Consumer Attention Allocation and Firm Strategies

Qitian Ren

Submitted in partial fulfillment of the requirements for

the degree of Doctor of Philosophy under

the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

Consumer Attention Allocation and Firm Strategies

Qitian Ren

Nowadays consumers can easily access to vast amounts of product information before making a purchase. Yet, limitations on the ability to process information force consumers to make choices regarding the subjects to which they pay more or less attention. In this dissertation, I study how a consumer optimally allocates attention to various product information before making a purchase decision and how a seller should design the marketing strategies taking into account the consumer's attention allocation decision. I find that either a consumer engages in “confirmatory” search under which she searches more information that favors her prior belief or the consumer engages in “disconfirmatory” search under which she searches more information that disfavors her prior belief. In particular, the consumer conducts more disconfirmatory search when the information processing cost is low, while she conducts more confirmatory search when the cost is high. This suggests that “confirmatory bias” widely studied in psychology literature could be optimal behavior coming out of people optimizing attention to different types of information, especially when people has high information processing costs. Furthermore, a consumer's purchase likelihood may vary with her information processing cost in a non-monotonic way, depending on the consumer's prior belief and the utilities of buying a matched product and a mismatched product. Moreover, I show that when more information becomes available or credible, the consumer would increase attention to negative information when the prior utility of the product is high but she would increase attention to positive information when the prior utility is low. In terms of seller's strategies, I find that when the consumer has a low information processing cost,

the seller would charge a relatively high price such that consumers always process information; but when the consumer has a high information processing cost, the seller would charge a relatively low price such that consumers purchase the product without any learning. The optimal price and profit would first decrease and then increase in consumer's information processing cost. In addition, offering the return policy induces the consumer to pay more attention to positive information and less attention to negative information, and the seller would offer the return policy except when the consumer has a very high information processing cost. Finally, when a seller can influence the information environment, he would have a lower incentive to suppress the negative information when the consumer has a lower prior belief about product fit. Moreover, a higher information processing cost for a consumer would increase or decrease a seller's incentive to suppress the negative information in the environment, depending on whether the seller can adjust the product price and whether the consumer has a high or low prior belief. Interestingly, the seller may charge a lower price when he can fully control the information environment than when he can not.

Table of Contents

List of Figures

Acknowledgments

I am deeply indebted to my dissertation advisor Professor Kinshuk Jerath, whose insightful guidance and generous support made it possible for me to finish my dissertation. Over the past four years, I have learned much from him. In particular, whenever I was discussing with him about thoughts and research ideas, his sharp mind and clear exposition style always make complex ideas immediately accessible, which is something I strive to emulate. He taught me not only how to develop research ideas but also how to become a mature scholar. I am extremely delighted that I had an opportunity to learn from him.

I am also sincerely thankful to other members in my wonderful dissertation committee: Professor Miklos Sarvary, Professor Andrea Prat, Professor Ran Kivetz and Professor Monic Sun. I greatly appreciate their generous help and intellectual comments when completing this dissertation.

I am grateful to Professor Donald Lehmann who saw my potential as a researcher and supported me in the very beginning.

I am also thankful to other professors in and outside the Marketing department who spent their precious time on practicing with me for my job talk and offered insightful comments that greatly improved the content of my dissertation.

I have benefited a lot when learning from my peers Ryan Dew, Yu Ding, Fei Long, Jia Liu. I enjoyed those relaxing and intellectual chats with them when we were having coffee and dinner together. They not only offered incredible help in developing my research ideas but also brought great joy throughout my doctoral studies.

I want to express my deepest gratitude to my parents for their unconditional support. They have taught me to become a wise and brave man who is able to appreciate the beauty of life and overcome every obstacle in life.

The best thing that happened during my doctoral studies is finding my beloved wife, Xiaohan Zhang, who is also my best friend, soulmate, and perfect partner. Until today, it is still surprising to me that we share so many things in common. We both enjoy each other’s company, we both love holding each other’s hands when walking across the street, and we both fall in love with the starry night. You brought so much fun, trust and encouragement to me, not only making it a joyful journey to complete this dissertation but also making me a more confident and mature man with young mind. I love you for always standing by me, supporting me and appreciating my dream. It is the most beautiful blessing to have you by my side for the rest of my life, holding your hands and hoping together that we can reach the stars far away from home.

