Thus, $\Pi_1(\mu_0) > \bar{\Pi}_{iter}(\mu_0)$ for any μ_0 such that iterative signals are feasible. □

Proof of Lemma 1. λ_0^{**} and λ_1^{**} are determined by the first order conditions: $-K'(\lambda_1^{**}) + p = 0$ and $-K'(\lambda_0^{**}) + p + K(\lambda_1^{**}) - p\lambda_1^{**} = 0$. $-K(0) + p \cdot 0 = 0\ \&\ -K'(\lambda) + p > 0$ for small $\lambda \Rightarrow -K(\lambda_1^{**}) + p\lambda_1^{**} > 0$. Therefore, $-K'(\lambda_0^{**}) + p + K(\lambda_1^{**}) - p\lambda_1^{**} = 0$ implies that $-K'(\lambda_0^{**}) + p > 0 = -K'(\lambda_1^{**}) + p, \Rightarrow K'(\lambda_0^{**}) < K'(\lambda_1^{**}) \Rightarrow \lambda_0^{**} < \lambda_1^{**}$. $-K(0) + p \cdot 0 + (1 - 0)\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] > 0$ and strict concavity (w.r.t. λ_0) of the objective function imply that $-K(\lambda_0^{**}) + p\lambda_0^{**} + (1 - \lambda_0^{**})\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] > 0$. □

Proof of Proposition 2. It can be proved in the same way as in the proof of Proposition 1 where we derive the optimal one-period strategy of the sender. □

Proof of Proposition 3.
(1) $v_g\lambda_1^{**} \le c < \widehat{c}$

If the sender provides information in both periods, the sender's constrained program is:

$$\max -K(\lambda_0) + p\lambda_0 + (1 - \lambda_0)\left[-K(\frac{c}{v_g}) + \frac{cp}{v_g}\right] \tag{P_{2H}}$$
$$\text{s.t. } (IR_0), (F_0), \mu_1 \ge \frac{c}{v_g}$$

We first transform (P_{2H}) into an equivalent program that is easier to analyze.

Lemma 3. *Suppose $v_g\lambda_1^{**} \leq c < \widehat{c}$. If $\mu_{0,1} \leq \mu_0 < \frac{2v_g-c}{(v_g)^2}c$, the sender provides information in one period. If $\mu_0 \geq \frac{2v_g-c}{(v_g)^2}c$, (P_{2H}) is equivalent to:*

$$\max -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\frac{c}{v_g}) + \frac{cp}{v_g}\right] \tag{P'_{2H}}$$
$$s.t.\ \lambda_0 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0 - (1+v_g)c}{-v_b v_g - c}\right]$$

Proof. The proof of the equivalence between (P_{2H}) and (P'_{2H}) is similar to that of Lemma 2. It is feasible for the sender to provide information at both periods if and only if the domain of λ_1 is non-empty: $\frac{c}{v_g} \leq \frac{v_g\mu_0-(2-p)c}{pv_g-c} \Leftrightarrow \mu_0 \geq \frac{2v_g-c}{(v_g)^2}c$. □

Denote the optimal λ_0 without constraints by $\lambda_{0,H}^{**}$. $\lambda_{0,H}^{**} = \arg\max_{\lambda_0} -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\frac{c}{v_g}) + \frac{cp}{v_g}\right]$. The following lemma summarizes the relative size of $\lambda_{0,H}^{**}$, λ_0^{**}, and $\frac{c}{v_g}$.

Lemma 4. $0 < \lambda_{0,H}^{**} < \lambda_1^{**} \leq \frac{c}{v_g}$.

Proof. $\lambda_1^{**} < \frac{c}{v_g}$ is the assumption. F.O.C $\Rightarrow K'(\lambda_{0,H}^{**}) = p + K(\frac{c}{v_g}) - \frac{cp}{v_g}$. From Lemma 1, $K'(\lambda_1^{**}) = p$. $-K(\frac{c}{v_g}) + \frac{cp}{v_g} > 0$ when $c < \widehat{c}$. So, $K'(\lambda_{0,H}^{**}) < K'(\lambda_1^{**}) \Rightarrow \lambda_{0,H}^{**} < \lambda_1^{**}$. $-K'(\lambda_{0,H}^{**}) + p + K(\frac{c}{v_g}) - \frac{cp}{v_g} = 0 \Rightarrow K'(\lambda_{0,H}^{**}) = p + K(\frac{c}{v_g}) - \frac{cp}{v_g} > p - \frac{cp}{v_g} = (1-\frac{c}{v_g})p > 0$, where the last inequality follows from the assumption that $c < v_g$. Thus, $\lambda_{0,H}^{**} > 0$. □

When it is feasible for the sender to provide information in both periods, $\mu_0 \geq \frac{2v_g-c}{(v_g)^2}c$, Lemma 4 and strict concavity of the objective function imply that the optimal two-period strategy of the sender is $(\lambda_t^*, \bar{\mu_t}^*) = (\frac{c}{v_g}, 1), t = 0, 1$. The sender surplus is $(2-\frac{c}{v_g})\left[-K(\frac{c}{v_g}) + \frac{cp}{v_g}\right] > -K(\frac{c}{v_g}) + \frac{cp}{v_g}$, the sender surplus of the optimal one-period strategy. Therefore, the sender will always provide information in both periods as long as it is feasible.

(2) $c < v_g\lambda_1^{**}$

If the sender provides information in both periods, we first show that we can restrict the domain of μ_1 to be $\leq \lambda_1^{**}$. The intuition is that the optimal second-period strategy is always $(\lambda_1^*, \bar{\mu}_1^*) = (\lambda_1^{**}, 1)$, $\forall \mu_1 \geq \lambda_1^{**}$. Therefore, choosing μ_1 above λ_1^{**} does not increase the second-period sender's payoff or relax the first-period constraints. Formally, when $\lambda_1^{**} \leq \mu_1 < \mu_0$, the sender's constrained program is:

$$\max -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right]$$
$$\text{s.t. } \lambda_0(\bar{\mu_0} + v_b) + (1-\lambda_0)[\lambda_1^{**}v_g - c] \geq c \tag{$\tilde{IR}'_0$}$$
$$(F_0), \mu_1 \in [\lambda_1^{**}, \mu_0)$$

$(\tilde{IR}'_0)$ & $(F_0) \Rightarrow \lambda_0 \leq \frac{-2c+\mu_0-\mu_1+v_g\lambda_1^{**}}{v_g\lambda_1^{**}-c-v_b-\mu_1} \leq \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}$ ("=" when $\mu_1 = \lambda_1^{**}$).

$(F_0) \Rightarrow \lambda_0 \geq \frac{\mu_0-\mu_1}{1-\mu_1}$. The domain of λ_0 is non-empty iff $\frac{\mu_0-\mu_1}{1-\mu_1} \leq \frac{-2c+\mu_0-\mu_1+v_g\lambda_1^{**}}{v_g\lambda_1^{**}-c-v_b-\mu_1} \Leftrightarrow \mu_1 \leq \frac{-2c+v_g\lambda_1^{**}+\mu_0[1-v_g\lambda_1^{**}+c+v_b]}{v_g-c}$. Therefore, smaller μ_1 means it is more likely for the domain of λ_0 to be non-empty and larger upper bound of λ_0. So, the optimal μ_1 will never $\in (\lambda_1^{**}, \mu_0)$. Hence, the optimal strategy in the last period is $(\lambda_1^*, \bar{\mu}_1^*) = \begin{cases} (\lambda_1^{**}, \frac{\mu_1}{\lambda_1^{**}}) & , \ if \ \mu_1 \in (c-v_b\lambda_1^{**}, \lambda_1^{**}] \\ (\frac{\mu_1-c}{-v_b}, \frac{-\mu_1 v_b}{\mu_1-c}) & , \ if \ \mu_1 \in [\frac{c}{v_g}, c-v_b\lambda_1^{**}] \end{cases}$ and the constrained program of the entire game is either[1]:

$$\begin{aligned} &\max -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] \qquad (P_{2S_+}) \\ &\text{s.t. } (IR_0), (F_0), \mu_1 \in [c - v_b\lambda_1^{**}, \lambda_1^{**}] \end{aligned}$$

or:

$$\begin{aligned} &\max -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{-v_b}\right] \qquad (P_{2S_0}) \\ &\text{s.t. } (IR_0), (F_0), \mu_1 \in [\frac{c}{v_g}, c - v_b\lambda_1^{**}] \end{aligned}$$

We consider the two programs above separately, and then compare the corresponding local solutions to pin down the global solution.

1. S_+ strategy (solution to (P_{2S_+}))

Proposition 24. *Suppose $c < v_g\lambda_1^{**}$ and $\mu_0 < \widehat{\mu_0} = 2c - v_b\lambda_1^{**} - [c + (1-\lambda_1^{**})v_b]\lambda_0^{**}$. If $\mu_0 > 2c - v_b\lambda_1^{**}$ and $\mu_0 \geq \frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c}$, (P_{2S_+}) is feasible with the following solution. $\lambda_0^* = \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}$; $\bar{\mu}_0^* = \begin{cases} \frac{(v_b-c)(c-v_b\lambda_1^{**})-v_b\mu_0}{\mu_0-2c+v_b\lambda_1^{**}} \in (-v_b, 1) & , if \ \widehat{\mu_1}(\mu_0) < c - v_b\lambda_1^{**} \\ 1 & , if \ \widehat{\mu_1}(\mu_0) \geq c - v_b\lambda_1^{**} \end{cases}$; $(\lambda_1^*, \bar{\mu}_1^*) = (\lambda_1^{**}, \frac{\mu_1}{\lambda_1^{**}})$; $\mu_1^* = \widehat{\mu_1}(\mu_0) \vee c - v_b\lambda_1^{**}$, where $\widehat{\mu_1}(\mu_0) = \frac{2c-v_b\lambda_1^{**}-(1+c-v_b\lambda_1^{**}+v_b)\mu_0}{c-v_b-\mu_0}$. The receiver gets zero surplus.*

Proof. We first transform (P_{2S_+}) into an equivalent program that is easier to analyze.

Lemma 5. *Suppose $c < v_g\lambda_1^{**}$ and $\mu_0 \geq \mu_{0,1}$. If $\mu_0 \leq 2c - v_b\lambda_1^{**}$ or $\mu_0 < \frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c}$, the sender provides information in one period. If $\mu_0 > 2c - v_b\lambda_1^{**}$ and $\mu_0 \geq \frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c}$, (P_{2S_+}) is equivalent to:*

$$\begin{aligned} &\max -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] \qquad (P''_{2S_+}) \\ &s.t. \ \lambda_0 \in \left(0, \frac{\mu_0 - 2c + v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**}) - c}\right] \end{aligned}$$

[1]We include $\mu_1 = c - v_b\lambda_1^{**}$ in (P_{2S_+}) as well to simplify the exposition.

Proof. We first show that, if $\mu_0 > 2c - v_b\lambda_1^{**}$ and $\mu_0 \geq \frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c}$, (P_{2S_+}) is equivalent to:

$$\begin{aligned}
&\max -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] \qquad (P'_{2S_+})\\
&\text{s.t. } \lambda_0 \in \left[\frac{\mu_0-\mu_1}{1-\mu_1}, \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c} \wedge \frac{\mu_0-\mu_1}{-v_b-\mu_1}\right]\\
&\mu_1 \in [c - v_b\lambda_1^{**} \vee \widehat{\mu_1}(\mu_0), \lambda_1^{**}]\\
&, \text{ where } \widehat{\mu_1}(\mu_0) = \frac{2c - v_b\lambda_1^{**} - (1 + c - v_b\lambda_1^{**} + v_b)\mu_0}{c - v_b - \mu_0}
\end{aligned}$$

$(F_0) \Rightarrow \lambda_0 = \frac{\mu_0-\mu_1}{\bar{\mu_0}-\mu_1} \in \left[\frac{\mu_0-\mu_1}{1-\mu_1}, \frac{\mu_0-\mu_1}{-v_b-\mu_1}\right]$. (IR_0) & $(F_0) \Rightarrow \lambda_0 \leq \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}$. For λ_0 to be positive, we need $\mu_0 > 2c - v_b\lambda_1^{**}$. The domain of λ_0 is non-empty iff $\frac{\mu_0-\mu_1}{1-\mu_1} \leq \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c} \Leftrightarrow \mu_1 \geq \widehat{\mu_1}(\mu_0)$. For $\mu_1 \leq \lambda_1^{**}$, we need $\widehat{\mu_1}(\mu_0) \leq \lambda_1^{**} \Leftrightarrow \mu_0 \geq \frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c}$. We also have that $\mu_1 = \underline{\mu}_0 < \mu_0$. Thus, the constraints in (P_{2S_+}) imply the constraints in (P'_{2S_+}).

For any (λ_0, μ_1) satisfying the constraints in (P'_{2S_+}), consider the following information structure: $(\lambda_0, \mu_1, \bar{\mu_0} = \frac{\mu_0-(1-\lambda_0)\mu_1}{\lambda_0}, \bar{\mu_1} = \frac{\mu_1}{\lambda_1^{**}} \wedge 1, \underline{\mu}_1 = \frac{\mu_1-\lambda_1^{**}\bar{\mu_1}}{1-\lambda_1^{**}})$. (IR_0) & (F_0) are satisfied by construction. $\bar{\mu_0} = \frac{\mu_0-(1-\lambda_0)\mu_1}{\lambda_0} > \frac{\mu_0-(1-\lambda_0)\mu_0}{\lambda_0} = \mu_0$. $\bar{\mu_0} = \frac{\mu_0-(1-\lambda_0)\mu_1}{\lambda_0} \leq \frac{\mu_0-(1-\frac{\mu_0-\mu_1}{1-\mu_1})\mu_1}{\frac{\mu_0-\mu_1}{1-\mu_1}} = 1$. One can verify that $\bar{\mu_1} \in (-v_b, 1]$, $\underline{\mu}_1 \in [0, -v_b)$. Therefore, the $(\lambda_0, \mu_1, \bar{\mu_0}, \bar{\mu_1}, \underline{\mu}_1)$ we constructed satisfies all the constraints in (P_{2S_+}) and is feasible. Therefore, the two programs are equivalent.

We then show that (P'_{2S_+}) is equivalent to (P''_{2S_+}). It is clear that the constraints in (P'_{2S_+}) imply the constraints in (P''_{2S_+}). We now show that for any $\lambda_0 \in \left(0, \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}\right]$, we can find a feasible (λ_0, μ_1) that satisfies the constraints in (P'_{2S_+}). Since $\lambda_0 = \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}$ maximizes the objective function among $\lambda_0 \in \left(0, \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}\right]$ when $\mu_0 < \widehat{\mu_0}$, we only need to verify (by construction) that $\lambda_0 = \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}$ can be obtained.

i) $\widehat{\mu_1}(\mu_0) < c - v_b\lambda_1^{**}$
Consider $\mu_1 = c - v_b\lambda_1^{**}, \lambda_0 = \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}, \bar{\mu_0} = \frac{\mu_0-(1-\lambda_0)\mu_1}{\lambda_0} = \frac{(v_b-c)(c-v_b\lambda_1^{**})-v_b\mu_0}{\mu_0-2c+v_b\lambda_1^{**}}$. By construction, (IR_0) and (F_0) are satisfied; μ_1's constraints are also satisfied. So, we just need to verify that $\bar{\mu_0} \in (p, 1)$. $\bar{\mu_0} < 1 \Leftrightarrow (v_b - c)(c - v_b\lambda_1^{**}) - v_b\mu_0 < \mu_0 - 2c + v_b\lambda_1^{**} \Leftrightarrow \widehat{\mu_1}(\mu_0) < c - v_b\lambda_1^{**}$, which is the assumption. $\bar{\mu_0} > -v_b \Leftrightarrow c < -v_b(1 - \lambda_1^{**})$, which holds because $\mu_0 > 2c - v_b\lambda_1^{**} \Rightarrow c < \frac{1}{2}(\mu_0 + v_b\lambda_1^{**}) \leq \mu_0 + v_b\lambda_1^{**} < -v_b + v_b\lambda_1^{**} = -v_b(1 - \lambda_1^{**})$.

ii) $\widehat{\mu_1}(\mu_0) \geq c - v_b\lambda_1^{**}$
Consider $\lambda_0 = \frac{\mu_0 - 2c + v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}, \bar{\mu_0} = 1, \mu_1 = \frac{\mu_0 - \lambda_0\bar{\mu_0}}{1-\lambda_0} = \widehat{\mu_1}(\mu_0)$. By construction, (IR_0) and (F_0) are satisfied; μ_1's constraints are also satisfied. So, we just need to verify that $\mu_1 = \widehat{\mu_1}(\mu_0) \in [c - v_b\lambda_1^{**}, \lambda_1^{**}]$. $\widehat{\mu_1}(\mu_0) \geq c - v_b\lambda_1^{**}$ is the assumption. $\mu_0 \geq \frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c} \Rightarrow \widehat{\mu_1}(\mu_0) \leq \lambda_1^{**}$.

□

When $\frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c} \geq \lambda_0^{**} (\Leftrightarrow \mu_0 \geq \widehat{\mu_0})$, the optimal λ_0 is λ_0^{**}. When $\frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c} < \lambda_0^{**}$, the optimal λ_0 is $\frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}$ due to strict concavity of the objective function (denote it by $J(\lambda_0)$). Since the one-period optimal sender surplus is $-K(\lambda_1^{**}) + p\lambda_1^{**} = J(0)$, $J(\cdot)$ is strictly concave and obtains the unique maximum value at $\lambda_0^{**} > \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}$, we have $J(\frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}) > J(0)$. So, the sender always provides information in both periods when it is feasible ($\mu_0 > 2c - v_b\lambda_1^{**}$ and $\mu_0 \geq \frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c}$). We will use this observation in the later proofs. According to the proof of Lemma 5, the receiver always gets zero surpluses when $\mu_0 < \widehat{\mu_0}$. $\mu_1 = \widehat{\mu_1}(\mu_0) \vee c - v_b\lambda_1^{**}$ is the smallest μ_1 that supports $\lambda_0^* = \frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}$, which gives the receiver the largest surplus in the first period. So, $\mu_1^* = \widehat{\mu_1}(\mu_0) \vee c - v_b\lambda_1^{**}$.

$$(F_0) \Rightarrow \bar{\mu}_0^* = \frac{\mu_0 - (1-\lambda_0)\mu_1}{\lambda_0} = \begin{cases} \frac{(v_b-c)(c-v_b\lambda_1^{**})-v_b\mu_0}{\mu_0-2c+v_b\lambda_1^{**}} \in (p,1) & , if\ \widehat{\mu_1}(\mu_0) < c - v_b\lambda_1^{**} \\ 1 & , if\ \widehat{\mu_1}(\mu_0) \geq c - v_b\lambda_1^{**} \end{cases}$$ □

2. S_0 strategy (solution to (P_{2S_0}))

Proposition 25. *Suppose $c \leq v_g\lambda_1^{**}$. When $\mu_0 \geq \frac{2v_g-c}{(v_g)^2}c$, (P'_{2S_0}) is feasible. λ_0^*, λ_1^*, and μ_1^* are continuous and increase in μ_0, while $\bar{\mu}_0^*$ and $\bar{\mu}_1^*$ are continuous and decrease in μ_0, in the solution to (P'_{2S_0}). The receiver gets zero surplus at each period.*

Proof. We first transform (P_{2S_0}) into an equivalent program that is easier to analyze.

Lemma 6. *Suppose $c < v_g\lambda_1^{**}$. If $\mu_{0,1} \leq \mu_0 < \frac{2v_g-c}{(v_g)^2}c$, the sender provides information in one period. If $\mu_0 \geq \frac{2v_g-c}{(v_g)^2}c$, (P_{2S_0}) is equivalent to:*

$$\max -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{-v_b}\right] \qquad (P'_{2S_0})$$
$$s.t.\ \lambda_0 \in \left[\frac{\mu_0-\mu_1}{1-\mu_1}, \frac{\mu_0-\mu_1-c}{-v_b-\mu_1}\right]$$
$$\mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0 - c}{v_g - c}\right]$$

Proof. Using the same argument as the proof of Lemma 5, one can show that (P_{2S_0}) is equivalent to:

$$\max -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\frac{\mu_1 - c}{-v_b}) + \frac{(\mu_1 - c)p}{-v_b}\right] \qquad (P''_{2S_0})$$
$$\text{s.t. } \lambda_0 \in \left[\frac{\mu_0-\mu_1}{1-\mu_1}, \frac{\mu_0-\mu_1-c}{-v_b-\mu_1}\right]$$
$$\mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0 - c}{v_g - c} \wedge c - v_b\lambda_1^{**}\right]$$

We just need to show that (P''_{2S_0}) is equivalent to (P'_{2S_0}). If $\frac{v_g\mu_0-c}{v_g-c} \leq c - v_b\lambda_1^{**}$, μ_1's constraint becomes $\mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0-c}{v_g-c}\right]$. So, (P''_{2S_0}) is equivalent to (P'_{2S_0}).