Chapter 1: Introduction

Consumers often acquire and process relevant product information to reduce uncertainty about a product's fit before making a purchase decision. For instance, they may inspect the product, read product descriptions and reviews, talk to friends, etc. Yet, limitations on the ability to process information force consumers to make choices regarding the subjects to which they pay more or less attention. In particular, consumers may make two types of mistakes about purchase: (1) Buying a product that they should have not bought and (2) rejecting a product that they should have bought. Processing information reduces the chance of making these mistakes and how consumers allocate attention to various product information determines which type of mistake is reduced more. To be more specific, a product may have several potentially good and bad aspects. Some information such as a negative product review is more concerned with the potential drawbacks of a product, while other information such as a positive product review is more concerned with the potential benefits. On the one hand, paying attention to the first type of information (referred to as "negative information") allows the consumer to realize some serious drawbacks when the product does not fit her,[1] thereby reducing the mistake of buying a mismatched product; on the other hand, paying attention to the second type of information (referred to as "positive information") allows the consumer to find out some good aspects of the product when the product fits her, thereby reducing the mistake of not buying a matched product. However, due to the costs of processing information, the consumer has to trade off these two types of mistakes by allocating attention to positive and negative product information.

For example, suppose a consumer is considering whether or not to buy an electric car. A priori, an electric car may save the consumer a lot of money on gas (potential benefit) but it may be hard for the consumer to charge the car (potential drawback). The consumer needs to process further information to reduce uncertainty about those aspects. However,

[1] In my dissertation, I refer to the consumer as "she" and to the seller as "he".

since processing information is costly for the consumer, she has to decide how much time and efforts to spend on investigating each aspect respectively. In particular, suppose the consumer chooses an information processing strategy under which she spends a lot of time and efforts on investigating the potential benefit but spends little time on the potential drawback. In this case, if the electric car indeed fits the consumer (e.g., it indeed saves the consumer a lot of money on gas and it is not very hard for the consumer to charge the car), then the consumer is very likely to find out that the product would fit her under such information processing strategy. This would increase the consumer's willingness to pay and reduces the mistake of not buying a matched product; but if the electric car does not fit (e.g., the charging station is far away from where the consumer lives and thus charging an electric car is very hard for this consumer), then she may not be able to find out such serious drawback because she pays little attention to it, thereby leading to a high risk of buying a mismatched product. Similarly, suppose the consumer instead chooses an opposite information processing strategy under which she pays a lot of attention to the potential drawback but pays little attention to the potential benefit. Under such strategy, she would instead have a low chance of mistakenly buying an electric car when it does not fit her, but she suffers from a high risk of mistakenly rejecting it when it indeed fits her.

Consider another example in which a consumer is reading the online product reviews of a product. If the consumer pays more attention to positive product reviews than negative ones, she is more likely to make a mistake of buying a mismatched product because she may not be able to find out its drawbacks mentioned in those negative reviews. However, if she instead pays more attention to negative reviews than positive ones, then she is instead more likely to reject a matched product because she may ignore some good aspects of the product.

The examples above show that the consumer's attention allocation decision directly determines the quality of her purchase decision, motivating the consumer to optimize her attention on various product information. To understand how a consumer allocates attention, I develop in Chapter 3 an analytical model based on rational inattention framework (Sims, 2003) in which the decision maker's attention allocation decision is captured by her choice of signal structure and cost of information is measured using information theory (Shannon 1948). To be more specific, I assume that there exist two possible states: either the product fits a consumer (i.e., the utility of the product is above the consumer's reservation price and thus the consumer should buy the product) or it does not fit (i.e., the utility of the product is below the consumer's reservation price and thus she should not buy the product). The consumer does not know the match state ex-ante, but information search generates a signal based on which a consumer can update her belief about product fit, and the consumer's attention allocation decision determines the signal accuracy, referred to as "signal structure", under both "fit" and "not fit" states. In particular, when the consumer pays relatively more attention to positive information than negative information, the signal is relatively more accurate under "fit" state than under "not fit" state, and vice versa. Therefore, the consumer's attention allocation decision is captured by her choice of signal structure, and she chooses the optimal signal structure by trading off the value and cost of information, where the value of a particular signal structure is quantified by the expected utility the consumer can obtain via making decisions based on the signal and the cost of a particular signal structure is quantified using the mutual information metric in information theory (Shannon 1948).

I obtain closed-form solution for optimal attention allocation decision and find that a consumer engages in either "confirmatory" search or "disconfirmatory" search. Under "confirmatory" search the consumer searches more information that favors her prior belief about product fit (e.g., paying more (less) attention to positive information than negative information under a high (low) prior belief), while under "disconfirmatory" search the consumer instead searches more information that disfavors her prior belief (e.g., paying more (less) attention to negative information than positive information under a high (low) prior belief). In particular, the consumer may do more disconfirmatory search when she has a low information processing cost, but she would do more confirmatory search when she has a high cost. Note that the confirmatory search resonates with the "confirmation bias" phenomenon widely studied in the psychology literature (e.g., Nickerson 1998) which experimentally shows that people tend to bias their information search by seeking more evidence that favors their prior beliefs in both judgment and decision making contexts. However, in contrast to the psychological explanations that typically perceive the confirmation bias as an inferential error of human reasoning (e.g., Evans, 1989, p.41), our result suggests that the "confirmatory bias" could be optimal behavior coming out of people optimizing attention to different types of information, especially when people has high cognitive limitations and thus high information processing costs. Moreover, since a consumer's attention allocation pattern may switch from disconfirmatory search to confirmatory search as her information processing cost becomes higher, her purchase likelihood thus may vary with the information processing cost in a non-monotonic way. For example, depending on her prior belief and other factors, the consumer's purchase likelihood may first increase and then decrease with her information processing cost, or it may first decrease and then increase. I will discuss these cases in details later in Chapter 3. Besides, there often exists some uncertainty that is not reducible. For example, some product information may not be available for consumers to process or it is not credible. Therefore, I further study how this irreducible uncertainty influences the consumer's attention allocation. Interestingly, I find that as more information becomes available or credible, if the prior utility of a product is high, then the consumer would increase attention to negative information and thus the purchase likelihood decreases; but if the prior utility is low, then the consumer would increase attention to positive information and the purchase likelihood increases.

Understanding the consumer's attention allocation decision is also important for a seller. For example, a seller may be able to influence the consumer's attention allocation decision by making product information harder or easier for consumers to process. But until a seller understands how a consumer allocates attention, it is not clear a priori that whether increasing or decreasing a consumer's information processing cost would induce the consumer to process relatively more positive information. Furthermore, some of the seller's marketing strategies can also influence the consumer's attention allocation decision. For example, if the seller charges a lower price or offers a return policy that allows the consumer to return mismatched product, then the disutility of buying a mismatched product becomes lower. This may induce the consumer to allocate more attention on the positive information. Given this, will the consumer's flexible attention allocation make the seller charge a higher or lower price? Is it always optimal for a seller to offer the return policy? In addition, a seller may also affect the consumer's attention allocation by choosing how much and which information to be available or unavailable for the consumer to process, which is referred to as "information design". For example, a software company can decide how many and which features are available for consumers to try in its free trial, and a seller may also suppress the negative product reviews on the reviews sites so that the consumers are less likely to find out the potential drawbacks of the product. The questions is: when dose the seller have a high incentive to suppress the negative information and when does he have a low incentive? To answer the questions above about firm strategies, in Chapter 4, I turn attention to study how a seller charges the price, offers return policy and designs the information environment, taking into account the consumer's attention allocation decision.

I obtain several interesting implications of consumer attention allocation on firm strategies. First, in terms of pricing, with low information processing cost, consumers have high incentives to process information about product fit, and if they do process and receive a positive signal, their willingness to pay becomes higher. This induces the seller to charge a high price such that consumers always process information ("learning-promoted" strategy). In particular, a higher price motivates consumers to increase attention to negative information of a product and decrease attention to positive information, and this effect becomes stronger under a higher information processing cost. Therefore, as information processing cost becomes higher, the demand sensitivity to price increases and thus the seller has to charge a lower price. However, with high enough information processing cost, consumers have low incentives to process information, and thus it becomes more profitable for the seller to charge a low price so that consumers always buy the product without costly search and learning ("learning-prevented" strategy). Under this strategy, a higher information processing cost reduces the consumer's incentive to learn and thus allows the seller to charge a higher price without inducing the consumer to process information. Overall, the seller prefers the "learning-promoted" pricing strategy when the consumer has a low information processing cost but prefers the "learning-prevented" strategy when the consumer has a high information processing cost. Therefore, the optimal price and profit first decrease and then increase with information processing cost.