If $\frac{v_g\mu_0-c}{v_g-c} > c - v_b\lambda_1^{**}$, denote the solution to (P'_{2S_0}) by (λ_0, μ_1).

a) $(\lambda_0^{**}, \lambda_1^{**})$ can be obtained $(\lambda_1 = \lambda_1^{**} \Leftrightarrow \mu_1 = c - v_b\lambda_1^{**})$ in (P''_{2S_0})
$\mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0-c}{v_g-c} \wedge c - v_b\lambda_1^{**}\right]$ is equivalent to $\mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0-c}{v_g-c}\right]$, as the optimal μ_1 under the latter (relaxed) constraint will be $c - v_b\lambda_1^{**}$. So, (P''_{2S_0}) is equivalent to (P'_{2S_0}).

b) $(\lambda_0^{**}, \lambda_1^{**})$ can not be obtained in (P''_{2S_0})
Suppose $\mu_1 > c - v_b\lambda_1^{**}$. If $\lambda_0 > \frac{\mu_0-\mu_1}{1-\mu_1}$, consider $(\lambda'_0 = \lambda_0, \mu'_1 = \mu_1 - \varepsilon)$. For small enough ε, it is feasible and gives the sender a strictly higher payoff. A contradiction! If $\lambda_0 = \frac{\mu_0-\mu_1}{1-\mu_1}$ instead, we have $\frac{\mu_0-\mu_1}{1-\mu_1} \geq \lambda_0^{**}$. A contradiction!
Therefore, $\mu_1 \leq c - v_b\lambda_1^{**}$ and thus (P''_{2S_0}) is equivalent to (P'_{2S_0}).

In sum, (P''_{2S_0}) is equivalent to (P'_{2S_0}). □

Lemma 7. *Suppose $c < v_g\lambda_1^{**}$. For $\mu_0 < \widehat{\mu_0}, \lambda_0$ is binding at the upper bound in the solution to (P'_{2S_0}).*

Proof. To solve (P'_{2S_0}), we consider several cases.

i) $\lambda_0 \leq \frac{\mu_0-\mu_1-c}{-v_b-\mu_1}$ is binding and μ_1's constraints are not binding.
The Lagrangian is:
$\mathcal{L} = -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{-v_b}\right] + \eta\left(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1} - \lambda_0\right)$
$s.t.\ \eta \geq 0, \eta\left(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1} - \lambda_0\right) = 0.$
$$\text{F.O.C.} \Rightarrow \begin{cases} -K'(\lambda_0) + p + K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{v_b} - \eta = 0 \\ (1-\lambda_0)\left[K'(\frac{\mu_1-c}{-v_b})\cdot\frac{1}{v_b} - \frac{p}{v_b}\right] + \eta\cdot\frac{\mu_0+v_b-c}{(v_b+\mu_1)^2} = 0 \end{cases}$$

Plug in $\lambda_0 = \frac{\mu_0-\mu_1-c}{-v_b-\mu_1}$. Dividing the second equality by $\frac{\mu_0+v_b-c}{(v_b+\mu_1)^2}$ and comparing with the first equality, we obtain:

$$\begin{aligned}\eta &= -(v_b+\mu_1)\left[K'(\frac{\mu_1-c}{-v_b})\cdot\frac{1}{v_b}-\frac{p}{v_b}\right]\\ &= -K'(\lambda_0)+p+K(\frac{\mu_1-c}{-v_b})+\frac{(\mu_1-c)p}{v_b}\\ \Rightarrow\ & K(\frac{\mu_1-c}{-v_b})+\frac{v_b+\mu_1}{v_b}K'(\frac{\mu_1-c}{-v_b})-K'(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1})-\frac{cp}{v_b}=0 \qquad (*)\end{aligned}$$

$\frac{\partial}{\partial\mu_1}\left[K(\frac{\mu_1-c}{-v_b})+\frac{v_b+\mu_1}{v_b}K'(\frac{\mu_1-c}{-v_b})\right] = -\frac{v_b+\mu_1}{(v_b)^2}K''(\frac{\mu_1-c}{-v_b}) > 0$. So, the sum of the first two terms of the LHS of $(*)$ strictly increases in μ_1. $\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}$ strictly decreases in μ_1, $K'(\cdot)$ strictly increases in μ_1. So, $-K'(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1})$ strictly increases in μ_1. Thus, the LHS of $(*)$ strictly increases in μ_1. When μ_0 increases, the LHS of $(*)$ is strictly negative if μ_1 is unchanged. Therefore, μ_1 also has to increase. So, the sum of the first two terms of the LHS of $(*)$ increases. As a result, the third term, $-K'(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}) = -K'(\lambda_0)$ has to decrease strictly. So, λ_0 has to increase strictly. In sum, the optimal λ_0 and μ_1 are strictly increasing in μ_0.

ii) $\mu_1 \le \frac{v_g\mu_0-c}{v_g-c}$ is binding.

When $\mu_1 = \frac{v_g\mu_0-c}{v_g-c}$, $\lambda_0 \in \{\frac{c}{v_g}\}$. So, λ_0 is binding at the upper bound.

iii) $\mu_1 \ge \frac{c}{v_g}$ is binding and λ_0 is not binding.

The Lagrangian is $\mathcal{L} = -K(\lambda_0)+p\lambda_0+(1-\lambda_0)\left[-K(\frac{\mu_1-c}{-v_b})+\frac{(\mu_1-c)p}{-v_b}\right]+\eta\left(\mu_1-\frac{c}{v_g}\right)$ $s.t.\ \eta\ge 0, \eta\left(\mu_1-\frac{c}{v_g}\right)=0$.

F.O.C. $\Rightarrow \begin{cases}-K'(\lambda_0)+p+K(\frac{\mu_1-c}{-v_b})+\frac{(\mu_1-c)p}{v_b}=0\\ (1-\lambda_0)\left[K'(\frac{\mu_1-c}{-v_b})\cdot\frac{1}{v_b}-\frac{p}{v_b}\right]+\eta=0\end{cases}$

The second equality $\Rightarrow \eta = -\frac{1-\lambda_0}{v_b}\left[K'(\frac{\mu_1-c}{-v_b})-p\right] \overset{c<v_g\lambda_1^{**}}{<} -\frac{1-\lambda_0}{v_b}\left[K'(\lambda_1^{**})-p\right]=0$. But $\eta\ge 0$. A contradiction! So, this case cannot happen.

iv) $\mu_1 \ge \frac{c}{v_g}$ is binding and $\lambda_0 \ge \frac{\mu_0-\mu_1}{1-\mu_1}$ is binding.

The Lagrangian is $\mathcal{L} = -K(\lambda_0)+p\lambda_0+(1-\lambda_0)\left[-K(\frac{\mu_1-c}{-v_b})+\frac{(\mu_1-c)p}{-v_b}\right]+\eta\left(\mu_1-\frac{c}{v_g}\right)+\xi\left(\lambda_0-\frac{\mu_0-\mu_1}{1-\mu_1}\right)$ $s.t.\ \eta\ge 0, \eta\left(\mu_1-\frac{c}{v_g}\right)=0, \xi\ge 0, \xi\left(\lambda_0-\frac{\mu_0-\mu_1}{1-\mu_1}\right)=0$.

F.O.C. $\Rightarrow \begin{cases}-K'(\lambda_0)+p+K(\frac{\mu_1-c}{-v_b})+\frac{(\mu_1-c)p}{v_b}+\xi=0\\ (1-\lambda_0)\left[K'(\frac{\mu_1-c}{-v_b})\cdot\frac{1}{v_b}-\frac{p}{v_b}\right]+\eta+\xi\frac{1-\mu_0}{(1-\mu_1)^2}=0\end{cases}$

Similar to the previous case, the LHS of the second equality >0. A contradiction! So, this case cannot happen.

v) $\lambda_0 \geq \frac{\mu_0-\mu_1}{1-\mu_1}$ is binding and μ_1 is not binding.
The Lagrangian is:
$\mathcal{L} = -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)\left[-K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{-v_b}\right] + \xi\left(\lambda_0 - \frac{\mu_0-\mu_1}{1-\mu_1}\right)$
$s.t.\ \xi \geq 0, \xi\left(\lambda_0 - \frac{\mu_0-\mu_1}{1-\mu_1}\right) = 0.$
F.O.C. $\Rightarrow \begin{cases} -K'(\lambda_0) + p + K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{-v_b} + \xi = 0 \\ (1-\lambda_0)\left[K'(\frac{\mu_1-c}{-v_b}) \cdot \frac{1}{v_b} - \frac{p}{v_b}\right] + \xi\frac{1-\mu_0}{(1-\mu_1)^2} = 0 \end{cases}$
Similar to the previous case, the LHS of the second equality > 0. A contradiction! So, this case cannot happen.

vi) both λ_0 and μ_1 are not binding.
The solution is the unconstrained optimal solution $(\lambda_0^{**}, \lambda_1^{**})$. But we have assumed that it is not feasible.

i) to vi) finish the proof of Lemma 7. □

According to Lemma 7, if $\mu_0 \geq \frac{2v_g-c}{(v_g)^2}c$, (P'_{2S_0}) is equivalent to:

$$\max -K(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}) + p\,\frac{\mu_0-\mu_1-c}{-v_b-\mu_1} + (1-\frac{\mu_0-\mu_1-c}{-v_b-\mu_1})\left[-K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{-v_b}\right] \quad (P'''_{2S_0})$$
$$\text{s.t. } \mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0-c}{v_g-c}\right]$$

The first order derivative of the objective function w.r.t. μ_1 is:

$$\begin{aligned} &D(\mu_0,\mu_1) \\ :=&\frac{\partial}{\partial\mu_1}\left\{-K(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}) + p\,\frac{\mu_0-\mu_1-c}{-v_b-\mu_1} + (1-\frac{\mu_0-\mu_1-c}{-v_b-\mu_1})\left[-K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{-v_b}\right]\right\} \\ =&\frac{\mu_0+v_b-c}{(\mu_1+v_b)^2}\left[K(\frac{\mu_1-c}{-v_b}) + \frac{v_b+\mu_1}{v_b}K'(\frac{\mu_1-c}{-v_b}) - K'(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}) - \frac{cp}{v_b}\right] \end{aligned}$$

The first term of $D(\mu_0,\mu_1)$, $\frac{\mu_0+v_b-c}{(\mu_1+v_b)^2}$, is always strictly negative. The second term, $K(\frac{\mu_1-c}{-v_b}) + \frac{v_b+\mu_1}{v_b}K'(\frac{\mu_1-c}{-v_b}) - K'(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}) - \frac{cp}{v_b}$, is the LHS of $(*)$, which has been shown to be strictly increasing in μ_1 in the proof of Lemma 7. One can see that $D(\mu_0,\mu_1)$ is strictly negative when μ_1 is large. Thus, $D(\mu_0,\mu_1)$ is always negative or positive for μ_1 small and negative for μ_1 large. Let $\mu_1^{**}(\mu_0)$ be the cutoff value such that $D(\mu_0,\mu_1) \geq 0$ for $\mu_1 \leq \mu_1^{**}(\mu_0)$ and $D(\mu_0,\mu_1) \leq 0$ for $\mu_1 \geq \mu_1^{**}(\mu_0)$ ($\mu_1^{**}(\mu_0) := -\infty$ if $D(\mu_0,\mu_1)$ is always negative). Since $\mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0-c}{v_g-c}\right]$, the optimal $\mu_1^*(\mu_0) = \left[\frac{c}{v_g} \vee \mu_1^{**}(\mu_0)\right] \wedge \frac{v_g\mu_0-c}{v_g-c}$. One can see that we can define $\tilde{\mu}_1^{**}(\mu_0) := \begin{cases} \frac{c}{v_g}, \ if\ D(\mu_0,\mu_1) \text{ is always negative} \\ \mu_1^{**}(\mu_0), \text{ otherwise} \end{cases}$. $\tilde{\mu}_1^{**}(\mu_0) \in (-\infty,+\infty)$ and $\mu_1^*(\mu_0) = \left[\frac{c}{v_g} \vee \tilde{\mu}_1^{**}(\mu_0)\right] \wedge \frac{v_g\mu_0-c}{v_g-c}$. Since $\tilde{\mu}_1^{**}(\mu_0)$ is continuous in μ_0, $\mu_1^*(\mu_0)$ is also continuous in μ_0. It then implies that $\lambda_0^*(\mu_0) = \frac{\mu_0-\mu_1^*(\mu_0)-c}{-v_b-\mu_1^*(\mu_0)}$, $\lambda_1^*(\mu_0) = \frac{\mu_1^*(\mu_0)-c}{-v_b}$, and $\bar{\mu}_1^*(\mu_0) = \frac{-\mu_1^*(\mu_0)v_b}{\mu_1^*(\mu_0)-c}$ are continuous in μ_0.

We have shown in the proof of Lemma 7 that λ_0^* and μ_1^* strictly increase in μ_0 when μ_1^* is the interior solution. Now we consider the case when μ_1 is binding. When $\mu_1^* = \frac{c}{v_g}$, $\lambda_0^* = \frac{\mu_0-\mu_1^*-c}{-v_b-\mu_1^*} = \frac{\mu_0-\frac{c}{v_g}-c}{-v_b-\frac{c}{v_g}}$ strictly increases in μ_0. When $\mu_1^* = \frac{v_g\mu_0-c}{v_g-c}$, it is strictly increasing in μ_0 and $\lambda_0^* = \frac{\mu_0-\mu_1^*-c}{-v_b-\mu_1^*} = \frac{c}{v_g}$ is constant. Together with the continuity property we just established, we have shown that λ_0^* and μ_1^* (weakly) increase in μ_0. Thus, $\lambda_1^* = \frac{\mu_1^*-c}{-v_b}$ (weakly) increases in μ_0 and $\bar{\mu}_1^* = \frac{-\mu_1^*v_b}{\mu_1^*-c}$ (weakly) decreases in μ_0. □

According to the proof of Proposition 1 and Lemma 6, the sender does not provide information iff $\mu_0 < \mu_{0,1}$ and provide information in one period if $\mu_{0,1} \le \mu_0 < \frac{2v_g-c}{(v_g)^2}c$. Thus, we just need to determine whether she provides information in one period or in both periods when $\mu_0 \ge \frac{2v_g-c}{(v_g)^2}c$ by comparing the sender surplus of the optimal one-period strategy and the optimal sender surplus of the S_0 strategy.

a) $c \le v_g\lambda_0^{**}$

Define $\mu_{1,2} := \inf\{\mu_0 \ge \frac{2v_g-c}{(v_g)^2}c : \Pi_{S_0}(\mu_0) \ge \Pi_1(\mu_0)\}$. One can see that $\mu_{1,2} \in [\frac{2v_g-c}{(v_g)^2}c, \widehat{\mu_0})$ and $\Pi_{S_0}(\mu_{1,2}) \ge \Pi_1(\mu_{1,2})$. According to Lemma 7,

$$\begin{aligned}
\Pi_1(\mu_0) =& -K(\frac{\mu_0-c}{-v_b} \wedge \lambda_1^{**}) + p(\frac{\mu_0-c}{-v_b} \wedge \lambda_1^{**}) \\
\Pi_{S_0}(\mu_0) =& \max_{\mu_1} -K(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}) + p\ \frac{\mu_0-\mu_1-c}{-v_b-\mu_1} + \\
& (1-\frac{\mu_0-\mu_1-c}{-v_b-\mu_1})\left[-K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{-v_b}\right] \\
& s.t.\ \mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0-c}{v_g-c}\right]
\end{aligned}$$

i) $\mu_{1,2} \ge c - v_b\lambda_1^{**}$

$\forall \mu_0 \in (\mu_{1,2}, \widehat{\mu_0}], \Pi_{S_0}(\mu_0) > \Pi_{S_0}(\mu_{1,2}) \ge \Pi_1(\mu_{1,2}) = \Pi_1(\mu_0)$.

ii) $\mu_{1,2} < c - v_b\lambda_1^{**}$

$\forall \mu_0 \in [\mu_{1,2}, c - v_b\lambda_1^{**})$, $\frac{d\Pi_1(\mu_0)}{d\mu_0} = K'(\frac{\mu_0-c}{-v_b})\frac{1}{v_b} - \frac{p}{v_b}$.

A. $\mu_1(\mu_0) = \mu_1^u(\mu_0) := \frac{v_g\mu_0-c}{v_g-c}$

In this case, $\lambda_0 = \frac{\mu_0-\mu_1(\mu_0)-c}{-v_b-\mu_1(\mu_0)} = \frac{c}{v_g}$.

So, $\Pi_{S_0}(\mu_0) = -K(\frac{c}{v_g}) + \frac{cp}{v_g} + (1-\frac{c}{v_g})[-K(\frac{\mu_1^u(\mu_0)-c}{-v_b}) + \frac{(\mu_1^u(\mu_0)-c)p}{-v_b}]$.

For $\Delta > 0$ small enough, we have $\mu_0+\delta < c-v_b\lambda_1^{**}$, $\forall \delta \in (0,\Delta)$. Consider $\mu_{0,\delta} = \mu_0 + \delta \in (\mu_0, \mu_0+\Delta)$, we have $\Pi_{S_0}(\mu_{0,\delta}) \ge \underline{\Pi_{S_0}}(\mu_{0,\delta}) := -K(\frac{c}{v_g}) +$

$\frac{cp}{v_g} + (1 - \frac{c}{v_g})[-K(\frac{\mu_1^u(\mu_{0,\delta})-c}{-v_b}) + \frac{(\mu_1^u(\mu_{0,\delta})-c)p}{-v_b}]$. Noticing that $\Pi_{S_0}(\mu_0) = \underline{\Pi_{S_0}}(\mu_0)$, we have

$$\frac{d\Pi_{S_0}(\mu_0)}{d\mu_0} \geq \frac{d\underline{\Pi_{S_0}}(\mu_0)}{d\mu_0} = K'(\frac{\mu_1^u(\mu_0)-c}{-v_b})\frac{1}{v_b} - \frac{p}{v_b} > \frac{d\Pi_1(\mu_0)}{d\mu_0}$$

So, $\Pi_{S_0}(\mu_0) \geq \Pi_1(\mu_0), \forall \mu_0 \in [\mu_{1,2}, c - v_b\lambda_1^{**})$, and the inequality is strict when $\mu_0 > \mu_{1,2}$.

B. $\mu_1(\mu_0) < \mu_1^u(\mu_0)$
Let $\lambda_0 = \frac{\mu_0 - \mu_1(\mu_0) - c}{-v_b - \mu_1(\mu_0)}$. For $\Delta > 0$ small enough, we have $\mu_0 + \delta < c - v_b\lambda_1^{**}$ and $\mu_1(\mu_0) + \frac{\delta}{1-\lambda_0} < \mu_1^u(\mu_0) < \mu_1^u(\mu_0 + \delta)$, $\forall \delta \in (0, \Delta)$. Consider $\mu_{0,\delta} = \mu_0 + \delta \in (\mu_0, \mu_0 + \Delta)$. Since $\frac{\mu_0 + \delta - (\mu_1^u(\mu_0) + \frac{\delta}{1-\lambda_0}) - c}{-v_b - (\mu_1^u(\mu_0) + \frac{\delta}{1-\lambda_0})} = \lambda_0$, we have $\Pi_{S_0}(\mu_{0,\delta}) \geq \widetilde{\underline{\Pi_{S_0}}}(\mu_{0,\delta}) := -K(\lambda_0) + p\lambda_0 + (1-\lambda_0)[-K(\frac{\mu_1(\mu_0) + \frac{\delta}{1-\lambda_0} - c}{p}) + \frac{(\mu_1(\mu_0) + \frac{\delta}{1-\lambda_0} - c)p}{-v_b}]$. Noticing that $\Pi_{S_0}(\mu_0) = \widetilde{\underline{\Pi_{S_0}}}(\mu_0)$, we have

$$\begin{aligned}\frac{d\Pi_{S_0}(\mu_0)}{d\mu_0} \geq \frac{d\widetilde{\underline{\Pi_{S_0}}}(\mu_0)}{d\mu_0} =& (1-\lambda_0)\left[-\frac{1}{p}K'(\frac{\mu_1(\mu_0)-c}{p})\frac{1}{1-\lambda_0} + \frac{1}{1-\lambda_0}\right] \\ =& 1 - \frac{1}{p}K'(\frac{\mu_1(\mu_0)-c}{p}) \\ \geq& 1 - \frac{1}{p}K'(\frac{\mu_1^u(\mu_0)-c}{p}) \\ >& \frac{d\Pi_1(\mu_0)}{d\mu_0}\end{aligned}$$

So, $\Pi_{S_0}(\mu_0) \geq \Pi_1(\mu_0)$ and the inequality is strict when $\mu_0 > \mu_{1,2}$.

In sum, $\Pi_{S_0}(\mu_0) > \Pi_1(\mu_0)$, $\forall \mu_0 \in (\mu_{1,2}, \widehat{\mu_0}]$.

b) $v_g\lambda_0^{**} < c < v_g\lambda_1^{**}$
The following lemma provides a closed-form solution to program (P_{2S_0}) when the search cost is intermediate:

Lemma 8. *Suppose $v_g\lambda_0^{**} < c < v_g\lambda_1^{**}$. $\lambda_0^* = \frac{c}{v_g}, \mu_1^* = \frac{v_g\mu_0 - c}{v_g - c}$ in the solution to (P_{2S_0}).*

Proof. According to the proof of Proposition 25, λ_0^* is binding at the upper bound and increases in μ_0 in the solution to (P'_{2S_0}), for $c < v_g\lambda_1^{**}$. One can see that $\lambda_0^* = \frac{c}{v_g}$ and $\mu_1^* = c - v_b\lambda_1^{**}$ for μ_0 large enough in the solution to (P'_{2S_0}). Because $\lambda_0^* \geq \frac{c}{v_g}$, the only way for λ_0^* to be increasing in μ_0 is for it to always be $\frac{c}{v_g}$. Given $\lambda_0^* = \frac{c}{v_g}$, the optimal $\mu_1^* = \frac{v_g\mu_0 - c}{v_g - c}$ for (P'_{2S_0}). Lemma 6 shows that (P'_{2S_0}) is equivalent to (P_{2S_0}). So, $\lambda_0^* = \frac{c}{v_g}, \mu_1^* = \frac{v_g\mu_0 - c}{v_g - c}$ are also the solutions to (P_{2S_0}). $\square$

Define $\mu_{1,2} := \inf\{\mu_0 \geq \frac{2v_g - c}{v_g^2}c : \Pi_{S_0}(\mu_0) \geq \Pi_1(\mu_0)$ or S_+ strategy is feasible$\}$. Note that $\mu_{1,2} \leq \mu_{2,+}$. If the S_+ strategy is feasible $\forall \mu_0 > \mu_{1,2}$, the 1-period sender surplus will always be dominated by the 2-period sender surplus $\forall \mu_0 > \mu_{1,2}$, as the optimal S_+ strategy generates a strictly higher sender surplus than the optimal 1-period strategy. We now consider the case in which the S_+ strategy is not feasible for some $\mu_0 > \mu_{1,2}$, which implies that $\Pi_{S_0}(\mu_{1,2}) \geq \Pi_1(\mu_{1,2})$.

$$\Pi_1(\mu_0) = -K(\frac{\mu_0 - c}{-v_b} \wedge \lambda_1^{**}) + p(\frac{\mu_0 - c}{-v_b} \wedge \lambda_1^{**})$$
$$\Pi_{S_0}(\mu_0) = -K(\frac{c}{v_g}) + \frac{cp}{v_g} + (1 - \frac{c}{v_g})\left[-K(\frac{\mu_1^u(\mu_0) - c}{-v_b}) + \frac{(\mu_1^u(\mu_0) - c)p}{-v_b}\right]$$

i) $\mu_{1,2} \geq c - v_b\lambda_1^{**}$
$\forall \mu_0 \in (\mu_{1,2}, \mu_{2,+}], \Pi_{S_0}(\mu_0) > \Pi_{S_0}(\mu_{1,2}) \geq \Pi_1(\mu_{1,2}) = \Pi_1(\mu_0)$.

ii) $\mu_{1,2} < c - v_b\lambda_1^{**}$
$\forall \mu_0 \in [\mu_{1,2}, c - v_b\lambda_1^{**}]$,

$$\frac{d\Pi_1(\mu_0)}{d\mu_0} = K'(\frac{\mu_0 - c}{-v_b})\frac{1}{v_b} - \frac{p}{v_b}$$
$$\frac{d\Pi_{S_0}(\mu_0)}{d\mu_0} = K'(\frac{\mu_1^u(\mu_0) - c}{-v_b})\frac{1}{v_b} - \frac{p}{v_b} > \frac{d\Pi_1(\mu_0)}{d\mu_0}$$

So, $\Pi_{S_0}(\mu_0) \geq \Pi_1(\mu_0)$ and the inequality is strict when $\mu_0 > \mu_{1,2}$. $\Pi_{S_0}(c - v_b\lambda_1^{**}) > \Pi_1(c - v_b\lambda_1^{**})$. $\forall \mu_0 \in [c - v_b\lambda_1^{**}, \mu_{2,+}], \Pi_{S_0}(\mu_0) \geq \Pi_{S_0}(c - v_b\lambda_1^{**}) > \Pi_1(c - v_b\lambda_1^{**}) = \Pi_1(\mu_0)$.

In sum, $\forall \mu_0 \in (\mu_{1,2}, \mu_{2,+}],\ \Pi_{S_0}(\mu_0) > \Pi_1(\mu_0)$.

One can see that the optimal S_+ strategy always generates a strictly higher (and strictly positive) sender surplus than the optimal 1-period strategy. Therefore, the sender always provides information in both periods when the S_+ strategy is feasible.[2] By Lemma 5, the S_+ strategy is feasible iff $\mu_0 > 2c - v_b\lambda_1^{**}$ and $\mu_0 \geq \frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c}$ when $c < v_g\lambda_1^{**}$. Hence, together with the above results on the S_0 strategy, there exists $\mu_{1,2} \in [\frac{2v_g - c}{(v_g)^2}c, 2c - v_b\lambda_1^{**} \vee \frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c}]$ such that the sender does not provide information if $\mu_0 < \mu_{0,1}$, provides information in one period if $\mu_0 \in [\mu_{0,1}, \mu_{1,2})$, and provides information in both periods if $\mu_0 > \mu_{1,2}$. □

Proof of Proposition 4.

(1) High Search Cost ($v_g\lambda_1^{**} \leq c < \widehat{c}$)

It has been shown in the proof of Proposition 3. Since the optimal strategy does not depend on the prior, the sender's payoff does not depend on the prior either.

[2]But the optimal 2-period strategy may be either S_+ or the S_0 strategy.

(2) Low Search Cost ($c \leq \tilde{c} = v_g K'^{-1}\left[\frac{K(\lambda_1^{**})}{\lambda_1^{**}}\right]$)

$$\begin{aligned}
\text{F.O.C. of } (P_b) \Rightarrow & K'(\lambda_0^{**}) = K(\lambda_1^{**}) + p(1-\lambda_1^{**}) \\
\Rightarrow & \lambda_1^{**} K'(\lambda_0^{**}) - K(\lambda_1^{**}) = (1-\lambda_1^{**})\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] > 0 \\
\Rightarrow & K'(\lambda_0^{**}) > \frac{K(\lambda_1^{**})}{\lambda_1^{**}} = K'(\frac{\tilde{c}}{v_g}) \\
\Rightarrow & \lambda_0^{**} > \frac{\tilde{c}}{v_g} \\
\Rightarrow & \tilde{c} < v_g \lambda_0^{**} < v_g \lambda_1^{**}
\end{aligned}$$

We now compare the optimal sender surplus between the solution to (P_{2S_+}) and the solution to (P_{2S_0}), and show that the optimal S_0 strategy is always preferred to the optimal S_+ strategy when both types of strategy are feasible.

Proposition 26. *Suppose $c \leq \tilde{c}$ and $\mu_0 < \widehat{\mu_0}$. The sender uses the S_0 strategy when she provides information in both periods.*

Proof. $\forall \mu_0 < \widehat{\mu_0}$ such that S_0 (S_+) strategy is feasible, denote the optimal sender surplus by $\Pi_{S_0}(\mu_0)$ ($\Pi_{S_+}(\mu_0)$).

i) $\mu_0' := \frac{(v_g-c)(c-v_b\lambda_1^{**})+c}{v_g} \leq \mu_0 < \widehat{\mu_0}$
$\frac{(v_g-c)(c-v_b\lambda_1^{**})+c}{v_g} \leq \mu_0 \Leftrightarrow \widehat{\mu_1}(\mu_0) \leq c - v_b\lambda_1^{**}$. According to Proposition 24, the information structure $(\lambda_{0,S_+}, \mu_{1,S_+}) = (\frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}, c - v_b\lambda_1^{**})$ gives $\Pi_{S_+}(\mu_0)$. $\mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0-c}{v_g-c} \wedge c - v_b\lambda_1^{**}\right]$ in (P''_{2S_0}) and $\frac{(v_g-c)(c-v_b\lambda_1^{**})+c}{v_g} \leq \mu_0 \Leftrightarrow \frac{v_g\mu_0-c}{v_g-c} \geq c - v_b\lambda_1^{**}$. Consider $(\lambda_0, \mu_1) = (\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}, c - v_b\lambda_1^{**})$, which satisfies the costraints in (P''_{2S_0}) and is identical to $(\lambda_{0,S_+}, \mu_{1,S_+})$. So, $\Pi_{S_0}(\mu_0) \geq \Pi_{S_+}(\mu_0)$.

ii) $\mu_0 < \mu_0'$
$\mu_0 < \mu_0' \Leftrightarrow \widehat{\mu_1}(\mu_0) > c - v_b\lambda_1^{**}$. According to Proposition 24, $(\lambda_{0,S_+}, \mu_{1,S_+}) = (\frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}, \widehat{\mu_1}(\mu_0))$ gives $\Pi_{S_+}(\mu_0)$. One can verify that $\frac{v_g\mu_0'-c}{v_g-c} = c - v_b\lambda_1^{**}$ and $\frac{\mu_0'-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c} = \frac{c}{v_g} \leq \lambda_0^{**}$. So, $\mu_0' \leq \widehat{\mu_0}$, which implies that $\lambda_{0,S_+}(\mu_0') \leq \lambda_0^{**}$.
Consider μ_0 such that both S_0 and S_+ strategies are feasible. Let $\mu_1^u(\mu_0) := \frac{v_g\mu_0-c}{v_g-c}$.

$$\Pi_{S_+}(\mu_0) = -K(\lambda_{0,S_+}(\mu_0)) + p\lambda_{0,S_+}(\mu_0) + (1-\lambda_{0,S_+}(\mu_0))[-K(\lambda_1^{**}) + p\lambda_1^{**}]$$

$$\Pi_{S_0}(\mu_0) \geq \underline{\Pi}_{S_0}(\mu_0) := -K(\frac{c}{v_g}) + \frac{cp}{v_g} + (1-\frac{c}{v_g})\left[-K(\frac{\mu_1^u(\mu_0)-c}{-v_b}) + \frac{(\mu_1^u(\mu_0)-c)p}{-v_b}\right]$$

Lemma 9. $\frac{\mu_1^u(\mu_0)-c}{-v_b} \geq \lambda_{0,S_+}(\mu_0)$, $\forall \mu_0 < \mu_0'$.

Proof. $\frac{d}{d\mu_0}[\frac{\mu_1^u(\mu_0)-c}{-v_b}] = \frac{v_g}{-v_b(v_g-c)}$, $\frac{d}{d\mu_0}[\lambda_{0,S_+}(\mu_0)] = \frac{1}{-v_b(1-\lambda_1^{**})-c}$.
$\frac{d}{d\mu_0}[\frac{\mu_1^u(\mu_0)-c}{-v_b}] \le \frac{d}{d\mu_0}[\lambda_{0,S_+}(\mu_0)] \Leftrightarrow c(-2v_b-1) \le -v_b v_g \lambda_1^{**}$ $(*)$
If $v_b \ge -1/2$, $(*)$ always holds. If $(-1 <)\ v_b < -1/2$, we have that $\frac{-v_b}{-2v_b-1} \ge 1 \Rightarrow c \le v_g\lambda_1^{**} \le \frac{-v_b v_g}{-2v_b-1}\lambda_1^{**} \Rightarrow (*)$ also holds. So, $\frac{d}{d\mu_0}[\frac{\mu_1^u(\mu_0)-c}{-v_b}] \le \frac{d}{d\mu_0}[\lambda_{0,S_+}(\mu_0)]$, $\forall \mu_0 < \mu_0'$. Note that $\frac{\mu_1^u(\mu_0')-c}{-v_b} = \lambda_1^{**} > \lambda_0^{**} \ge \lambda_{0,S_+}(\mu_0')$. This concludes the proof. □

Now we calculate the increasing rate of the sender surplus as a function of μ_0:

$$\frac{d\Pi_{S_+}(\mu_0)}{d\mu_0} = \frac{K'(\lambda_{0,S_+}(\mu_0)) - K(\lambda_1^{**}) + \frac{cp}{v_b}}{v_b(1-\lambda_1^{**})+c} - \frac{p}{v_b}$$

$$\frac{d\underline{\Pi}_{S_0}(\mu_0)}{d\mu_0} := \frac{K'(\frac{\mu_1^u(\mu_0)-c}{-v_b})}{v_b} - \frac{p}{v_b}$$

$$\frac{d\Pi_{S_+}(\mu_0)}{d\mu_0} \ge \frac{d\underline{\Pi}_{S_0}(\mu_0)}{d\mu_0} \Leftrightarrow \frac{K'(\frac{\mu_1^u(\mu_0)-c}{-v_b})}{-v_b} + \frac{K(\lambda_1^{**}) - \frac{cp}{v_b}}{-v_b(1-\lambda_1^{**})-c} \ge \frac{K'(\lambda_{0,S_+}(\mu_0))}{-v_b(1-\lambda_1^{**})-c} \quad (\star)$$

$$\begin{aligned}
& c \le \tilde{c} = v_g K'^{-1}\left[\frac{K(\lambda_1^{**})}{\lambda_1^{**}}\right] \\
\Leftrightarrow & K(\lambda_1^{**}) \ge \lambda_1^{**} K'(\frac{c}{v_g}) \\
\Rightarrow & -v_b\left[K(\lambda_1^{**}) - \frac{cp}{v_b}\right] \ge (c - v_b\lambda_1^{**})K'(\frac{c}{v_g}) \left(c \le v_g\lambda_1^{**} \Rightarrow K'(\frac{c}{v_g}) \le K'(\lambda_1^{**}) = p\right) \\
\Rightarrow & -v_b\left[K(\lambda_1^{**}) - \frac{cp}{v_b}\right] \ge (c - v_b\lambda_1^{**})K'(\lambda_{0,S+}(\mu_0)) \left(\lambda_{0,S+}(\mu_0) < \lambda_{0,S+}(\mu_0') = \frac{c}{v_g}\right) \\
\Leftrightarrow & \frac{K'(\lambda_{0,S+}(\mu_0))}{-v_b} + \frac{K(\lambda_1^{**}) - \frac{cp}{v_b}}{-v_b(1-\lambda_1^{**})-c} \ge \frac{K'(\lambda_{0,S_+}(\mu_0))}{-v_b(1-\lambda_1^{**})-c} \\
\overset{Lemma 9}{\Rightarrow} & \frac{K'(\frac{\mu_1^u(\mu_0)-c}{-v_b})}{-v_b} + \frac{K(\lambda_1^{**}) - \frac{cp}{v_b}}{-v_b(1-\lambda_1^{**})-c} \ge \frac{K'(\lambda_{0,S_+}(\mu_0))}{-v_b(1-\lambda_1^{**})-c} \\
\overset{(\star)}{\Leftrightarrow} & \frac{d\Pi_{S_+}(\mu_0)}{d\mu_0} \ge \frac{d\underline{\Pi}_{S_0}(\mu_0)}{d\mu_0}
\end{aligned}$$

One can verify that $\lambda_{0,S_+}(\mu_0') = \frac{c}{v_g}$, $\frac{\mu_1^u(\mu_0')-c}{-v_b} = \lambda_1^{**}$. So, $\Pi_{S_+}(\mu_0') = \underline{\Pi}_{S_0}(\mu_0')$. Therefore, $\underline{\Pi}_{S_0} \ge \Pi_{S_+}(\mu_0)$. $\Pi_{S_0}(\mu_0) \ge \underline{\Pi}_{S_0} \Rightarrow \Pi_{S_0}(\mu_0) \ge \Pi_{S_+}(\mu_0)$.

Now we show that the S_0 strategy is feasible whenever the S_+ strategy is feasible, which concludes the proof of Proposition 26.

Lemma 10. *Suppose $c < v_g\lambda_1^{**}$. For any μ_0 such that the S_+ strategy is feasible, the S_0 strategy is also feasible.*

Proof. Suppose there exists μ_0 such that the S_+ strategy is feasible while the S_0 strategy is not feasible. Then, $2c - v_b\lambda_1^{**} < \frac{2v_g - c}{(v_g)^2}$ and $\frac{(2-\lambda_1^{**})c}{v_g(1-\lambda_1^{**})+c} < \frac{2v_g - c}{(v_g)^2}$, which is equivalent to $\lambda_1^{**} < \frac{c^2}{v_b(v_g)^2} + \frac{2c}{v_g}$ and $c > v_g\lambda_1^{**}$, which is not possible as we assumed that $c < v_g\lambda_1^{**}$. □

□

Proposition 26 tells us that we can limit our attention to the S_0 strategy when $c \le \tilde{c}$. Proposition 25 has characterized the optimal S_0 strategy, which implies that the sender's payoff strictly increases in the prior belief.

□

Proposition 27. ***(Comparative Statics W.r.t. the Prior Belief When the Search Cost Is Intermediate)*** *When the search cost is intermediate, $v_g\lambda_0^{**} < c < v_g\lambda_1^{**}$, and the sender provides information in both periods. There exists $\mu_{2,+} \in [\frac{2v_g - c}{(v_g)^2}c, \widehat{\mu_0})$ and $\mu_{2,0} \in [\frac{2v_g - c}{(v_g)^2}c, \mu_{2,+}]$. The probability of a positive signal in the first period, λ_0^*, remains the same when $\mu_0 < \mu_{2,0}$ and strictly increases in the prior when $\mu_0 > \mu_{2,+}$. The probability of a positive signal in the second period, λ_1^*, strictly increases in the prior when $\mu_0 < \mu_{2,0}$ and remains the same when $\mu_0 > \mu_{2,+}$. The belief after observing a positive signal in the second period, $\bar{\mu}_1^*$, strictly decreases in the prior when $\mu_0 < \mu_{2,0}$ or $\mu_0 > \mu_{2,+}$. A positive signal always fully reveals the state in the first period, $\bar{\mu}_0^* \equiv 1$.*

When the expected receiver surplus in the second period is zero (the S_0 strategy), the minimum amount of information for the receiver to search in the first period is already too high. Under the S_+ strategy, the receiver anticipates that the sender will provide favorable information in the second period, which relaxes the first-period participation constraint. Therefore, the receiver is willing to search even if the sender provides less information in the first period. This benefits the sender. However, the S_+ strategy has the disadvantage of inducing higher expected receiver search costs.

When the prior is low, the disadvantage of the S_+ strategy dominates the advantage. The sender prefers the S_0 strategy to the S_+ strategy. She faces information over-provision in the first period and under-provision in the second period. In the first period, she provides the minimum amount of information for the receiver to search. So, the probability of a positive signal in the first period, λ_0^*, does not depend on the prior. In the second period, she provides the maximum amount of information. More frequent positive signals are feasible when the prior is higher. Even if the receiver becomes less certain about the state being good after observing a positive signal, he will still search as long as the likelihood of receiving a positive signal and earning a strictly positive surplus increases. In equilibrium, the sender trades off the precision of a positive signal for frequency as the prior increases.

When the prior is high, the advantage of the S_+ strategy dominates the disadvantage. The sender prefers the S_+ strategy to the S_0 strategy. She faces information under-provision in the first period and no distortion in the second period. In the first period, she provides the maximum amount of feasible information, which strictly increases in the prior. In the second period, she provides the optimal amount of information, which does not depend on the prior. When the prior increases, the participation constraint in the first period is relaxed. To persuade the receiver to search in the first period, the sender can provide information less favorable to the receiver in the second period. She trades off the precision of a positive signal in the second period for the frequency of positive signals in the first period. So, the belief after observing a positive signal, $\bar{\mu}_1^*$, strictly decreases in the prior. The optimal strategy is discontinuous when the sender switches the types of strategy (as illustrated in Figure 1.6).

Proof of Proposition 27. We compare the sender surplus between the solution to (P_{2S_+}) and the solution to (P_{2S_0}). The following result shows that the optimal S_+ strategy generates a strictly higher sender surplus than the optimal S_0 strategy when the prior is high.