In addition to pricing, a seller often offers return policy which allows the consumer to return mismatched product. Regarding such a return policy, I find that offering return policy would induce consumers to pay more attention to positive information and less attention to negative information, and the seller should offer a return policy except when information processing cost is very high. Specifically, when the information processing cost is high, the seller prefers "learning-prevented" pricing strategy under which the seller discourages the consumer from learning. In this case, offering a return policy further reduces the consumer's learning incentive, which allows the seller to charge a higher price. However, when the information processing cost is very high, this price increasing effect of return policy is small and cannot offset the demand decreasing effect (the demand decreases under return policy because the consumer can always return the mismatched product), and thus offering a return policy in this case would reduce the firm's profit. Overall, the seller prefers offering return policy when the information processing cost is not very high.

Finally, when a seller can design the information environment by choosing how much information and which information to be available for consumers to process, I find that a lower prior belief about product fit hold by the consumer decreases the seller's incentive to suppress the negative information in the environment. Briefly speaking, a lower prior belief motivates the consumer to pay relatively more attention to negative information than positive one, forcing the seller to guarantee enough negative information in the environment. Otherwise, the information search would not be useful for the consumer and she would thus not start to process any information at all, which is often bad for the seller. Besides, I also find that a higher information processing cost would increase or decrease the seller's incentive to suppress negative information, depending on whether the seller can adjust the product price and whether the consumer has a high or low prior belief. Interestingly, the seller may charge a lower price when it can fully design the information environment than when it can not. This is because a lower price can increase the consumer's incentive of processing information, allowing the seller to further suppress the negative information to increase the consumer's purchase likelihood.

The rest of the paper is organized as follows. In Chapter 2, I discuss related literature on consumer search. In Chapter 3, I present the consumer attention allocation model and analyze the optimal attention allocation decision by the consumer. In Chapter 4, I study firm's marketing strategies taking into account the consumer's attention allocation decision. Finally, I make conclusions in Chapter 5.

Chapter 2: Related Literature

2.1 Consumer Search

The search cost and its implications on firm's strategies have been extensively studied in literature. In a large part of this literature, consumers pay a fixed cost to learn: lowest price (e.g., Diamond 1971, Stahl 1989), quality information (e.g., Mayzlin and Shin 2011, Gardete 2013), best-fit alternative (e.g., Bakos 1997, Anderson and Renault 1999, Villas-Boas 2009) or their own preferences (e.g., Guo and Zhang 2012). In these studies, consumers are either assumed to learn perfectly by paying a fixed search cost or assumed to receive an informative signal with exogenous precision.

Specifically, most of the early work in this literature focuses on consumer search for price information and its implications on market. For example, Diamond (1971) provides an extreme example about how consumer search for price influences the market, which is called as the Diamond Paradox . There are two parts to the paradox. On the seller side, as long as the consumers have a positive search cost, all firms set the monopoly price in equilibrium regardless of the number of firms. On the consumer side, consumers do not search. The basic intuition is that if some certain seller sets a price below the monopoly price, then this seller always has an incentive to raise its price slightly by an amount smaller than the search cost, because doing so would not induce any existing consumers of this seller to leave in search of other firms' prices and the seller can get more margin from each existing consumer. This incentive to raise price implies that all firms must charge the monopoly price in equilibrium. Furthermore, consumers have no reason to search because they rationally anticipate that all firms charge the same monopoly price. Therefore, Diamond's model is a search model without search and it shows that even modest search costs can lead to a market outcome that equilibrium prices are strikingly higher than marginal costs, which is in contrast to the traditional economic model such as Bertrand (1883) where consumers are perfectly informed about prices and firms charges a price equal to the marginal cost.

Bridging the Diamond's monopoly pricing and Bertrand's marginal-cost pricing, Stahl (1989) considers a sequential search model in which there are two types of consumers: "shoppers" with zero search costs and "non-shoppers" with positive search costs. If the fraction of "shoppers" tends to be one, almost all consumers in the market have a zero search cost and thus they can easily discover the lowest price in the market, which motivates the firms to undercut price intensively so that equilibrium prices approach the marginal cost (Bertrand, 1883); However, if the fraction of"shoppers" tends to be zero, then almost all consumers in the market have a positive search cost and thus all firms will charge the monopoly price for the reason I discussed earlier (Diamond, 1971). Moreover, Stahl (1989) shows that equilibrium prices increase in the number of firms in the market. The intuition is the following. As more and more competitors come to the market, firms in a mixed pricing strategy equilibrium would have an smaller probability of being the cheapest seller and thus have an incentive to charge higher prices.