Lemma 11. *If $v_g\lambda_0^{**} < c < v_g\lambda_1^{**}$, the receiver gets strictly positive surplus in the second period when μ_0 is close to $\widehat{\mu}_0$.*

Proof. According to Proposition 24, the sender achieves the benchmark payoff $-K(\lambda_0^{**}) + p\lambda_0^{**} + (1-\lambda_0^{**})\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right]$ when $\mu_0 \to \widehat{\mu}_0$ by the S_+ strategy. According to Lemma 8, the sender surplus of the S_0 strategy is $\leq -K(\frac{c}{v_g}) + \frac{cp}{v_g} + (1-\frac{c}{v_g})\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] < -K(\lambda_0^{**}) + p\lambda_0^{**} + (1-\lambda_0^{**})\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right]$ as $\lambda_0^{**} < \frac{c}{v_g} = \lambda_0$. The difference of the benchmark sender surplus and the sender surplus of the S_0 strategy is larger than a strictly positve constant. So, when $\mu_0 \to \widehat{\mu}_0$, the S_+ strategy gives the sender strictly higher payoff. $v_g\lambda_0^{**} < c \Leftrightarrow \widehat{\mu}_1(\widehat{\mu}_0) > c - v_b\lambda_1^{**} \Rightarrow$ the receiver gets strictly positive surplus from the S_+ strategy when $\mu_0 \to \widehat{\mu}_0$. Thus, the receiver gets strictly positive surplus in the second period in equilibrium when $\mu_0 \to \widehat{\mu}_0$. Continuity of the optimal strategy and sender surplus then implies that there exists a neighborhood $[\widehat{\mu}_0 - \delta, \widehat{\mu}_0]$ for some $\delta > 0$ such that the receiver gets strictly positive surplus in the second period in equilibrium when $\mu_0 \in [\widehat{\mu}_0 - \delta, \widehat{\mu}_0]$. □

When $\frac{c}{v_g} \leq \mu_0 < \mu_{1,2}$, the sender provides information in one period. When $\mu_{1,2} \leq \mu_0 < \widehat{\mu}_0$, the sender provides information in both periods. Proposition 24, and Lemma 8 imply the explicit form of the optimal strategy. When $\mu_{1,2} \leq \mu_0 < \mu_{2,+}$, $(\lambda_0^*, \bar{\mu}_0^*) = (\frac{c}{v_g}, 1)$, $(\lambda_1^*, \bar{\mu}_1^*) = (\frac{v_g(\mu_0-c)-c(1-c)}{p(v_g-c)}, \frac{[v_g\mu_0-c]p}{v_g(\mu_0-c)-c(1-c)})$, $\mu_1^* = \frac{v_g\mu_0-c}{v_g-c} < c - v_b\lambda_1^{**}$. The receiver gets zero surplus in each period. When $\mu_{2,+} \leq \mu_0 < \widehat{\mu}_0$, $(\lambda_0^*, \bar{\mu}_0^*) = (\frac{\mu_0-2c+v_b\lambda_1^{**}}{-v_b(1-\lambda_1^{**})-c}, 1)$, $(\lambda_1^*, \bar{\mu}_1^*) = (\lambda_1^{**}, \frac{\mu_1}{\lambda_1^{**}})$, $\mu_1^* = \widehat{\mu}_1(\mu_0) > c - v_b\lambda_1^{**}$, where $\widehat{\mu}_1(\mu_0) = \frac{2c-v_b\lambda_1^{**}-(1+c-v_b\lambda_1^{**}+v_b)\mu_0}{c-v_b-\mu_0}$. The receiver gets strictly positive surplus in the second period and zero total surplus. □

Proof of Propostion 5. When the search cost is high, $v_g\lambda_1^{**} \leq c < \widehat{c}$, the optimal strategy of the sender is $(\lambda_t^*, \bar{\mu}_t{}^*) = (\frac{c}{v_g}, 1)$ according to Proposition 4, which does not depend on η.

When the search cost is low, $c < \tilde{c} < v_g\lambda_0^{**}$. The boundary solution does not depend on η. Consider the interior solution to (P_{2S_0}), we have: $\eta\tilde{K}(\frac{\mu_1-c}{-v_b}) + \frac{p-\mu_1}{p}\eta\tilde{K}'(\frac{\mu_1-c}{-v_b}) - \eta\tilde{K}'(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}) + c = 0$. The LHS strictly increases in μ_1 and strictly decreases in η. So, $\eta\uparrow \Rightarrow \mu_1^* \uparrow \Rightarrow \lambda_1^* = \frac{\mu_1^*-c}{p}\uparrow, \lambda_0^* = \frac{\mu_0-\mu_1^*-c}{p-\mu_1^*}\downarrow$. □

Proposition 28. ***(Efficient Strategy in the Last Period)*** *At the second period, the social planner does not provide information when $\mu_1 < \mu_{0,1}$. When $\mu_1 \geq \mu_{0,1}$, the efficient signal fully reveals the state when a positive signal arrives, $\bar{\mu}_{1,e} = 1$; the probability of a positive signal, $\lambda_{1,e}$, depends on the search cost:*

1. *if $c \geq v_g\lambda_1^{**}$, then there exists a unique $\widehat{\widehat{c}} \in (v_g\lambda_1^{**}, \mu_1 v_g]$ such that the social planner does not provide information if $c > \widehat{\widehat{c}}$ and $\lambda_{1,e} = \frac{c}{v_g} \vee (\tilde{\lambda_1} \wedge \mu_1)$ if $c \leq \widehat{\widehat{c}}$.*
2. *if $c \in [\mu_1 + v_b\lambda_1^{**}, v_g\lambda_1^{**})$, then $\lambda_{1,e} = \mu_1$.*
3. *if $c < \mu_1 + v_b\lambda_1^{**} \wedge v_g\lambda_1^{**}$, then $\lambda_{1,e} = \tilde{\lambda_1} \wedge \mu_1$.*

Proof of Proposition 28. We first introduce a benchmark problem, in which the receiver is forced to participate and the social planner can generate any signal that fully reveals the state when a positive signal arrives. The social planner chooses the information structure to maximize total welfare. We will use the solution throughout the remaining section.

$$\max_{\lambda_1} -K(\lambda_1) + \lambda_1 \tag{E_b}$$

Lemma 12. *The optimal solution to (E_b) exists and is unique. Denote it by $\tilde{\lambda_1}$. The objective function under $\tilde{\lambda_1}$ is strictly positive. $\tilde{\lambda_1}$ does not depend on the search cost c and $\tilde{\lambda_1} > \lambda_1^{**}$, the solution to the payoff-maximizing benchmark problem.*

Proof. All the results follow from the same argument as the proof of Lemma 1 except $\tilde{\lambda_1} > \lambda_1^{**}$. The F.O.C.'s imply $K'(\tilde{\lambda_1}) = 1 > p = K'(\lambda_1^{**}) \Rightarrow \tilde{\lambda_1} > \lambda_1^{**}$. □

As the social planner wants to maximize the total welfare, she always wants to make the precision of the signal, $\bar{\mu_1}$, as high as possible (subject to the feasibility constraint) given μ_1 and λ_1, to increase receiver surplus while holding sender surplus fixed. So, if the social planner provides information, then $\bar{\mu_1} = \frac{\mu_1}{\lambda_1} \wedge 1$.

1. if $c \geq v_g\lambda_1^{**}$ (*i.e.* $\lambda_1^{**} \leq \frac{c}{v_g}$), then $TS = \begin{cases} -K(\lambda_1) + \mu_1 - c, \ if \ \lambda_1 \geq \mu_1 \\ -K(\lambda_1) + \lambda_1 - c, \ if \ \lambda_1 < \mu_1 \end{cases} \Rightarrow \lambda_{1,e} = \frac{c}{v_g} \vee (\tilde{\lambda_1} \wedge \mu_1)$ when the social planner provides information. Since $\mu_1 \geq \mu_{0,1}$, we have $\lambda_{1,e} \leq \mu_1$ and therefore $\bar{\mu}_{1,e} = 1$. The social planner provides information if the total surplus is non-negative. Similar to the proof of Proposition 2, one can show that there exists a unique $\widehat{\widehat{c}} \in (v_g\lambda_1^{**}, \mu_1 v_g]$ such that the social planner does not provide information iff $c > \widehat{\widehat{c}}$. Moreover, if $\widehat{\widehat{c}} \geq v_g\tilde{\lambda_1}$, $\lambda_{1,e} = \frac{c}{v_g} \Rightarrow \bar{\mu}_{1,e} = 1 \Rightarrow TS = -K(\lambda_{1,e}) + p\lambda_{1,e} \Rightarrow \widehat{\widehat{c}} = \widehat{c}$.
2. if $c \in [\mu_1 + v_b\lambda_1^{**}, v_g\lambda_1^{**})$ (*i.e.* $\lambda_1^{**} \geq \frac{\mu_1 - c}{-v_b} > \frac{c}{v_g}$), then by the previous argument, $\lambda_{1,e} = \frac{c}{v_g} \vee (\tilde{\lambda_1} \wedge \mu_1) = \mu_1 \Rightarrow \bar{\mu}_{1,e} = 1$.

3. if $c < \mu_1 + v_b\lambda_1^{**} \wedge v_g\lambda_1^{**}$ (*i.e.* $\lambda_1^{**} \in \left(\frac{c}{v_g}, \frac{\mu_1 - c}{-v_b}\right)$), then the total welfare $TS =$ $\begin{cases} -K(\lambda_1) + \mu_1 - c, \ if \ \lambda_1 \geq \mu_1 \\ -K(\lambda_1) + \lambda_1 - c, \ if \ \lambda_1 < \mu_1 \end{cases} \Rightarrow \lambda_{1,e} = \tilde{\lambda_1} \wedge \mu_1 \Rightarrow \bar{\mu}_{1,e} = 1.$

□

Proof of Proposition 6. When$v_g\tilde{\lambda_1} \leq c \leq \widehat{\widehat{c}}$, the constrained program of the social planner is:

$$\max -K(\lambda_0) + \lambda_0\bar{\mu_0} - c + (1-\lambda_0)\left[-K(\frac{c}{v_g}) + \frac{cp}{v_g}\right] \qquad (E_{2H})$$
$$\text{s.t. } (IR_0), (F_0), \mu_1 \geq \frac{c}{v_g}$$

Using similar methods of finding the optimal sender strategy in the main text, one can show that the solution to (E_{2H}) is $(\lambda_{0,e}, \lambda_{1,e}) = (\frac{c}{v_g}, \frac{c}{v_g})$. Therefore, there is no information distortion.

When $v_g\lambda_1^{**} \leq c < v_g\tilde{\lambda_1} \wedge \widehat{\widehat{c}}$, the constrained program of the social planner is:

$$\max -K(\lambda_0) + \lambda_0\bar{\mu_0} - c + (1-\lambda_0)\left[-K(\tilde{\lambda_1} \wedge \mu_1) + \tilde{\lambda_1} \wedge \mu_1 - c\right] \qquad (E_{2I})$$
$$\text{s.t. } (IR_0), (F_0), \mu_1 \geq \frac{c}{v_g}$$

Using similar methods of finding the optimal sender strategy in the main text, one can show that the solution to (E_{2I}) is $(\lambda_{0,e}, \lambda_{1,e}) = (\frac{c}{v_g}, \frac{c}{v_g})$. Therefore, there is no information distortion.

When $c < v_g\lambda_1^{**}$, one can see that we can restrict μ_1 to be less than or equal to $\tilde{\lambda_1}$ without loss of generality. The constrained program of the social planner is:

$$\max -K(\lambda_0) + \lambda_0\bar{\mu_0} - c + (1-\lambda_0)\left[-K(\mu_1) + \mu_1 - c\right] \qquad (E_{2S_0})$$
$$\text{s.t. } (IR_0), (F_0), \mu_1 \geq \frac{c}{v_g}$$

Using similar methods of finding the optimal sender strategy in the main text, one can show that the solution to (E_{2S_0}) is $(\lambda_{0,e}, \lambda_{1,e}) = (\frac{\mu_0 - \frac{c}{v_g} - c}{p - \frac{c}{v_g}}, \frac{c}{v_g})$. Therefore, the sender provides too much information (relative to the efficient solution) in the second period and too little information in the first period. □

Proof of Proposition 7. One can see that if the expected surplus of the receiver in the second period is 0 in the optimal solution to (P_{dc}), then $(\lambda_1^*, \bar{\mu}_1^*)$ solves (P_1). Otherwise, the sender can strictly increase the payoff by using the same $(\lambda_0^*, \bar{\mu}_0^*)$ and replacing $(\lambda_1^*, \bar{\mu}_1^*)$ by the optimal solution to (P_1), holding the same μ_1. Hence, if dynamic commitment power strictly increases the sender surplus, the solution to (P_{dc}) must satisfy: $\mathbb{E}$[receiver surplus at $t =$

$1] = (1-\lambda_0)[\lambda_1(\bar{\mu_1}+v_b)-c] > 0$. Denote the optimal sender surplus when the sender does not have dynamic commitment power by Π_{wo}. The corresponding optimal strategy of the sender is $(\lambda_t^*, \bar{\mu_t}^*) = (\frac{c}{v_g}, 1)$ according to Proposition 4. Consider the following strategy: $(\lambda_0, \bar{\mu_0}, \lambda_1, \bar{\mu_1}) = (\frac{c-\delta v_g}{(1-\delta)v_g}, 1, \frac{c}{v_g}+\delta, 1)$. Denote the corresponding sender surplus by $\Pi(\delta)$. One can verify that it is feasible when $\delta > 0$ is small and the sender has dynamic commitment power, and leads to a payoff no larger than the optimal sender surplus with dynamic commitment (denote it by Π_w).

$$
\begin{aligned}
\Pi(\delta) =& -K\left[\frac{c-\delta v_g}{(1-\delta)v_g}\right] + p\frac{c-\delta v_g}{(1-\delta)v_g} + \left[1-\frac{c-\delta v_g}{(1-\delta)v_g}\right]\left[-K(\frac{c}{v_g}+\delta) + \frac{cp}{v_g} + \delta p\right] \\
\frac{d\Pi(\delta)}{d\delta} =& \frac{v_g - c}{(1-\delta)^2 v_g}\left[I_1(\delta) + I_2(\delta)\right] \\
\text{, where } I_1(\delta) =& K'\left[\frac{c-\delta v_g}{(1-\delta)v_g}\right] - (1-\delta)K'(\frac{cp}{v_g}+\delta) \\
I_2(\delta) =& \frac{cp}{v_g} - K(\frac{c}{v_g}+\delta)
\end{aligned}
$$

$I_1(\delta) \to 0,\ I_2(\delta) \to \frac{cp}{v_g} - K(\frac{c}{v_g}) > 0$ as $\delta \to 0$. Therefore, $\exists \widehat{\delta} > 0\ s.t.\ \frac{d\Pi(\delta)}{d\delta} > 0,\ \forall \delta \in (0, \widehat{\delta}]$. Since $\Pi(\delta)$ is continuous in δ and $\Pi(0) = \Pi_{wo}$, we have $\Pi_w \geq \Pi(\delta) > \Pi_{wo},\ \forall \delta \in (0, \widehat{\delta}]$.

We now show that the benefit of dynamic commitment power vanishes as the search cost approaches zero in several steps. Denote the optimal sender surplus when the sender has (does not have) dynamic commitment power and the search cost is c by $\Pi_{w,c}$ ($\Pi_{wo,c}$). Note that $\Pi_{w,c} \geq \Pi_{wo,c}, \forall c \geq 0$.

Lemma 13. *When the search cost is zero, the optimal sender surpluses with and without dynamic commitment power are the same,* $\Pi_{w,0} = \Pi_{wo,0}$.

Proof. Suppose the sender provides information in both periods. When the search cost is 0 and the sender uses one-shot signals, one can see that (IR_0) and (IR_1) are equivalent to $\bar{\mu}_0 \geq -v_b$ and $\bar{\mu}_1 \geq -v_b$. Therefore, even if the receiver obtains strictly positive surplus in the second period, the first-period participation constraint is not relaxed. Hence, the second-period strategy of the sender maximizes her second-period payoff in the solution to the program with dynamic commitment, (P_{dc}), and thus satisfies the constraints of the program without dynamic commitment, (P_2). To show that dynamic commitment power does not improve the sender's payoff when $c = 0$, we just need to show that iterative signals are not optimal when the sender has dynamic commitment power.[3] When $c = 0 < v_g\lambda_1^{**}$ and $\mu_1 > c - v_b\lambda_1^{**} = -v_b\lambda_1^{**}$, the optimal strategy with and without dynamic commitment power coincides, as $(\lambda_1^*, \bar{\mu}_1^*) = (\lambda_1^{**}, \frac{\mu_1}{\lambda_1^{**}} \wedge 1)$. In the proof of Proposition 1, we have shown

[3] We have shown in Proposition 1 that iterative signals are not optimal when the sender does not have dynamic commitment power.

that iterative signals are not optimal. So, we just need to show that iterative signals are not optimal when $\mu_1 \leq -v_b\lambda_1^{**}$.

If $\mu_0 \geq -v_b\lambda_1^{**}$, then $\mu_1 = \bar{\mu}_0 > \mu_0 \geq -v_b\lambda_1^{**}$. So, iterative signals are not optimal.

If $\mu_0 < -v_b\lambda_1^{**}$, we have:

$$\begin{aligned}\Pi_1(\mu_0) =& -K(\frac{\mu_0}{-v_b}) + \frac{\mu_0 p}{-v_b}\\ \Pi_{iter}(\mu_0) =& \max_{\lambda_0,\mu_1} -K(\lambda_0) + \lambda_0\left[-K(\frac{\mu_1}{-v_b}) + \frac{\mu_1 p}{-v_b}\right]\\ &\text{s.t. } (IR_{0,iter}), (IR_{1,iter}), (F_0), (F_1), \mu_1 = \bar{\mu_0}\end{aligned}$$

Denote the optimal soution when the sender uses iterative signals by $(\tilde{\lambda_0}, \tilde{\mu_1})$. $\Pi_{iter}(\mu_0) \leq \Pi_1(\mu_0) \Leftrightarrow -K(\tilde{\lambda_0}) + \tilde{\lambda_0}\left[-K(\frac{\tilde{\mu_1}}{-v_b}) + \frac{\tilde{\mu_1}p}{-v_b}\right] \leq -K(\frac{\mu_0}{-v_b}) + \frac{\mu_0 p}{-v_b}$. To show that iterative signals are not optimal, it suffices to show that $-\tilde{\lambda_0}K(\frac{\tilde{\mu_1}}{-v_b}) + \frac{\tilde{\lambda_0}\tilde{\mu_1}p}{-v_b} \leq -K(\frac{\mu_0}{-v_b}) + \frac{\mu_0 p}{-v_b}$. Strict convexity of $K(\cdot) \Rightarrow \tilde{\lambda_0}K(\frac{\tilde{\mu_1}}{-v_b}) = \tilde{\lambda_0}K(\frac{\tilde{\mu_1}}{-v_b}) + (1-\tilde{\lambda_0})K(0) \geq K(\frac{\tilde{\lambda_0}\tilde{\mu_1}}{-v_b})$. Thus, $-\tilde{\lambda_0}K(\frac{\tilde{\mu_1}}{-v_b}) \leq -K(\frac{\tilde{\lambda_0}\tilde{\mu_1}}{-v_b})$. It suffices to show that $-K(\frac{\tilde{\lambda_0}\tilde{\mu_1}}{-v_b}) + \frac{\tilde{\lambda_0}\tilde{\mu_1}p}{-v_b} \leq -K(\frac{\mu_0}{-v_b}) + \frac{\mu_0 p}{-v_b}$, which hold because $(F_0) \Rightarrow \tilde{\lambda_0}\tilde{\mu_1} \leq \mu_0$ and we know from the F.O.C. that $-K(\lambda) + p\lambda$ strictly increases in λ when $\lambda < \lambda_1^{**}$ (here, $\lambda \leq \frac{\mu_0}{-v_b} < \lambda_1^{**}$). □

We now make the following observation. Since $\Pi_{w,c} \geq \Pi_{wo,c}, \forall c \geq 0$ and $\Pi_{w,0} = \Pi_{wo,0}$, we must have $\Pi_{wo,c} \to \Pi_{w,c}$ as $c \to 0$ if $\Pi_{wo,c} \to \Pi_{wo,0}$ as $c \to 0$.[4] The next result confirms that it is indeed the case and thus finishes the proof.

Lemma 14. $\Pi_{wo,c} \to \Pi_{wo,0}$ *as* $c \to 0$.

Proof. Proposition 26 shows that the optimal strategy is the S_0 strategy when the search cost is low. So, according to Lemma 6 and 7, for any c small enough, the sender's problem is:

$$\begin{aligned}\Pi_{wo,c} =& \max_{\mu_1} -K(\frac{\mu_0-\mu_1-c}{-v_b-\mu_1}) + p\,\frac{\mu_0-\mu_1-c}{-v_b-\mu_1} + (1-\frac{\mu_0-\mu_1-c}{-v_b-\mu_1})\left[-K(\frac{\mu_1-c}{-v_b}) + \frac{(\mu_1-c)p}{-v_b}\right] \quad (P'''_{2S_0})\\ &\text{s.t. } \mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0-c}{v_g-c}\right]\end{aligned}$$

Denote the solution when the search cost is 0 by $\mu_{1,0}^*$. Define $\mu_{1,c}$ to be the closest value to $\mu_{1,0}^*$ among $\left[\frac{c}{v_g}, \frac{v_g\mu_0-c}{v_g-c}\right]$ and denote the corresponding sender surplus by $\underline{\Pi}_{wo,c}$. One can see that $\underline{\Pi}_{wo,c} \leq \Pi_{wo,c}$. Since $\mu_{1,0}^* \in \left[0, \frac{v_g\mu_0}{v_g}\right]$, we have $\mu_{1,c} \to \mu_{1,0}^*$ as $c \to 0$. Therefore, $\underline{\Pi}_{wo,c} \to \Pi_{wo,0}$ as $c \to 0$. Since $\underline{\Pi}_{wo,c} \leq \Pi_{wo,c} \leq \Pi_{wo,0}$, we also have $\Pi_{wo,c} \to \Pi_{wo,0}$ as $c \to 0$. □

□

Proof of Proposition 8. When $c \geq \widehat{c}$, the sender does not provide information for any δ.

[4] One can easily show this observation formally by the triangle inequality.

(1) $v_g\lambda_1^{**} \leq c < \widehat{c}$

By the same argument as the proof of Proposition 4, one can see that the optimal strategy of the sender does not depend on δ. For low prior, $\mu_0 \in [\frac{c}{v_g}, \frac{2v_g - c}{(v_g)^2}c)$, the sender provides information in one period, $(\lambda^*, \bar{\mu}^*) = (\frac{c}{v_g}, 1)$. For high prior, $\mu_0 \geq \frac{2v_g-c}{(v_g)^2}c$, the sender provides information in both periods, $(\lambda_t^*, \bar{\mu_t}^*) = (\frac{c}{v_g}, 1), t = 0, 1$.

(2) $c \leq \tilde{c}$

The sender's problem can be divided into 2 cases:

$$\max -K(\lambda_0) + p\lambda_0 + \delta(1-\lambda_0)\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] \qquad (P^{\delta}_{2S_+})$$
$$\text{s.t. } (IR_{0,\delta}), (F_0), \mu_1 \in [c - v_b\lambda_1^{**}, \lambda_1^{**}]$$

$$\max -K(\lambda_0) + p\lambda_0 + \delta(1-\lambda_0)\left[-K(\frac{\mu_1 - c}{-v_b}) + \frac{(\mu_1 - c)p}{-v_b}\right] \qquad (P^{\delta}_{2S_0})$$
$$\text{s.t. } (IR_{0,\delta}), (F_0), \mu_1 \in [\frac{c}{v_g}, c - v_b\lambda_1^{**}]$$

Consider the solution to $(P^{\delta}_{2S_0})$. By the same argument as the proof of Lemma 6, $(P^{\delta}_{2S_0})$ is equivalent to

$$\max -K(\lambda_0) + p\lambda_0 + \delta(1-\lambda_0)\left[-K(\frac{\mu_1 - c}{-v_b}) + \frac{(\mu_1 - c)p}{-v_b}\right] \qquad (P^{\delta'}_{2S_0})$$
$$\text{s.t. } \lambda_0 \in \left[\frac{\mu_0 - \mu_1}{1-\mu_1}, \frac{\mu_0 - \mu_1 - c}{-v_b - \mu_1}\right]$$
$$\mu_1 \in \left[\frac{c}{v_g}, \frac{v_g\mu_0 - c}{v_g - c}\right]$$

By arguments similar to the proof of Lemma 7, $\lambda_0 \leq \frac{\mu_0 - \mu_1 - c}{-v_b - \mu_1}$ is binding for $\mu_0 < \widehat{\mu_0}$. Suppose that μ_1's constraints are not binding.

The Lagrangian is:

$\mathcal{L} = -K(\lambda_0) + p\lambda_0 + \delta(1-\lambda_0)\left[-K(\frac{\mu_1 - c}{-v_b}) + \frac{(\mu_1 - c)p}{-v_b}\right] + \eta\left(\frac{\mu_0 - \mu_1 - c}{-v_b - \mu_1} - \lambda_0\right)$

$s.t.\ \eta \geq 0, \eta\left(\frac{\mu_0 - \mu_1 - c}{-v_b - \mu_1} - \lambda_0\right) = 0.$

F.O.C. $\Rightarrow$

$$\eta = \delta(p - \mu_1) - \delta\frac{p - \mu_1}{p}K'(\frac{\mu_1 - c}{-v_b})$$
$$= -K'(\frac{\mu_0 - \mu_1 - c}{-v_b - \mu_1}) + p + \delta\left[K(\frac{\mu_1 - c}{-v_b}) - \mu_1 + c\right]$$
$$\Rightarrow \delta\left[K(\frac{\mu_1 - c}{-v_b}) + \frac{p - \mu_1}{p}K'(\frac{\mu_1 - c}{-v_b}) - p + c\right] - K'(\frac{\mu_0 - \mu_1 - c}{-v_b - \mu_1}) + p = 0 \qquad (*_\delta)$$

The sum of the first two terms of the LHS of $(*_\delta)$ strictly increases in δ, and the LHS of $(*_\delta)$ strictly increases in μ_1. So, the optimal λ_0 strictly decreases in δ; the optimal μ_1 and λ_1 strictly increase in δ. When one of μ_1's constraints is binding, λ_0 and λ_1 does not depend on δ.

We finish the proof of this case by showing that the S_0 strategy dominates the S_+ strategy. Denote the sender surplus of the optimal S_0 (S_+) strategy when the discount factor is $\delta \in (0,1)$ by $\Pi_{S_0}(\delta)$ ($\Pi_{S_+}(\delta)$) and denote the corresponding optimal S_0 (S_+) strategy at time t by $\lambda_{t,S_0}(\delta)$ ($\lambda_{t,S_+}(\delta)$). We have:

$$
\begin{aligned}
\Pi_{S_0}(\delta) =& -K(\lambda_{0,S_0}(\delta)) + p\lambda_{0,S_0}(\delta) + \delta(1-\lambda_{0,S_0}(\delta))\left[-K(\lambda_{1,S_0}(\delta)) + p\lambda_{1,S_0}(\delta)\right] \\
\geq& -K(\lambda_{0,S_0}(1)) + p\lambda_{0,S_0}(1) + \delta(1-\lambda_{0,S_0}(1))\left[-K(\lambda_{1,S_0}(1)) + p\lambda_{1,S_0}(1)\right] \\
=& -K(\lambda_{0,S_0}(1)) + p\lambda_{0,S_0}(1) + (1-\lambda_{0,S_0}(1))\left[-K(\lambda_{1,S_0}(1)) + p\lambda_{1,S_0}(1)\right] \\
& -(1-\delta)(1-\lambda_{0,S_0}(1))\left[-K(\lambda_{1,S_0}(1)) + p\lambda_{1,S_0}(1)\right] \\
\overset{(\dagger)}{\geq}& -K(\lambda_{0,S_+}(1)) + p\lambda_{0,S_+}(1) + (1-\lambda_{0,S_+}(1))\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] \\
& -(1-\delta)(1-\lambda_{0,S_+}(1))\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] \\
=& -K(\lambda_{0,S_+}(1)) + p\lambda_{0,S_+}(1) + \delta(1-\lambda_{0,S_+}(1))\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] \\
\geq& -K(\lambda_{0,S_+}(\delta)) + p\lambda_{0,S_+}(\delta) + \delta(1-\lambda_{0,S_+}(\delta))\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right] \\
=& \Pi_{S_+}(\delta)
\end{aligned}
$$

, where the inequality (†) holds because Proposition 26 implies that $-K(\lambda_{0,S_0}(1)) + p\lambda_{0,S_0}(1) + (1-\lambda_{0,S_0}(1))\left[-K(\lambda_{1,S_0}(1)) + p\lambda_{1,S_0}(1)\right] \geq -K(\lambda_{0,S_+}(1)) + p\lambda_{0,S_+}(1) + (1-\lambda_{0,S_+}(1))\left[-K(\lambda_1^{**}) + p\lambda_1^{**}\right]$ and we have $\lambda_{0,S_0}(1) \geq \lambda_{0,S_+}(1),\ -K(\lambda_{1,S_0}(1)) + p\lambda_{1,S_0}(1) \leq -K(\lambda_1^{**}) + p\lambda_1^{**}$.

□

Proof of Proposition 9.

(1) High search cost $v_g\lambda_1^{**} \leq c < \widehat{c}$

Consider an arbitrary period t in which the receiver takes action G. Suppose the belief at the beginning of period t is μ_t.[5] One can see that the receiver must take action G after observing a positive signal and take action B or S after observing a negative signal in period t. Denote the sender's (receiver's) continuation value after the receiver observes a negative signal in period t by V_t (W_t). $V_t, W_t \geq 0$ and (weakly) increase in $\underline{\mu}_t$. The

[5] It is possible that there are more than one initial beliefs at the beginning of period t if the receiver searches for information regardless of the signal realization in a previous period. In that case, μ_t is any one of them. We will show that the optimal information structure is one-shot. So, there is actually one initial belief in any period.

sender's problem in period t is:

$$\max_{\lambda_t,\bar{\mu}_t} -K(\lambda_t)+p\lambda_t+(1-\lambda_t)V_t$$
$$s.t.\ \lambda_t(\bar{\mu}_t+v_b)+(1-\lambda_t)W_t \geq c \qquad (IR_t)$$
$$\lambda_t\bar{\mu}_t+(1-\lambda_t)\underline{\mu}_t=\mu_t \qquad (F_t)$$

Denote the optimal λ_t without constraints by $\lambda^{**}_{t,H}$. $\lambda^{**}_{t,H}=\arg\max_{\lambda_t} -K(\lambda_t)+p\lambda_t+(1-\lambda_t)V_t$. The F.O.C. of $-K(\lambda_t)+p\lambda_t+(1-\lambda_t)V_t \Rightarrow K'(\lambda^{**}_{t,H})=p-V_t<p=K'(\lambda^{**}_1)\Rightarrow \lambda^{**}_{t,H}<\lambda^{**}_1\leq c/v_g$. For any information structure in which $\lambda_t>c/v_g$, by reducing λ_t (and potentially increasing $\bar{\mu}_t$ to satisfy the participation constraint), the sender can increase her payoff if V_t is fixed. One can see that $\underline{\mu}_t$ and W_t will (weakly) increase, as long as we keep (IR_t) binding. Hence, V_t will also (weakly) increase. So, the sender's payoff will be even higher. So, under the optimal information structure, λ^*_t must be no greater than c/v_g. Since it holds for any period in which the receiver takes action G, and the receiver surplus from that period is $\lambda^*_t(\bar{\mu}_t+v_b)-c\leq \frac{c}{v_g}(1+v_b)-c=0$. The receiver gets zero surplus in each period, $W_t=0$. Therefore, the receiver takes action G immediately after observing a positive signal in any period he searches for information (otherwise, the expected gain from search is strictly negative and he will not search). This implies that the optimal information structure is k periods of one-shot signals. To satisfy the receiver's particiaption constraints, $(\lambda^*_t,\bar{\mu}^*_t)=(c/v_g,1)$, for $t=0,1,...,k-1$.

The sender's expected payoff of providing k periods of such information is: $\sum_{i=0}^{k-1}(1-\frac{c}{v_g})^i\left[\frac{cp}{v_g}-K(\frac{c}{v_g})\right]$, which increases in k. Thus, the sender will provide as many periods of information as possible. Now we characterize the maximum number of periods.

Denote the initial belief at the beginning of period t by $\mu_t:=\underline{\mu}_{t-1}$. The feasibility costraint (F_t) and $(\lambda^*_t,\bar{\mu}^*_t)=(c/v_g,1)$ imply that $\mu_t=\frac{\mu_{t-1}-\frac{c}{v_g}}{1-\frac{c}{v_g}}$. By induction, one can show that $\mu_k=\frac{\mu_0-1+(1-\frac{c}{v_g})^k}{(1-\frac{c}{v_g})^k}$. For it to be feasible to provide information in k periods, we need $\mu_{k-1}\geq c/v_g \Leftrightarrow k\leq \frac{ln(1-\mu_0)}{ln(1-c/v_g)}$. Hence, the maximum number of periods is $\lfloor\frac{ln(1-\mu_0)}{ln(1-c/v_g)}\rfloor$.

(2) Low search cost

The receiver needs to decide between G and B at the end of the game. We first derive an upper bound on the probability that the receiver decides on G under any feasible information structure. Denote the probability that the receiver takes action G in period t by q_t. Because the belief must be greater than or equal to $-v_b$ if the receiver takes action G, the mean-preserving property of the beliefs implies that $\mu_0\geq\sum_{t=0}^{+\infty}q_t(-v_b)\Rightarrow \sum_{t=0}^{+\infty}q_t\leq -\frac{\mu_0}{v_b}$. The probability that the receiver takes action G eventually is bounded from above by $-\frac{\mu_0}{v_b}$. Thus, the sender's payoff is bounded from above by $-\frac{\mu_0 p}{v_b}$, even if the persuasion cost is zero. Now we show that the sender can achieve that payoff as the search cost vanishes. This means that she can obtain the equilibrium payoff as if the persuasion cost were zero.

Consider the following strategy: Given a search cost c, the sender provides the same one-shot signals for T consecutive periods ($t = 0, 1, ..., T-1$), where $(\lambda_t, \bar{\mu}_t) = (\lambda, \bar{\mu}) = (\sqrt{c}, -v_b + \sqrt{c})$ and $T = \frac{ln\left(1-\frac{\mu_0}{-v_b+\sqrt{c}}\right)}{ln(1-\sqrt{c})}$. In each period, the receiver's expected payoff from searching is $\lambda_t(\bar{\mu}_t + v_b) - c = 0$. So, the receiver will keep searching if he observes a negative signal, except in the last period. By setting $\underline{\mu}_{T-1} = 0$, one can verify that the mean-preserving property of the belief is satisfied, and that the variables are well-defined for c small. The probability that the receiver takes action G eventually is $\sum_{t=0}^{T-1} \lambda(1-\lambda)^t = 1 - (1-\lambda)^T = \frac{\mu_0}{-v_b+\sqrt{c}} \to -\frac{\mu_0}{v_b}$ as $c \to 0$.

The sender only incurs the persuasion cost if the receiver has not received a good signal. The expected total persuasion cost of the sender is bounded from above by the costs of always providing the information in T periods, which is $TK(\sqrt{c})$.

$$\lim_{c\to 0} TK(\sqrt{c}) = \lim_{c\to 0} \frac{ln\left(1-\frac{\mu_0}{-v_b+\sqrt{c}}\right)}{ln(1-\sqrt{c})} K(\sqrt{c}) = ln\left(1+\frac{\mu_0}{v_b}\right) \lim_{c\to 0} \frac{K(\sqrt{c})}{ln(1-\sqrt{c})} \overset{\text{L'Hospital's rule}}{=} 0$$

Hence, the sender's payoff approaches $-\frac{\mu_0 p}{v_b}$ as $c \to 0$. The receiver's expected payoff from searching (net of the search cost) is zero in every period given the above strategy. One can see that the sender's payoff under the optimal strategy is no lower than that payoff, and it is bounded from above by $-\frac{\mu_0 p}{v_b}$. So, it also approaches $-\frac{\mu_0 p}{v_b}$ as $c \to 0$.

Now we show that the sender adds noise to positive signals when the search cost is low. Suppose not. The mean-preserving property of the beliefs implies that $\mu_0 \geq \sum_{t=0}^{+\infty} q_t \cdot 1 \Rightarrow \sum_{t=0}^{+\infty} q_t \leq \mu_0 < -\frac{\mu_0}{v_b}$. Then, the payoff of the sender is bounded from above by $\mu_0 p < -\frac{\mu_0 p}{v_b}$. But, we have shown that the sender's payoff approaches $-\frac{\mu_0 p}{v_b}$ as $c \to 0$. So, it cannot be optimal for small c. Therefore, the sender adds noise to positive signals when the search cost is low.

□

A.2 Appendix to Chapter 2

Proof of Proposition 10. We have derived (D_1) in the main text. It implies immediately that $\underline{\mu}'(\mu) < 0$ for $\mu \in (\mu^*, \mu^{**}]$. For $\mu \in [\mu^{**}, 1]$, by the implicit function theorem, we have:

$$\begin{aligned}
\begin{bmatrix} \bar{\mu}'(\mu) \\ \underline{\mu}'(\mu) \end{bmatrix} &= - \begin{bmatrix} -\phi'(\bar{\mu}(\mu)) & \phi'(\underline{\mu}(\mu)) \\ -\psi'(\bar{\mu}(\mu)) & \psi'(\underline{\mu}(\mu)) \end{bmatrix}^{-1} \cdot \begin{bmatrix} 0 \\ \frac{1}{2\sigma^2 c} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2\sigma^2 c} \frac{1}{\phi'(\bar{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]} \\ \frac{1}{2\sigma^2 c} \frac{1}{\phi'(\underline{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\phi(\underline{\mu}(\mu))-\phi(\bar{\mu}(\mu))}{\phi'(\bar{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]} < 0 \\ \frac{\phi(\underline{\mu}(\mu))-\phi(\bar{\mu}(\mu))}{\phi'(\underline{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]} < 0 \end{bmatrix}
\end{aligned}$$

This gives us the expression for $(\overline{D_2})$ and $(\underline{D_2})$. One can see from the negative sign of the derivative that both $\underline{\mu}(\mu)$ and $\bar{\mu}(\mu)$ strictly decrease in μ.

We now look at the width of the search region.

$$\begin{aligned}
&[\bar{\mu}(\mu) - \underline{\mu}(\mu)]' \\
=&\frac{\phi(\underline{\mu}(\mu)) - \phi(\bar{\mu}(\mu))}{\phi'(\bar{\mu}(\mu))[\bar{\mu}(\mu) - \underline{\mu}(\mu)]} - \frac{\phi(\underline{\mu}(\mu)) - \phi(\bar{\mu}(\mu))}{\phi'(\underline{\mu}(\mu))[\bar{\mu}(\mu) - \underline{\mu}(\mu)]} \\
=&\frac{\phi(\underline{\mu}(\mu)) - \phi(\bar{\mu}(\mu))}{\bar{\mu}(\mu) - \underline{\mu}(\mu)} \left[1/\phi'(\bar{\mu}(\mu)) - 1/\phi'(\underline{\mu}(\mu))\right] \\
=&\frac{\phi(\underline{\mu}(\mu)) - \phi(\bar{\mu}(\mu))}{\bar{\mu}(\mu) - \underline{\mu}(\mu)} \left[\underline{\mu}(\mu)^2(1 - \underline{\mu}(\mu))^2 - \bar{\mu}(\mu)^2(1 - \bar{\mu}(\mu))^2\right]
\end{aligned}$$

One can see that $\frac{\phi(\underline{\mu}(\mu))-\phi(\bar{\mu}(\mu))}{\bar{\mu}(\mu)-\underline{\mu}(\mu)} > 0$. So, $[\bar{\mu}(\mu) - \underline{\mu}(\mu)]' > 0 \Leftrightarrow \underline{\mu}(\mu)^2(1 - \underline{\mu}(\mu))^2 > \bar{\mu}(\mu)^2(1 - \bar{\mu}(\mu))^2 \Leftrightarrow \underline{\mu}(\mu)(1 - \underline{\mu}(\mu)) > \bar{\mu}(\mu)(1 - \bar{\mu}(\mu)) \Leftrightarrow |\underline{\mu}(\mu) - 1/2| < |\bar{\mu}(\mu) - 1/2|$. Thus, the width of the search region, $\bar{\mu}(\mu) - \underline{\mu}(\mu)$, increases in the belief, μ, if and only if the quitting boundary is closer to 1/2 than the purchasing boundary. We know that $\forall \mu \geq \mu^{**},\ p = \mu + \frac{\bar{\mu}(\mu)+\underline{\mu}(\mu)}{2}$ due to the symmetry of the one-dimensional learning problem.[6] Therefore,

[6]More specifically, the sum of the purchasing and quitting thresholds is zero when the price is zero in the one-dimensional optimal search strategy, as shown by Branco et al. (2012). It implies that the price equals to the average of the two boundaries. In our two-dimensional problem, the consumer only searches the more uncertain attribute when $\mu \geq \mu^{**}$. So, it can be translated to a one-dimensional search problem with the price p normalized to $p - \mu$.

$$
\begin{aligned}
&\frac{\bar{\mu}(\mu)+\underline{\mu}(\mu)}{2} = p-\mu \geq 3/2-1 = 1/2\\
\Rightarrow\ &\frac{\bar{\mu}(\mu)+\underline{\mu}(\mu)}{2} \geq 1/2\\
\Leftrightarrow\ &\bar{\mu}(\mu)+\underline{\mu}(\mu) > 1\\
\Leftrightarrow\ &|\underline{\mu}(\mu)-1/2| < |\bar{\mu}(\mu)-1/2|,\ \forall \mu \geq \mu^{**}
\end{aligned}
$$

Thus, the width of search region, $\bar{\mu}(\mu)-\underline{\mu}(\mu)$, always increases in the belief μ. Now suppose that $\underline{\mu}(\mu) \geq 1/2$, then $\forall \mu \in (\mu^*, \mu^{**}]$, we have

$$
\begin{aligned}
\underline{\mu}'(\mu) &\overset{(D_1)}{=} \frac{\phi(\underline{\mu}(\mu))-\phi(\mu)}{\phi'(\underline{\mu}(\mu))[\mu-\underline{\mu}(\mu)]}\\
&= \frac{-\phi'(\xi_1(\mu))[\mu-\underline{\mu}(\mu)]}{\phi'(\underline{\mu}(\mu))[\mu-\underline{\mu}(\mu)]}\ (\xi_1(\mu) \in (\underline{\mu}(\mu), \mu))\\
&= -\frac{\phi'(\xi_1(\mu)}{\phi'(\underline{\mu}(\mu))}\\
&< -1
\end{aligned}
$$

, where the last inequality comes from the fact that the absolute value of $\phi'(x) = -\frac{1}{x^2(1-x)^2}$ is strictly increasing in x for $x \geq 1/2$. Similarly, $\forall \mu \in [\mu^{**}, 1]$, we have

$$
\begin{aligned}
\underline{\mu}'(\mu) &\overset{(\underline{D_2})}{=} \frac{\phi(\underline{\mu}(\mu))-\phi(\bar{\mu}(\mu))}{\phi'(\underline{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]}\\
&= \frac{-\phi'(\xi_2(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]}{\phi'(\underline{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]}\ (\xi_2(\mu) \in (\underline{\mu}(\mu), \bar{\mu}(\mu)))\\
&= -\frac{\phi'(\xi_2(\mu)}{\phi'(\underline{\mu}(\mu))}\\
&< -1\\
\bar{\mu}'(\mu) &\overset{(\overline{D_2})}{=} \frac{\phi(\underline{\mu}(\mu))-\phi(\bar{\mu}(\mu))}{\phi'(\bar{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]}\\
&= \frac{-\phi'(\xi_3(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]}{\phi'(\bar{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]}\ (\xi_3(\mu) \in (\underline{\mu}(\mu), \bar{\mu}(\mu)))\\
&= -\frac{\phi'(\xi_3(\mu))}{\phi'(\bar{\mu}(\mu))}\\
&> -1
\end{aligned}
$$

□

Proof of Theorem 1. By symmetry, we only need to prove the case of $\mu_1 \geq \mu_2$. We first show that the viscosity solution of the HJB equation (A.3) exists and is unique. Since the value function is a viscosity solution of (A.3), the viscosity solution of (A.3) must be the value function by uniqueness. We then just need to verify that the learning strategy we conjectured indeed generates a viscosity solution to (A.3). So, the conjectured strategy is optimal.