Following the studies above about consumer search for price, later work in the consumer search literature starts to investigate the case where the consumer searches for both price and product fit information. This stream of research can be traced back to Bakos (1997) who studies the role of buyer search costs in markets with differentiated product offerings. He shows that when a consumer incurs a cost to search both price and fit, a lower search cost can promote price competition and reduce the market power of sellers. However, if searching for fit has a separate cost from searching for price, then a lower search cost for fit would actually increase the prices and the sellers' profits. This is because with a lower search cost for fit, the consumer can easily find a product with a better fit and thus can easily become captive of the seller who offers a better fit, which increases the seller's market power. Anderson and Renault (1999) also study the implications of consumer's costly search for both price and fit on price competition in a differentiated market. They show that prices may initially decrease with the degree of product differentiation. This is because more product diversity induces the consumer to search more, leading to more competition. But if the product diversity becomes very high, firms would have a higher market power and charge a higher price, and thus prices would increase with the degree of product differentiation.

Aside from pricing, some other work investigates the impact of consumer search on other kinds of firm strategies. For example, Villas-Boas (2009) studies the impact of consumer search on product line length and shows that the optimal number of products offered is decreasing in the evaluation costs. Guo and Zhang (2012) studies optimal product line design when consumers need to incur costly deliberation to uncover their valuations for quality. They show that in order to invite the consumer to deliberate, a seller must maintain quality dispersion and cut the price of the high-end product.

Unlike the above literature which investigates the firm's strategies in response to consumer search, one stream of research focuses on characterizing the specific procedure of consumer search. For instance, Weitzman (1979) characterizes the solution to the problem of searching for the best alternative. The optimal strategy is an elementary reservation price rule, where the reservation prices depends only on the features of that alternative. The selection rule is to search next the unsampled alternative with highest reservation price. The stopping rule is to terminate search whenever the maximum sampled reward is above the reservation price of every unsampled altenative. Branco et al. (2012, 2016) considered a gradual learning model in which consumers incur search costs to learn product attribute information sequentially, and update the expected utility of the product at each search occasion. The optimal stopping rule for the consumer consists of an upper bound ("purchase threshold") and a lowerbound ("exit threshold") on the consumer's expected valuation. In particular, when the consumer's expected valuation of the product hits the purchase threshold, the consumer stops searching and purchases the product. When the expected valuation hits the exit threshold, the consumer stops searching and does not purchase the product. When the expected valuation is in between the two bounds, the consumer continues to search and updates the expected utility of the product. Following Branco et al. (2012, 2016), Ke et al. (2016) further characterizes the consumer search for information on multiple products and the optimal search strategy is characterized by an optimal consideration set and a purchase threshold structure. Specifically, a product is only considered for search or purchase if it has a sufficiently high expected utility. Given multiple products in the consumer 's consideration set, the consumer only stops searching for information and purchases a product if the difference between the expected utilities of the top two products is greater than some threshold.

Although the literature above recognizes the fact that people often needs to incur nontrivial costs to process information, it does not investigate how people selectively pays attention to various product information. As mentioned in Chapter 1, consumers may spend more time and efforts on investigating the potentially good aspects of a product or spend more time and efforts on investigating the potentially bad aspects (e.g., they may pay more or less attention to positive customer reviews than negative reviews.). Such attention allocation decision can be modeled by the rational inattention framework proposed by Sims (1998, 2006), which is discussed in the following section.

2.2 Rational Inattention

There is a vast amount of information that is in principle available to decision-makers (e.g., consumers can easily obtain various product information through internet such as the online customer reviews). However, due to limited attention, it is impossible for people to attention to all of the information. Therefore, the idea of rational inattention is that decision-makers have to decide how to optimally allocate their limited attention, specifically, which information to attend to carefully, which information to pay less attention to, and which information to ignore.