Lemma 15. *The viscosity solution of the HJB equation* (A.3) *exists and is unique.*

Proof. Since the consumer can guarantee a payoff of zero by quitting immediately and cannot achieve a payoff higher than $\sup\{\mu_1 + \mu_2 - p\} = 1 + 1 - p \leq 2$, the value function is bounded and thus exists. This implies the existence of the viscosity solution because the value function is a viscosity solution to (A.3).

The proof of the uniqueness uses a modification of a comparison principle in Crandall et al. (1992). Given that it very much resembles the proof of Lemma 1 in Ke and Villas-Boas (2019), we refer the reader to their proof. □

To verify that the learning strategy we conjectured indeed generates a viscosity solution to the HJB equation (A.3):

$$\max\left\{\max_{i=1,2}\left[\frac{\mu_i^2(1-\mu_i)^2}{2\sigma^2}V_{\mu_i\mu_i}(\mu_1,\mu_2) - c\right], \max\left[\mu_1+\mu_2-p, 0\right] - V(\mu_1,\mu_2)\right\} = 0$$

We just need to show that (everything else holds by our construction):

$$\begin{aligned} &\frac{\mu_1^2(1-\mu_1)^2}{2\sigma^2}V_{\mu_1\mu_1}(\mu_1,\mu_2) - c \leq 0 \\ \Leftrightarrow\ &\mu_1^2(1-\mu_1)^2 V_{\mu_1\mu_1}(\mu_1,\mu_2)/2\sigma^2 c \leq 1 \end{aligned} \tag{A.3}$$

if $\mu_1 + \mu_2 > 1$, $\mu_1 \geq \mu_2$, and $\underline{\mu}(\mu_1) < \mu_2 < \bar{\mu}(\mu_1)$.

For $\mu_1 \in (\mu^*, \mu^{**}]$, we have

$$\begin{aligned}
&V_{\mu_1}(\mu_1,\mu_2)/2\sigma^2 c\\
=&\phi'(\underline{\mu}(\mu_1))\underline{\mu}'(\mu_1)[\mu_2-\underline{\mu}(\mu_1)]\\
\overset{(D_1)}{=}&\frac{\phi(\underline{\mu}(\mu_1))-\phi(\mu_1)}{\mu_1-\underline{\mu}(\mu_1)}[\mu_2-\underline{\mu}(\mu_1)]\\
\Rightarrow\ &V_{\mu_1\mu_1}(\mu_1,\mu_2)/2\sigma^2 c\\
=&\phi'(\underline{\mu}(\mu_1))\underline{\mu}'(\mu_1)[\mu_2-\underline{\mu}(\mu_1)]\\
\overset{(D_1)}{=}&\frac{\phi'(\underline{\mu}(\mu_1))\underline{\mu}'(\mu_1)-\phi'(\mu_1)}{\mu_1-\underline{\mu}(\mu_1)}[\mu_2-\underline{\mu}(\mu_1)]+[\phi(\underline{\mu}(\mu_1))-\phi(\mu_1)]\frac{(\mu_2-\mu_1)\underline{\mu}'(\mu_1)+\underline{\mu}(\mu_1)-\mu_2}{[\mu_1-\underline{\mu}(\mu_1)]^2}\\
=&-\frac{\phi'(\mu_1)[\mu_2-\underline{\mu}(\mu_1)]}{\mu_1-\underline{\mu}(\mu_1)}+(\mu_2-\mu_1)\frac{[\phi(\underline{\mu}(\mu_1))-\phi(\mu_1)]^2}{\phi'(\underline{\mu}(\mu_1))[\mu_1-\underline{\mu}(\mu_1)]^3}\\
\Rightarrow\ &\mu_1^2(1-\mu_1)^2V_{\mu_1\mu_1}(\mu_1,\mu_2)/2\sigma^2 c\\
=&\frac{\mu_2-\underline{\mu}(\mu_1)}{\mu_1-\underline{\mu}(\mu_1)}+(\mu_1-\mu_2)\mu_1^2(1-\mu_1)^2\frac{\underline{\mu}(\mu_1)^2[1-\underline{\mu}(\mu_1)]^2}{[\mu_1-\underline{\mu}(\mu_1)]^3}[\phi(\underline{\mu}(\mu_1))-\phi(\mu_1)]^2
\end{aligned}$$

So,

$$\begin{aligned}
&\mu_1^2(1-\mu_1)^2V_{\mu_1\mu_1}(\mu_1,\mu_2)/2\sigma^2 c\le 1\\
\Leftrightarrow&\mu_1^2(1-\mu_1)^2\frac{\underline{\mu}(\mu_1)^2[1-\underline{\mu}(\mu_1)]^2}{[\mu_1-\underline{\mu}(\mu_1)]^2}[\phi(\underline{\mu}(\mu_1))-\phi(\mu_1)]^2\le 1\\
\Leftrightarrow&\mu_1(1-\mu_1)\frac{\underline{\mu}(\mu_1)^2[1-\underline{\mu}(\mu_1)]}{[\mu_1-\underline{\mu}(\mu_1)]}[\phi(\underline{\mu}(\mu_1))-\phi(\mu_1)]\le 1\\
\Leftrightarrow&H(\mu_1):=\mu_1(1-\mu_1)[\phi(\underline{\mu}(\mu_1))-\phi(\mu_1)]-\frac{\mu_1-\underline{\mu}(\mu_1)}{\underline{\mu}(\mu_1)[1-\underline{\mu}(\mu_1)]}\le 0
\end{aligned}$$

Observe that $H(\mu^*)=0$. Ignoring the subscript 1 for notational ease, we have:

$$\begin{aligned}
H'(\mu)=&(1-2\mu)[\phi(\underline{\mu}(\mu))-\phi(\mu)]+\frac{\mu(1-\mu)}{\mu-\underline{\mu}(\mu)}[\phi(\underline{\mu}(\mu))-\phi(\mu)]+\frac{1}{\mu(1-\mu)}\\
&-\frac{1}{\underline{\mu}(\mu)(1-\underline{\mu}(\mu))}+\frac{\phi(\underline{\mu}(\mu))-\phi(\mu)}{\mu-\underline{\mu}(\mu)}[-\mu+2\mu\underline{\mu}(\mu)-\underline{\mu}(\mu)^2]\\
=&[1-3\mu+\underline{\mu}(\mu)][\phi(\underline{\mu}(\mu))-\phi(\mu)]+\frac{1}{\mu(1-\mu)}-\frac{1}{\underline{\mu}(\mu)(1-\underline{\mu}(\mu))}
\end{aligned}$$

Suppose (A.3) does not hold. There would exist $\widehat{\mu}$ such that $H(\widehat{\mu})=0$ and $H'(\widehat{\mu})>0$.

$$\text{(A.3)}\Rightarrow\phi(\underline{\mu}(\widehat{\mu}))-\phi(\widehat{\mu})=\frac{\widehat{\mu}-\underline{\mu}(\widehat{\mu})}{\widehat{\mu}(1-\widehat{\mu})\underline{\mu}(\widehat{\mu})[1-\underline{\mu}(\widehat{\mu})]}$$

Hence, we get an expression for $\frac{1}{\widehat{\mu}(1-\widehat{\mu})}$ and $\frac{1}{\underline{\mu}(\mu)[1-\underline{\mu}(\mu)]}$. Plugging these expressions into the previous expression for $H'(\mu)$, we have:

$$H'(\widehat{\mu}) = -2[\phi(\underline{\mu}(\mu)) - \phi(\mu)][\mu - \underline{\mu}(\mu)] \leq 0$$

A contradiction! So, (A.3) holds for $\forall \mu_1 \in [\mu^*, \mu^{**}]$.

For $\mu_1 \in [\mu^{**}, 1]$, we have

$$\begin{aligned}
V_{\mu_1}(\mu_1, \mu_2)/2\sigma^2 c =& \phi'(\underline{\mu}(\mu_1))\underline{\mu}'(\mu_1)[\mu_2 - \underline{\mu}(\mu_1)] \\
\overset{(D_2)}{=} & \frac{\mu_2 - \underline{\mu}(\mu_1)}{\bar{\mu}(\mu_1) - \underline{\mu}(\mu_1)} \\
V_{\mu_1\mu_1}(\mu_1, \mu_2)/2\sigma^2 c =& \frac{-\underline{\mu}'(\mu_1)[\bar{\mu}(\mu_1) - \underline{\mu}(\mu_1)] - [\bar{\mu}'(\mu_1) - \underline{\mu}'(\mu_1)][\mu_2 - \underline{\mu}(\mu_1)]}{[\bar{\mu}(\mu_1) - \underline{\mu}(\mu_1)]^2} \\
=& \frac{1}{2\sigma^2 c} \frac{1}{[\bar{\mu}(\mu_1) - \underline{\mu}(\mu_1)]^3} [\frac{\mu_2 - \bar{\mu}(\mu_1)}{\phi'(\underline{\mu}(\mu_1))} - \frac{\mu_2 - \underline{\mu}(\mu_1)}{\phi'(\bar{\mu}(\mu_1))}] \\
\Rightarrow V_{\mu_1\mu_1}(\mu_1, \mu_2) =& \frac{1}{[\bar{\mu}(\mu_1) - \underline{\mu}(\mu_1)]^3} [\frac{\mu_2 - \bar{\mu}(\mu_1)}{\phi'(\underline{\mu}(\mu_1))} - \frac{\mu_2 - \underline{\mu}(\mu_1)}{\phi'(\bar{\mu}(\mu_1))}]
\end{aligned}$$

Since $\frac{\partial V_{\mu_1\mu_1}(\mu_1,\mu_2)}{\partial \mu_2} < 0$, we only needs to show that (A.3) holds for $\mu_2 = \underline{\mu}(\mu_1)$:

$$\begin{aligned}
& \mu_1^2(1-\mu_1)^2 V_{\mu_1\mu_1}(\mu_1, \underline{\mu}(\mu_1))/2\sigma^2 c \leq 1 \\
\Leftrightarrow & \frac{\mu_1^2(1-\mu_1)^2}{[\bar{\mu}(\mu_1) - \underline{\mu}(\mu_1)]^2} \frac{-1}{\phi'(\underline{\mu}(\mu_1))} \leq 1 \qquad \text{(A.4)}
\end{aligned}$$

Let's first show that $\underline{\mu}(\mu^{**}) \leq 1/2$ by contradiction. Suppose instead $\underline{\mu}(\mu^{**}) > 1/2$.

$$\begin{aligned}
p - \mu^{**} &= \frac{\bar{\mu}(\mu^{**}) + \underline{\mu}(\mu^{**})}{2} \\
\Leftrightarrow \; p - \mu^{**} &= \frac{\mu^{**} + \underline{\mu}(\mu^{**})}{2} \\
\Leftrightarrow \; \underline{\mu}(\mu^{**}) &= 2p - 3\mu^{**}
\end{aligned}$$

Hence, $2p - 3\mu^{**} > 1/2 \Rightarrow \mu^{**} < \frac{2}{3}p - \frac{1}{6}$. Since $\phi(x)$ is strictly decreasing in x, the first equation of (2.8) implies

$$\begin{aligned}
\frac{1}{2\sigma^2 c} =& \phi(\underline{\mu}(\mu^{**})) - \phi(\mu^{**}) \\
<& \phi(1/2) - \phi(\frac{2}{3}p - \frac{1}{6}) \\
\Leftrightarrow c >& \frac{1}{2\sigma^2[\phi(1/2) - \phi(\frac{2}{3}p - \frac{1}{6})]}
\end{aligned}$$

A contradiction! Therefore, $\underline{\mu}(\mu^{**}) \leq 1/2$. Since $\underline{\mu}(\mu_1)$ is decreasing in μ_1, we have $\underline{\mu}(\mu_1) \leq 1/2$, $\forall \mu \in [\mu^{**}, 1]$. One can see that the LHS of (A.4), $\frac{\mu_1^2(1-\mu_1)^2}{[\bar{\mu}(\mu_1)-\underline{\mu}(\mu_1)]^2}\frac{-1}{\phi'(\underline{\mu}(\mu_1))}$, decreases in $\mu_1 \in [\mu^{**}, 1]$. And we know that (A.4) holds for $\mu_1 = \mu^{**}$ (we have shown that (A.3) and thus (A.4) hold for $\forall \mu_1 \in [\mu^*, \mu^{**}]$). Therefore, (A.4) and thus (A.3) hold for $\forall \mu_1 \in [\mu^{**}, 1]$.

□

Proof of Proposition 11.

(1) Comparative statics w.r.t. p

We first consider $\bar{\mu}(\mu)$. Fixing an arbitrary $\mu \in (\mu^{**}, 1]$, recall the system of equations (2.7):

$$\phi(\underline{\mu}(\mu)) - \phi(\bar{\mu}(\mu)) = \frac{1}{2\sigma^2 c}$$
$$\psi(\underline{\mu}(\mu)) - \psi(\bar{\mu}(\mu)) = \frac{p-\mu}{2\sigma^2 c}$$

By the implicit function theorem, we obtain:

$$\begin{bmatrix} \frac{\partial \bar{\mu}(\mu)}{\partial p} \\ \frac{\partial \underline{\mu}(\mu)}{\partial p} \end{bmatrix} = - \begin{bmatrix} -\phi'(\bar{\mu}(\mu)) & \phi'(\underline{\mu}(\mu)) \\ -\psi'(\bar{\mu}(\mu)) & \psi'(\underline{\mu}(\mu)) \end{bmatrix}^{-1} \cdot \begin{bmatrix} 0 \\ -\frac{1}{2\sigma^2 c} \end{bmatrix}$$
$$= \begin{bmatrix} -\frac{\phi(\bar{\mu}(\mu))-\phi(\bar{\mu}(\mu))}{\phi'(\underline{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]} & > 0 \\ -\frac{\phi(\underline{\mu}(\mu))-\phi(\bar{\mu}(\mu))}{\phi'(\underline{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]} & > 0 \end{bmatrix}$$

We now consider $\underline{\mu}(\mu)$. Suppose there exists $p_1 > p_2$ with the corresponding quitting boundaries $(\mu_1, \underline{\mu}_{p_1}(\mu_1))$ and $(\mu_1, \underline{\mu}_{p_2}(\mu_1))$, respectively. Denote the cutoff beliefs by $(\mu^*_{p_1}, \mu^{**}_{p_1})$ for price p_1 and by $(\mu^*_{p_2}, \mu^{**}_{p_2})$ for price p_2. Fixing an arbitrary $\mu_1 \in (\mu^*_{p_1}, 1]$, we know that the consumer is indifferent between quitting and searching for information when her belief is $(\mu_1, \underline{\mu}_{p_1}(\mu_1))$ and the price is p_1. Since $p_2 < p_1$, one can see that the value of searching for information when her belief is $(\mu_1, \underline{\mu}_{p_1}(\mu_1))$ and the price is p_2 is strictly higher than zero. So, the consumer will keep searching for information. Thus, $\underline{\mu}_{p_2}(\mu_1) < \underline{\mu}_{p_1}(\mu_1)$.

Therefore, the entire search region shifts upwards as the price increases.

(2) Comparative statics w.r.t. c

We first consider $\bar{\mu}(\mu)$. Fixing an arbitrary $\mu \in (\mu^{**}, 1]$, recall the system of equations (2.7):

$$\phi(\underline{\mu}(\mu)) - \phi(\bar{\mu}(\mu)) = \frac{1}{2\sigma^2 c}$$
$$\psi(\underline{\mu}(\mu)) - \psi(\bar{\mu}(\mu)) = \frac{p-\mu}{2\sigma^2 c}$$

By the implicit function theorem, we obtain:

$$\begin{bmatrix}\frac{\partial\bar{\mu}(\mu)}{\partial c}\\ \frac{\partial\underline{\mu}(\mu)}{\partial c}\end{bmatrix} = -\begin{bmatrix}-\phi'(\bar{\mu}(\mu)) & \phi'(\underline{\mu}(\mu))\\ -\psi'(\bar{\mu}(\mu)) & \psi'(\underline{\mu}(\mu))\end{bmatrix}^{-1}\cdot\begin{bmatrix}\frac{1}{2\sigma^2c^2}\\ \frac{p-\mu}{2\sigma^2c^2}\end{bmatrix}$$
$$= \frac{1}{2\sigma^2c^2\phi'(\bar{\mu}(\mu))\phi'(\underline{\mu}(\mu))[\bar{\mu}(\mu)-\underline{\mu}(\mu)]}\cdot\begin{bmatrix}\phi'(\underline{\mu}(\mu))(p-\mu-\underline{\mu}(\mu))\\ \phi'(\bar{\mu}(\mu))(p-\mu-\bar{\mu}(\mu))\end{bmatrix}$$

The consumer purchases the product when the belief is $(\mu, \bar{\mu}(\mu))$. So, $\mu + \bar{\mu}(\mu) - p > 0$. The consumer stops searching and does not purchase the product when the belief is $(\mu, \underline{\mu}(\mu))$. So, $\mu + \underline{\mu}(\mu) - p < 0$. We also have $\phi'(x) = -\frac{1}{x^2(1-x)^2} \Rightarrow \phi'(x) < 0, \forall x$. Thus, we obtain:

$$\begin{bmatrix}\frac{\partial\bar{\mu}(\mu)}{\partial c} < 0\\ \frac{\partial\underline{\mu}(\mu)}{\partial c} > 0\end{bmatrix}$$

We now consider $\underline{\mu}(\mu)$. Suppose there exists $c_1 > c_2$ with the corresponding quitting boundaries $(\mu_1, \underline{\mu}_{c_1}(\mu_1))$ and $(\mu_1, \underline{\mu}_{c_2}(\mu_1))$, respectively. Denote the cutoff beliefs by $(\mu^*_{c_1}, \mu^{**}_{c_1})$ for price c_1 and by $(\mu^*_{c_2}, \mu^{**}_{c_2})$ for price c_2. Fixing an arbitrary $\mu_1 \in (\mu^*_{c_1}, 1]$, we know that the consumer is indifferent between quitting and searching for information when her belief is $(\mu_1, \underline{\mu}_{c_1}(\mu_1))$ and the price is c_1. Since $c_2 < c_1$, one can see that the value of searching for information when her belief is $(\mu_1, \underline{\mu}_{c_1}(\mu_1))$ and the price is c_2 is strictly higher than zero. So, the consumer will keep searching for information. Thus, $\underline{\mu}_{c_2}(\mu_1) < \underline{\mu}_{c_1}(\mu_1)$.

(3) Comparative statics w.r.t. σ^2

c and σ^2 always appear together as $2\sigma^2 c$ in the equations. So, the qualitative result of the comparative statics w.r.t. σ^2 is the same as the comparative statics w.r.t. c.

□

Proof of Proposition 12. We first consider $\mu_1 \in [\mu^{**}, 1]$ and $\mu_1 \geq \mu_2$. Under this circumstance, the consumer only learns about attribute two until μ_2 hits either the purchasing boundary or the quitting boundary. As μ_2 is a martingale, by Dynkin's formula, we get:

$$P(\mu_1, \mu_2) := \mathbb{P}[\text{purchasing}|\text{starting at } (\mu_1, \mu_2)] = \frac{\mu_2 - \underline{\mu}(\mu_1)}{\bar{\mu}(\mu_1) - \underline{\mu}(\mu_1)} \tag{A.5}$$

Now we consider $\mu_1 \in [\mu^*, \mu^{**}]$ and $\mu_1 \geq \mu_2$. The belief either hits (μ^{**}, μ^{**}) and the consumer purchases the good or the belief hits $\{(x, \underline{\mu}(x)) : x \in [\mu_1, \mu^{**})\} \cup \{(\underline{\mu}(x), x) : x \in [\mu_1, \mu^{**})\}$ and the consumer quits. To calculate the purchasing likelihood, let's first calculate the likelihood of the belief hitting $(\mu_1, \underline{\mu}(\mu_1))$ before hitting the main diagonal (μ_1, μ_1), $q(\mu_1, \mu_2)$.

$$q(\mu_1, \mu_2) = \frac{\mu_1 - \mu_2}{\mu_1 - \underline{\mu}(\mu_1)}$$

Now we calculate the probability of purchasing given belief (μ, μ), $\tilde{P}(\mu)$ by consider the infinitesimal learning on attribute two. Noticing that $q(\mu, \mu) = 0, \frac{\partial q}{\partial \mu_1}|_{\mu_1=\mu_2=\mu} = \frac{1}{\mu - \underline{\mu}(\mu)}, \frac{\partial q}{\partial \mu_2}|_{\mu_1=\mu_2=\mu} = -\frac{1}{\mu - \underline{\mu}(\mu)}$, we have:

$$\begin{aligned}
\tilde{P}(\mu) &= \frac{1}{2}\mathbb{P}[\text{purchasing}|(\mu,\mu), d\mu \geq 0] + \frac{1}{2}\mathbb{P}[\text{purchasing}|(\mu,\mu), d\mu < 0] \\
&= \frac{1}{2}[1 - q(\mu + |d\mu|, \mu)]\tilde{P}(\mu + |d\mu|) + \frac{1}{2}[1 - q(\mu - |d\mu|, \mu)]\tilde{P}(\mu) \\
&= \tilde{P}(\mu) + \frac{|d\mu|}{2}\tilde{P}'(\mu) + |d\mu|\frac{\tilde{P}(\mu)}{\underline{\mu}(\mu) - \mu} + o(d\mu) \\
\Rightarrow 0 &= \frac{|d\mu|}{2}\left[\tilde{P}'(\mu) + 2\frac{\tilde{P}(\mu)}{\underline{\mu}(\mu) - \mu}\right] + o(d\mu) \\
\Rightarrow \frac{\tilde{P}'(\mu)}{\tilde{P}(\mu)} &= -\frac{2}{\underline{\mu}(\mu) - \mu}, \ \forall \mu \in (\mu^*, \mu^{**})
\end{aligned}$$

, where the last equality comes from dividing the previous equation by $|d\mu|$ and take the limit of $d\mu$ to 0. Together with the initial condition $\tilde{P}(\mu^{**}) = 1$, we obtain:

$$\tilde{P}(\mu) = e^{-\int_{\mu}^{\mu^{**}} \frac{2}{x - \underline{\mu}(x)} dx}$$

In sum, the purchasing likelihood when $\mu_1 \geq \mu_2$ and $\mu_1 \in (\mu^*, \mu^{**})$ is:

$$P(\mu_1, \mu_2) = \mathbb{P}[\text{purchasing}|\text{starting at } (\mu_1, \mu_2)] = [1 - q(\mu_1, \mu_2)]\tilde{P}(\mu_1) = h(\mu_1, \mu_2)\tilde{P}(\mu_1) \tag{A.6}$$

, where $h(\mu_1, \mu_2) = \frac{\mu_2 - \underline{\mu}(\mu_1)}{\mu_1 - \underline{\mu}(\mu_1)}$.

By symmetry, the purchasing likelihood when $\mu_1 < \mu_2$ and $\mu_2 \in (\mu^*, \mu^{**})$ is:

$$P(\mu_1, \mu_2) = P(\mu_2, \mu_1) = [1 - q(\mu_2, \mu_1)]\tilde{P}(\mu_2) = h(\mu_2, \mu_1)\tilde{P}(\mu_2) \tag{A.7}$$

□

Proof of Proposition 13. One can see that the consumer will not purchase the product if $\mu_1 \leq \underline{\mu}(1)$, even if the firm advertises one attribute which turns out to be good. So, the firm does not advertise if $\mu_1 \leq \underline{\mu}(1)$. Also, the consumer will purchase the product for sure if $\mu_2 \geq \bar{\mu}(\mu_1)$ without advertising. So, the firm does not advertise if $\mu_2 \geq \bar{\mu}(\mu_1)$. We now look at other cases.

(1) $\mu_1 > \underline{\mu}(1)$ and $\mu_2 \leq \underline{\mu}(1)$ (Region I_1 and I_2)

The consumer will never purchase the product if the firm advertises attribute one or does not advertise. In contrast, the consumer may purchase theproduct if the firm advertises on attribute two. The consumer will not purchase if attribute two is bad. However, if attribute two is good, the consumer will purchase the product immediately in the region I_2, and will search for information about attribute one in the region I_1. In the region I_1, the consumer will purchase the product after receiving enough positive information. So, the purchasing likelihood is strictly positive. Hence, the firm advertises attribute two.

(2) $\mu_1 \in (\underline{\mu}(1), \bar{\mu}(1)]$ and $\mu_2 > \underline{\mu}(1)$ (Region I_3)

The purchasing probability is zero if the firm does not advertise, and is positive if the firm advertises either attriutes. Thus, we need to compare the purchasing likelihoods between advertising attribute one and two. We use $P_i(\mu_1, \mu_2)$ to denote the purchasing probability when the prior belief is (μ_1, μ_2) and the firm advertises attribute i.

$$\begin{aligned} P_1(\mu_1, \mu_2) = &\ \mu_1 \cdot \frac{\mu_2 - \underline{\mu}(1)}{\bar{\mu}(1) - \underline{\mu}(1)} \\ P_2(\mu_1, \mu_2) = &\ \mu_2 \cdot \frac{\mu_1 - \underline{\mu}(1)}{\bar{\mu}(1) - \underline{\mu}(1)} \\ &\overset{\mu_1 \geq \mu_2}{\geq} P_1(\mu_1, \mu_2) \end{aligned}$$

, where the inequality is strict if $\mu_1 > \mu_2$. So, the firm advertises attribute two.

(3) $\mu_1 > \bar{\mu}(1)$ and $\mu_2 \in (\underline{\mu}(1), \bar{\mu}(\mu))$ (Region I_4, the diagonal striped black region, and the white search region)

To characterize the advertising strategy, we need to determine two things. First, whether the firm wants to advertise. Second, whether the firm prefers advertising attribute one or two, conditional on advertising.

We first compare advertising attribute one and two.

$$\begin{aligned} P_1(\mu_1, \mu_2) = &\ \mu_1 \cdot \frac{\mu_2 - \underline{\mu}(1)}{\bar{\mu}(1) - \underline{\mu}(1)} \\ P_2(\mu_1, \mu_2) = &\ \mu_2 \\ P_1(\mu_1, \mu_2) > P_1(\mu_1, \mu_2) \Leftrightarrow &\ \frac{\mu_2 - \underline{\mu}(1)}{\bar{\mu}(1) - \underline{\mu}(1)} > \frac{\mu_2}{\mu_1} \\ \Leftrightarrow &\ \mu_2 > \frac{\underline{\mu}(1)\mu_1}{\mu_1 - \bar{\mu}(1) + \underline{\mu}(1)} := \tilde{\mu}(\mu_1) \end{aligned}$$

So, the firm prefers advertising attribute one to advertising attribute two if and only if $\mu_2 > \tilde{\mu}(\mu_1)$. One can see that $\tilde{\mu}(\mu_1)$ decreases in μ_1.

We then determine whether the firm wants to advertise or not. If the belief is below the purchasing boundary, the firm always prefers advertising because the consumer will never purchase without advertising. Now suppose the belief is in the search region, $\mu_1 \in [\mu^{**}, 1]$ and $\mu_2 \in [\bar{\mu}(\mu_1), \mu_1]$. According to Proposition 12, the purchasing likelihood without advertising is:

$$P(\mu_1, \mu_2) = \frac{\mu_2 - \underline{\mu}(\mu_1)}{\bar{\mu}(\mu_1) - \underline{\mu}(\mu_1)}$$

If the firm advertises attribute one, the purchasing likelihood is:

$$P_1(\mu_1, \mu_2) = \begin{cases} \mu_1 \cdot \frac{\mu_2 - \underline{\mu}(1)}{\bar{\mu}(1) - \underline{\mu}(1)}, \text{ if } \mu_2 < \bar{\mu}(1) \\ \mu_1, \text{ if } \mu_2 \geq \bar{\mu}(1) \end{cases}$$

If the firm advertises attribute two, the purchasing likelihood is:

$$P_2(\mu_1, \mu_2) = \mu_2$$

Observe that $P(\mu_1, \underline{\mu}(\mu)) = 0$, $P(\mu_1, \underline{\mu}(\mu)) = 1$, $P_1(\mu_1, \underline{\mu}(1)) = 0$, $P_1(\mu_1, \bar{\mu}(\mu)) = \mu_1$, and $\underline{\mu}(1) \leq \underline{\mu}(\mu)$. By (quasi-) linearity of the purchasing likelihood, one can see that $P(\mu_1, \mu_2)$ crosses $P_1(\mu_1, \mu_2) \vee P_2(\mu_1, \mu_2)$ once as μ_2 increases, fixing a μ_1. Hence, there exists $\widehat{\mu}(\mu_1) \in [\tilde{\mu}(\mu_1), \bar{\mu}(\mu_1))$ such that the firm does not advertise if and only if $\mu_2 \geq \widehat{\mu}(\mu_1)$.

□

A.3 Appendix to Chapter 3

Proof of Proposition 14. The consumer's expected ex-ante payoff by choosing to reveal η proportion of information is:

$$U_0(\eta) = \begin{cases} -\mu_s \eta u_b + \frac{v^2}{4(1-\eta)t}, \ if \ \eta \leq 1 - \frac{v}{t} \\ -\mu_s \eta u_b + \frac{(1-\eta)t}{4}, \ if \ \eta > 1 - \frac{v}{t} \end{cases}$$

$U_0(\eta)$ decreases in η for $\eta > 1 - \frac{v}{t}$, so the consumer will not reveal more than $1 - \frac{v}{t}$ proportion of information. Consider $\eta \in [0, 1 - \frac{v}{t}]$.

$$\frac{dU_0(\eta)}{d\eta} = -\mu_s u_b + \frac{v^2}{4t(1-\eta)^2}$$

, which increases in η. So, the optimal η is either 0 or $1 - \frac{v}{t}$. $U_0(1 - \frac{v}{t}) \geq U_0(0) \Leftrightarrow \mu_s \leq \widehat{\mu}$, where $\widehat{\mu} = \frac{v}{4u_b}$. □

Proof of Proposition 15. Since the rational type does not sell the data, a signal $s = y$ implies the firm is the behavioral type. So, $\mu_{t+1} = 1$. Now consider $s = n$. By Baye's rule,

$$\begin{aligned} \mathbb{P}[type\ B|s=n] &= \frac{\mathbb{P}[s=n|type\ B]\mathbb{P}[type\ B]}{\mathbb{P}[s=n|type\ B]\mathbb{P}[type\ B] + \mathbb{P}[s=n|type\ R]\mathbb{P}[type\ R]} \\ &= \frac{(1-q)\mu_t}{(1-q)\mu_t + 1\cdot(1-\mu_t)} \\ &= \frac{1-q}{1-q\mu_t}\mu_t \end{aligned}$$

By induction, we have $\mu_{t+1} = \frac{(1-q)\mu_t}{(1-q)\mu_t+1-\mu_t}$ after receiving signal n once. Suppose $\mu_{t+k} = \frac{(1-q)^k}{(1-q)^k\mu_t+1-\mu_t}\mu_t$ after receiving signal n for k consecutive periods. After receiving signal n for $k+1$ consecutive periods, we have $\mu_{t+k+1} = \frac{(1-q)\mu_{t+k}}{(1-q)\mu_{t+k}+1-\mu_{t+k}} = \frac{(1-q)^{k+1}}{(1-q)^{k+1}\mu_t+1-\mu_t}\mu_t$. So, it shows that $\mu_{t+k} = \frac{(1-q)^k}{(1-q)^k\mu_t+1-\mu_t}\mu_t$ after receiving signal n for k consecutive periods.

One can see that $\frac{(1-q)^k}{(1-q)^k\mu_t+1-\mu_t}\mu_t$ approaches 0 as $k \to +\infty$. □

Proof of Proposition 16. Let $\widehat{k} = \left\lceil \frac{ln\frac{v(1-\mu_0)}{(4u_b-v)\mu_0}}{ln(1-q)} \right\rceil$. We first show that the consumer reveals $\eta = 1 - v/t$ proportion of information after $\widehat{k}$ periods, if the firm never sells the data in equilibrium. By Proposition 15, the belief after not selling data for k consecutive periods is $\mu_k = \frac{(1-q)^k}{(1-q)^k\mu_0+1-\mu_0}\mu_0$. By Proposition 14, consumer reveals $\eta = 1 - v/t$ proportion of information if and only if $\mu_k \leq \widehat{\mu} \Leftrightarrow k \geq \frac{ln\frac{v(1-\mu_0)}{(4u_b-v)\mu_0}}{ln(1-q)}$.

We now show that the rational firm has no incentive to deviate to selling data at any time. The game is continuous at infinity because of discounting. So, we can use the single-deviation property. Suppose the firm deviates once at period t when the belief is μ_t. There are two cases.

1. $\mu \leq \widehat{\mu}$

   The value function of the equilibrium strategy (never sell data) is:

   $V(\mu_t) = (1-\delta)\frac{v}{2}\frac{1}{1-\delta} = \frac{v}{2}$

   The value function of deviating once at period t is (assuming the firm sells data when the belief is 1, which maximizes the payoff):

   $\tilde{V}(\mu_t) = (1-\delta)[\frac{v}{2} + D(1-\frac{v}{t}) + \delta(q\frac{\frac{v^2}{2t}+D(0)}{1-\delta} + (1-q)\frac{\frac{v}{2}}{1-\delta})]$

   The rational firm will not deviate if $V(\mu_t) > \tilde{V}(\mu_t) \Leftrightarrow \frac{\delta}{1-\delta} > \frac{D(1-\frac{v}{t})}{q(v/2-v^2/2t-D(0))}$. One can see that $\exists \delta_1 \in (0,1)$ s.t. the inequality holds for any $\delta \geq \delta_1$.

2. $\mu > \widehat{\mu}$

   The value function of the equilibrium strategy (never sell data) is:

   $V(\mu_t) = (1-\delta)[\sum_{k=0}^{\widehat{k}-1}\delta^k\frac{v^2}{2t} + \sum_{k=\widehat{k}}^{+\infty}\delta^k\frac{v}{2}]$

   The value function of deviating once at period t is (assuming the firm sells data when the belief is 1, which maximizes the payoff):

   $\tilde{V}(\mu_t) = (1-\delta)[\frac{v^2}{2t} + D(0) + \delta(q\frac{\frac{v^2}{2t}+D(0)}{1-\delta} + (1-q)[\sum_{k=0}^{\widehat{k}-2}\delta^k\frac{v^2}{2t} + \sum_{k=\widehat{k}-1}^{+\infty}\delta^k\frac{v}{2}])]$

   The rational firm will not deviate if $V(\mu_t) > \tilde{V}(\mu_t) \Leftrightarrow \frac{\delta^{\widehat{k}}}{(1-\delta)[1-(1-q)\delta]} > \frac{D(0)}{q(v/2-v^2/2t)}$. One can see that $\exists \delta_2 \in (0,1)$ s.t. the inequality holds for any $\delta \geq \delta_2$.

Let $\widehat{\delta} = \max\{\delta_1, \delta_2\}$. One can see that $\widehat{\delta} < 1$ and for any $\delta > \widehat{\delta}$, the firm never sells consmer data, $\eta = 0$ in the first $\widehat{k}$ periods, and $\eta = 1 - v/t$ after $\widehat{k}$ periods is a MPE. □

Proof of Proposition 17. By Baye's rule, for a given firm,

$$
\begin{aligned}
&\mathbb{P}[type\ B|s=n] \\
=&\frac{\mathbb{P}[s=n|type\ B]\mathbb{P}[type\ B]}{\mathbb{P}[s=n|type\ B]\mathbb{P}[type\ B] + \mathbb{P}[s=n|type\ R]\mathbb{P}[type\ R]} \\
=&\frac{(1-q)[1\cdot(1-\mu_t)+(1-q)\mu_t]^{N-1}\mu_t}{(1-q)[1\cdot(1-\mu_t)+(1-q)\mu_t]^{N-1}\mu_t + 1\cdot[1\cdot(1-\mu_t)+(1-q)\mu_t]^{N-1}(1-\mu_t)} \\
=&\frac{1-q}{1-q\mu_t}\mu_t, \text{ which does not depend on } N.
\end{aligned}
$$

$$\begin{aligned}
&\mathbb{P}[type\ B|s=y] \\
=&\frac{\mathbb{P}[s=y|type\ B]\mathbb{P}[type\ B]}{\mathbb{P}[s=y|type\ B]\mathbb{P}[type\ B]+\mathbb{P}[s=y|type\ R]\mathbb{P}[type\ R]} \\
=&\frac{\left[1-(1-q)[1\cdot(1-\mu_t)+(1-q)\mu_t]^{N-1}\right]\mu_t}{\left[1-(1-q)[1\cdot(1-\mu_t)+(1-q)\mu_t]^{N-1}\right]\mu_t+\left[1-1\cdot[1\cdot(1-\mu_t)+(1-q)\mu_t]^{N-1}\right](1-\mu_t)} \\
=&\frac{1-(1-q)(1-q\mu_t)^{N-1}}{1-(1-q\mu_t)^N}\mu_t, \text{ which decreases in } N \text{ by checking the derivative.}
\end{aligned}$$

□

Proof of Proposition 18. Fix $\delta \in (0,1)$. Suppose $\forall N_\delta$, $\exists N \geq N_\delta$ s.t. there exists a MPE in which a rational firm (label it by firm 1 WLOG) does not sell the data at $t=0$. Denote the equilibrium strategy of all the firms by σ and the value function of firm 1 by $V_1(\cdot)$. The prior belief is $\vec{\mu}_0 = (\mu_0, \mu_0, ..., \mu_0)$. Denote the posterior belief upon observing signal y (n) by $\vec{\mu}^y$ ($\vec{\mu}^n$) when the initial belief is $\vec{\mu}$ and the equilibrium strategy is σ.

Suppose $\mu_0 > \widehat{\mu}$.

$$V_1(\vec{\mu_0}) = (1-\delta)\frac{v^2}{2t} + \delta\left[\mathbb{P}(s=y|\sigma)V_1(\vec{\mu}_0^y) + \mathbb{P}(s=n|\sigma)V_1(\vec{\mu}_0^n)\right]$$

, where $\begin{cases} \mathbb{P}(s=n|\sigma) \leq (1-q\mu_0)^{N-1} \\ \mathbb{P}(s=y|\sigma) = 1-\mathbb{P}(s=n|\sigma) \geq 1-(1-q\mu_0)^{N-1} \end{cases}$

The upper bound of the probability of signal n, $\mathbb{P}(s=n|\sigma)$, is obtained when no rational firm sells data under σ given belief $\vec{\mu}_0$. The value function of firm 1 if it deviates once in the first period (denote the strategy by σ') is:

$$V_{1,dev}(\vec{\mu}_0) = (1-\delta)\left(\frac{v^2}{2t} + D(0)\right) + \delta\left[\mathbb{P}(s=y|\sigma')V_1(\vec{\mu}_0^y) + \mathbb{P}(s=n|\sigma')V_1(\vec{\mu}_0^n)\right]$$

, where $\begin{cases} \mathbb{P}(s=n|\sigma') = (1-q)\mathbb{P}(s=n|\sigma) \\ \mathbb{P}(s=y|\sigma') = 1-\mathbb{P}(s=n|\sigma') \end{cases}$ Therefore, we have:

$$V_{1,dev}(\mu_0) - V_1(\mu_0) = (1-\delta)D(0) - \delta\left[V_1(\mu_0^n) - V_1(\mu_0^y)\right]q\mathbb{P}(s=n|\sigma)$$

Since $V_1(\cdot) \in [\frac{v^2}{2t}, \frac{v}{2} + D(1-\frac{v}{t})]$, $V_1(\mu_0^n) - V_1(\mu_0^y) \leq \frac{v}{2} + D(1-\frac{v}{t}) - \frac{v^2}{2t}$, which is a constant. $\mathbb{P}(s=n|\sigma) \leq (1-q\mu_0)^{N-1} \to 0$ $(N \to +\infty)$. Hence, given δ, $\exists N_\delta$ s.t. $\forall N \geq N_\delta$, $V_{1,dev}(\mu_0) - V_1(\mu_0) > 0$. But we assume that firm 1 does not sell the data at $t=0$. A contradiction.

In sum, firms always sell data and consumers reveal nothing at $t=0$ in equilibrium. Anticipating that, the consumer's posterior belief about each firm's type is μ_0. This repeats in every period. Therefore, for any $\delta \in (0,1), \exists N_\delta$ *s.t.* $\forall N \geq N_\delta$, firms always sell data and consumers reveal nothing in the unique MPE. Consumer's belief about the firm's type is always μ_0.

Suppose $\mu_0 \leq \widehat{\mu}$. The first period payoff will be $\frac{v}{2}$ in equilibrium and $\frac{v}{2} + D(1-\frac{v}{t})$ if the firm deviates. All the remaining proof is the same as above. □

Proof of Proposition 19. Consider firm 1 WLOG. We first list the updated belief after one signal:

$$\begin{cases} \mu^y = \mathbb{P}(\text{firm 1 is bad type}|s=y,\ \text{initial belief is } \mu) = \frac{1-(1-q)(1-q\mu)}{1-(1-q\mu)^2}\mu \\ \mu^n = \mathbb{P}(\text{firm 1 is bad type}|s=n,\ \text{initial belief is } \mu) = \frac{1-q}{1-q\mu}\mu \end{cases}$$

Both μ^y and μ^n increase in μ. $\mu^y \geq 1/2,\ \forall \mu$.

Suppose there exists an equilibrium in which rational firms never sells the data. Then consumers have identical beliefs for both firms. Denote the corresponding value function by $V(\cdot)$. Consider a belief $\mu > \widehat{\mu}$.

$$V(\mu) = (1-\delta)\frac{v^2}{2t} + \delta\left[q\mu V(\mu^y) + (1-q\mu)V(\mu^n)\right]$$

The value function of deviating once in the current period is:

$$V_{dev}(\mu) = (1-\delta)\left(\frac{v^2}{2t} + D(0)\right) + \delta\left[q[1+(1-q)\mu]V(\mu^y) + (1-q)(1-q\mu)V(\mu^n)\right]$$

$$V_{dev}(\mu) - V(\mu) = (1-\delta)D(0) - \delta\left[V(\mu^n) - V(\mu^y)\right]q(1-q\mu) \tag{A.8}$$

$v < 2u_b \Rightarrow \widehat{\mu} < 1/2$. $\mu^y \geq 1/2,\ \forall \mu$ implies that consumer will reveal no information after one signal y, which gives the rational firm a stage equilibrium payoff of $\frac{v^2}{2t}$. If the signal is n and $\mu^n \leq \widehat{\mu}$, rational firm gets a stage payoff of $v/2$; If the signal is n and $\mu^n > \widehat{\mu}$, rational firm gets a stage payoff of $\frac{v^2}{2t}$. So, we gets an upper bound of $V(\mu^n)$ by assuming that the belief is always no greater than $\widehat{\mu}$:

$$\begin{aligned} V(\mu^n) &\leq (1-\delta)\left[\frac{v}{2} + \sum_{k=1}^{+\infty}\delta^k[q\mu\frac{v^2}{2t} + (1-q\mu)\frac{v}{2}]\right] \\ &= (1-\delta)\frac{v}{2} + \delta[q\mu\frac{v^2}{2t} + (1-q\mu)\frac{v}{2}] \end{aligned}$$

Always selling consumer data gives a lower bound on the value function:

$$V(\mu^y) \geq (1-\delta)\sum_{k=0}^{+\infty}\delta^k[\frac{v^2}{2t} + D(0)] = \frac{v^2}{2t} + D(0)$$

Hence, we have:

$$V(\mu^n) - V(\mu^y) \leq (1-\delta)\frac{v}{2} + \delta[q\mu\frac{v^2}{2t} + (1-q\mu)\frac{v}{2}] - [\frac{v^2}{2t} + D(0)]$$

Plug it back to (A.8), we have:

$$\begin{aligned} &V_{dev}(\mu) - V(\mu) \\ \geq &(1-\delta)[D(0) - \delta q(1-q\mu)\frac{v}{2}] - \delta q(1-q\mu)\left[\delta[q\mu\frac{v^2}{2t} + (1-q\mu)\frac{v}{2}] - [\frac{v^2}{2t} + D(0)]\right] \end{aligned} \tag{A.9}$$

With a strictly positive probability, the signal will be y for k consecutive periods, $\forall k$. Denote the belief after k consecutive signal y by μ^{y^k}. One can see that $\mu^y \in (\mu, 1), \forall \mu \in (0,1)$. So, μ^{y^k} strictly increases in k and is bounded by 1. Thus, $\{\mu^{y^k}\}_{k=1}^{+\infty}$ has a limit. Denote the limit by $\mu^{y^{+\infty}}$. We have $(\mu^{y^{+\infty}})^y = \mu^{y^{+\infty}} \Rightarrow \mu^{y^{+\infty}} = 1$. So, μ^{y^k} could be arbitrarily close to 1 with a strictly positive probability. If $(1-q)(v/2 - v^2/2t) < D(0)$, for large enough δ and μ, we have $\delta[q\mu\frac{v^2}{2t} + (1-q\mu)\frac{v}{2}] - [\frac{v^2}{2t} + D(0)] < 0$. If $q(1-q)v/2 < D(0)$, for large enough δ and μ, we have $D(0) - \delta q(1-q\mu)\frac{v}{2} > 0$. Together, we get that $(1-\delta)[D(0) - \delta q(1-q\mu)\frac{v}{2}] - \delta q(1-q\mu)\left[\delta[q\mu\frac{v^2}{2t} + (1-q\mu)\frac{v}{2}] - [\frac{v^2}{2t} + D(0)]\right] > 0, \overset{(A.9)}{\Rightarrow} V_{dev}(\mu) - V(\mu) > 0$. Therefore, rational firm will sell the data when the belief is μ and the discount factor is high enough. A contradiction. □

Proof of Proposition 20. Suppose there exists such an equilibrium.

(1) $\mu \leq \widehat{\mu}$

Without any monitoring effort, the consumer does not get an additional signal. So, the consumer reveal $1 - v/t$ amount of information according to Proposition 14. The expected consumer surplus is:

$$CS(0,\mu) := -\mu u_b(1 - v/t) + v/4$$

By incuring effort h, the consumer receives an extra signal s_h. The updated belief will be:

$$\begin{cases} \frac{1-h}{1-h\mu}\mu, & if\ s_h = n \text{ (with probability } 1-\mu h) \\ 1, & if\ s_h = y \text{(with probability } \mu h) \end{cases}$$

The expected consumer surplus is:[7]

$$CS(h,\mu) := -c(h) - \frac{1-h}{1-h\mu}\mu u_b(1 - \frac{v}{t})(1-\mu h) + \frac{v^2}{4t}\mu h + \frac{v}{4}(1-\mu h),\ h \in [0, \bar{h}]$$

The difference of the expected consumer surplus between incurring monitoring effort h and no effort is:

$$\Delta CS(h,\mu) := CS(h,\mu) - CS(0,\mu) = -c(h) + \mu\frac{v}{4t}(t-v)h\left[-1 + \frac{4u_b(1-\mu)}{v(1-\mu h)}\right],\ h \in [0, \bar{h}] \tag{A.10}$$

[7]Technically, the domin is $h \in (0, \bar{h}]$. But, one can check that the expression holds for $h = 0$ as well.

The consumer incurs a strictly monitoring effort if and only if $\Delta CS(h,\mu) > 0$ for some $h \in (0,\bar{h}]$. Notice that $\Delta CS(0,\mu) = 0$. So, a sufficient condition for the consumer to incur a strictly monitoring effort is $\frac{\partial \Delta CS(h,\mu)}{\partial h}|_{h=0} > 0$.

$$\frac{\partial \Delta CS(h,\mu)}{\partial h} = -c'(h) + \frac{\mu v(t-v)}{4t}\left[-1 + \frac{4u_b(1-\mu)}{v(1-\mu h)^2}\right]$$
$$\frac{\partial \Delta CS(h,\mu)}{\partial h}|_{h=0} > 0 \Leftrightarrow -1 + \frac{4u_b(1-\mu)}{v} > 0$$
$$\Leftrightarrow \mu < 1-\widehat{\mu}$$

If $v < 2u_b$, $\widehat{\mu} < 1/2 \Rightarrow \mu \leq \widehat{\mu} < 1-\widehat{\mu}$, $\forall \mu \leq \widehat{\mu}$. So, the consumer always incurs a strictly positive monitoring effort when $\mu < \widehat{\mu}$.

Equation (A.10) implies that $h^*(\mu) \to 0$ as $\mu \to 0$, since $\Delta CS(h^*(\mu),\mu) \geq 0$

(2) $\mu > \widehat{\mu}$

Without any monitoring effort, the consumer does not get an additional signal. So, the consumer reveal nothing according to Proposition 14. The expected consumer surplus is:

$$\widetilde{CS}(0,\mu) := v^2/4t$$

By incuring effort h, the consumer receives an extra signal s_h. The updated belief will be:

$$\begin{cases} \frac{1-h}{1-h\mu}\mu, & if\ s_h = n \text{ (with probability } 1-\mu h) \\ 1, & if\ s_h = y \text{(with probability } \mu h) \end{cases}$$

If μ is high enough such that $\frac{1-\bar{h}}{1-\bar{h}\mu}\mu > \widehat{\mu}$, the consumer will not reveal anything regardless of the signal realization. So, there is no gain from an additional signal and the consumer will not acquire an extra signal. For the consumer to incur costly monitoring, she must reveal some information ($\eta = 1 - v/t$) if $s_h = n$.[8] The expected consumer surplus is the same as the first case:

$$\widetilde{CS}(h,\mu) = CS(h,\mu) = -c(h) - \frac{1-h}{1-h\mu}\mu u_b(1-\frac{v}{t})(1-\mu h) + \frac{v^2}{4t}\mu h + \frac{v}{4}(1-\mu h),\ h \in (0,\bar{h}]$$

The difference of the expected consumer surplus between incurring monitoring effort h and no effort is:

$$\begin{aligned} \Delta\widetilde{CS}(h,\mu) :=\ & \widetilde{CS}(h,\mu) - \widetilde{CS}(0,\mu) \\ =\ & -c(h) + (1-\mu h)\frac{v(t-v)}{4t}\left[1 - \frac{4u_b(1-h)\mu}{v(1-\mu h)}\right],\ h \in (0,\bar{h}] \end{aligned}$$

[8]This may be worse than revealing nothing for the consumer. But the latter is always dominated by not incurring monitoring costs.

The consumer incurs a strictly monitoring effort if and only if $\Delta\widetilde{CS}(h,\mu) > 0$ for some $h \in (0, \bar{h}]$.

$$\Delta\frac{\partial\widetilde{CS}(h,\mu)}{\partial h} = -\,c'(h) + \frac{\mu v(t-v)}{4t}\left(\frac{4u_b}{v} - 1\right)$$

Since $c'(0) = 0$, $c(\cdot)$ is convex, and $\lim_{h\to\bar{h}} c'(h) = +\infty$, we have $\max_{0<h\leq\bar{h}} \Delta\widetilde{CS}(h,\mu) = \Delta\widetilde{CS}(\widehat{h}(\mu),\mu)$, where $\widehat{h}(\mu)(>0)$ is determined by the first order condition:

$$c'(\widehat{h}(\mu)) = \frac{\mu v(t-v)}{4t}\left(\frac{4u_b}{v} - 1\right) \tag{A.11}$$

The consumer's optimal effort is either 0 or $\widehat{h}(\mu)$. This leads to the following lemma.

Lemma 16. *Suppose the belief is $\mu > \widehat{\mu}$. The consumer incurs monitoring effort $\widehat{h}(\mu)$ if and only if $\Delta\widetilde{CS}(\widehat{h}(\mu),\mu) > 0$.*

The next lemma characterize the optimal effort of the consumer.

Lemma 17. *There exists a $\widehat{\widehat{\mu}} > \widehat{\mu}$ such that the consumer incurs efforts in monitoring if and only if $\mu \leq \widehat{\widehat{\mu}}$. For $\mu \in [\widehat{\mu}, \widehat{\widehat{\mu}}]$, the optimal effort $h^*(\mu)$ strictly increases in μ.*

Proof. We first show that the optimal effort follows a cutoff strategy. Suppose the consumer incurs monitoring effort $\widehat{h}(\mu_1) > 0$ when the belief is $\mu_1 > \widehat{\mu}$. $\Delta\widetilde{CS}(h)$ and $\widehat{h}$ depend on μ. Lemma 16 implies that $\Delta\widetilde{CS}(\widehat{h}(\mu_1),\mu_1) > 0$. $\forall\mu_2 \in (\widehat{\mu},\mu_1)$. Since $\Delta\widetilde{CS}(h,\mu)$ strictly decreases in μ, we have that $\Delta\widetilde{CS}(\widehat{h}(\mu_2),\mu_2) \geq \widetilde{CS}(\widehat{h}(\mu_1),\mu_2) > \Delta\widetilde{CS}(\widehat{h}(\mu_1),\mu_1) > 0$, where the first inequality is implied by the optimality of $\widehat{h}(\mu_2)$ for $\widetilde{CS}(h,\mu_2)$, $h \in (0,\bar{h}]$. Therefore, there exists a $\widehat{\widehat{\mu}} \geq \widehat{\mu}$ such that the consumer incurs efforts in monitoring if and only if $\mu \leq \widehat{\widehat{\mu}}$. For $\mu \in [\widehat{\mu}, \widehat{\widehat{\mu}}]$, the optimal effort $h^*(\mu) = \widehat{h}(\mu)$. According to equation (A.11), $\widehat{h}(\mu)$ strictly increases in μ.

We now show that $\widehat{\widehat{\mu}} > \widehat{\mu}$.This can be shown by consider $\mu = \widehat{\mu} + \varepsilon, h = \sqrt{\varepsilon}$. Taking Taylor expansion in the expression for $\Delta\widetilde{CS}(h,\mu)$ and let $\varepsilon \to 0$ gives the result.

So, we finish the proof of the lemma. □

Consumer optimality has been shown in the above analyses. Noticing that the benefit for not selling data and the penalty for selling data are higher under endogenous monitoring. One can see that the rational firm has no incentive to deviate when it is patient enough by similar arguments as the proof of Proposition 16.

□

Proof of Proposition 21. The proof of Proposition 18 applies to this case as well. □

Proof of Proposition 22. Fix $\delta \in (0,1)$. Suppose $\forall N_\delta,\ \exists N \geq N_\delta$ s.t. there exists a MPE in which a rational firm j does not sell the data at $t=0$. Denote the equilibrium strategy of all the firms by σ and the value function of firm i by $V_i(\cdot)$. The prior belief is $\vec{\mu}_0 = (\mu_0, \mu_0, ..., \mu_0)$. Denote the posterior belief upon observing signal y (n) by $\vec{\mu}^y$ ($\vec{\mu}^n$) when the initial belief is $\vec{\mu}$ and the equilibrium strategy is σ.

Suppose $\mu_0 > \widehat{\mu}$. There are two possibilities:

(1) $j = 1$

$$V_1(\vec{\mu_0}) = (1-\delta)\frac{v^2}{2t} + \delta\left[\mathbb{P}(s=y|\sigma)V_1(\vec{\mu}_0^y) + \mathbb{P}(s=n|\sigma)V_1(\vec{\mu}_0^n)\right]$$

, where $\begin{cases} \mathbb{P}(s=n|\sigma) \leq (1-\alpha q\mu_0)^{N-1} \\ \mathbb{P}(s=y|\sigma) = 1-\mathbb{P}(s=n|\sigma) \geq 1-(1-\alpha q\mu_0)^{N-1} \end{cases}$

The upper bound of the probability of signal n, $\mathbb{P}(s=n|\sigma)$, is obtained when no rational firm sells data under σ given belief $\vec{\mu}_0$. The value function of firm 1 if it deviates once in the first period (denote the strategy by σ') is:

$$V_{1,dev}(\vec{\mu}_0) = (1-\delta)\left(\frac{v^2}{2t} + D(0)\right) + \delta\left[\mathbb{P}(s=y|\sigma')V_1(\vec{\mu}_0^y) + \mathbb{P}(s=n|\sigma')V_1(\vec{\mu}_0^n)\right]$$

, where $\begin{cases} \mathbb{P}(s=n|\sigma') = (1-q)\mathbb{P}(s=n|\sigma) \\ \mathbb{P}(s=y|\sigma') = 1-\mathbb{P}(s=n|\sigma') \end{cases}$ Therefore, we have:

$$V_{1,dev}(\mu_0) - V_1(\mu_0) = (1-\delta)D(0) - \delta\left[V_1(\mu_0^n) - V_1(\mu_0^y)\right] q\mathbb{P}(s=n|\sigma)$$

Since $V_1(\cdot) \in [\frac{v^2}{2t}, \frac{v}{2} + D(1-\frac{v}{t})]$, $V_1(\mu_0^n) - V_1(\mu_0^y) \leq \frac{v}{2} + D(1-\frac{v}{t}) - \frac{v^2}{2t}$, which is a constant. $\mathbb{P}(s=n|\sigma) \leq (1-\alpha q\mu_0)^{N-1} \to 0$ ($N \to +\infty$). Hence, given δ, $\exists N_\delta$ s.t. $\forall N \geq N_\delta$, $V_{1,dev}(\mu_0) - V_1(\mu_0) > 0$. But we assume that firm 1 does not sell the data at $t=0$. A contradiction.

(2) $j \neq 1$

$$V_j(\vec{\mu_0}) = (1-\delta)\frac{v^2}{2t} + \delta\left[\mathbb{P}(s=y|\sigma)V_j(\vec{\mu}_0^y) + \mathbb{P}(s=n|\sigma)V_j(\vec{\mu}_0^n)\right]$$

, where $\begin{cases} \mathbb{P}(s=n|\sigma) \leq (1-q\mu_0)(1-\alpha q\mu_0)^{N-2} \\ \mathbb{P}(s=y|\sigma) = 1-\mathbb{P}(s=n|\sigma) \geq 1-(1-q\mu_0)(1-\alpha q\mu_0)^{N-2} \end{cases}$

The upper bound of the probability of signal n, $\mathbb{P}(s=n|\sigma)$, is obtained when no rational firm sells data under σ given belief $\vec{\mu}_0$. The value function of firm j if it deviates once in the first period (denote the strategy by σ') is:

$$V_{j,dev}(\vec{\mu}_0) = (1-\delta)\left(\frac{v^2}{2t} + D(0)\right) + \delta\left[\mathbb{P}(s=y|\sigma')V_j(\vec{\mu}_0^y) + \mathbb{P}(s=n|\sigma')V_j(\vec{\mu}_0^n)\right]$$

, where $\begin{cases} \mathbb{P}(s=n|\sigma') = (1-\alpha q)\mathbb{P}(s=n|\sigma) \\ \mathbb{P}(s=y|\sigma') = 1-\mathbb{P}(s=n|\sigma') \end{cases}$ Therefore, we have:

$$V_{j,dev}(\mu_0) - V_j(\mu_0) = (1-\delta)D(0) - \delta\left[V_j(\mu_0^n) - V_j(\mu_0^y)\right]\alpha q \mathbb{P}(s=n|\sigma)$$

Since $V_j(\cdot) \in [\frac{v^2}{2t}, \frac{v}{2} + D(1-\frac{v}{t})]$, $V_j(\mu_0^n) - V_j(\mu_0^y) \leq \frac{v}{2} + D(1-\frac{v}{t}) - \frac{v^2}{2t}$, which is a constant. $\mathbb{P}(s=n|\sigma) \leq (1-q\mu_0)(1-\alpha q\mu_0)^{N-2} \to 0$ $(N \to +\infty)$. Hence, given δ, $\exists N_\delta$ s.t. $\forall N \geq N_\delta$, $V_{j,dev}(\mu_0) - V_j(\mu_0) > 0$. But we assume that firm j does not sell the data at $t=0$. A contradiction.

In sum, firms always sell data and consumers reveal nothing at $t=0$ in equilibrium. Anticipating that, the consumer's posterior belief about each firm's type is μ_0. This repeats in every period. Therefore, for any $\delta \in (0,1), \exists N_\delta$ $s.t.$ $\forall N \geq N_\delta$, firms always sell data and consumers reveal nothing in the unique MPE. Consumer's belief about the firm's type is always μ_0.

Suppose $\mu_0 \leq \widehat{\mu}$. The first period payoff will be $\frac{v}{2}$ in equilibrium and $\frac{v}{2} + D(1-\frac{v}{t})$ if the firm deviates. All the remaining proof is the same as above. □