To implement the idea of rational inattention, Sims (1998, 2003, 2006) proposes to model attention as an information flow and limited attention is formulated as an upper bound on information flow. In the following, I introduce the general framework of rational inattention problem proposed by Sims (1998, 2003, 2006). Generally speaking, a decision maker needs to choose an action a from a discrete or continuous choice set A and she has a prior belief G about the state of the world denoted by $x \epsilon X$ which leads to the payoff $v(x, a)$ if the decision maker takes the action a. The decision maker observes a signal S on the state, then updates her belief about the state based on the signal realization $s \epsilon S$, and finally chooses an action as a Bayesian expected utility maximizer. In particular, before receiving the signal, the decision maker is able to choose an information-processing strategy which determines the joint distribution of the signal and the state, $f(x, s)$ (equivalently, given some prior belief distribution $G(X)$, an information-processing strategy determines the conditional distribution $f(x|s)$ or $f(s|x)$). To be consistent with the following chapters, I consider here the discrete distribution case and thus the rational inattention problem can be formulated as follows:

$$\max_{f(x,s)} \sum_x \sum_s V^*(s) f(x, s)$$

$$\begin{aligned} s.t. \sum_s f(x, s) &= G(x) \quad all\ x \\ I(X, S) = \underbrace{-\sum_x G(x) \log(G(x))}_{H(X)} - \underbrace{E_s(-\sum_s f(x|s) \log(f(x|s)))}_{H(X|S)} &\leq \kappa, \end{aligned} \tag{1}$$

where $V^*(s) = \max_a \sum_x v(x, a) f(x|s)$ is the maximum utility the consumer can obtain given the decision maker receives a signal realization $S = s$, and κ is the maximum information flow rate between X and S (the lower κ is, the less attention the decision maker has.). The unique feature of rational inattention framework above is that it uses the information theory (Shannon, 1948) to measure the rate of information flow as shown in (1). Therefore, I briefly introduce information theory in the following and discuss its application in the context of attention allocation. Before that, note that a closely related formulation (e.g., Matějka and

McKay, 2015) assumes that capacity is variable, at a cost. That is, the decision maker solves the following problem:

$$\max_{f(x,s)} \sum_x \sum_s V^*(s) f(x,s) - \lambda I(X,S) \tag{2}$$

$$s.t. \sum_s f(x,s) \quad = \quad G(x) \quad all\ x,$$

where λ is the unit cost of processing one piece of information. This alternative formulation implies that the decision maker can not only choose how to allocate his or her attention but also choose how much information to process. In Chapter 3 and 4, I will use this formulation to model the consumer's attention allocation.

Information Theory

The basic idea of information theory is to measure the rate of information flow as the rate of uncertainty reduction. It therefore starts with a measure of uncertainty, called entropy. Specifically, the term $H(X) \equiv -\sum_x G(x)\log(G(x))$ is called the entropy of the random variable X and is a measure of the uncertainty about the state of X based on the prior belief $G(X)$. The definition of entropy as a measure of uncertainty can be basically derived from two intuitive axioms. One can perceive entropy as the average amount of 'surprise' associated with a set of probable events. Based on this, the two intuitive axioms can be described as follows: (1) the less probable an event is (i.e., $G(x)$ is small), the more surprising when it occurs; (2) The amount of surprise of seeing two independent events simultaneously should be the sum of the amounts of surprise of seeing each event separately. These two axioms imply that the surprise of an event is proportional to $\log p(x)$, with the proportionality constant determined by the base of logarithms. [2] Now, averaging over all events according to their respective probabilities, one can get the expression for $H(X)$. Next, the term $H(X|S)$ is called conditional entropy of X given signal S and is a measure of the uncertainty about the state of X given the information from signal S. Information reduces uncertainty, and the mutual information of X and S, denoted by $I(X,S) \equiv H(X) - H(X|S)$, is a measure of the reduction of uncertainty about the true state of X due to receiving the signal S and it quantifies the amount of information contained in a signal S (Shannon, 1948). Note that one can show the following relationship:

$$I(X,S) \equiv H(X) - H(X|S) = H(S) - H(S|X), \tag{3}$$

where $H(S) \equiv -\sum_s p(s)\log(p(s))$ and $H(S|X) \equiv -\sum_x p(x) \sum_s p(s|x)\log(p(s|x))$.

The measure $I(X,S)$ has some appealing properties. First, it is zero if and only if X and S are independent, and it is always non-negative. Second, given a sequence of observations, say on S_1 and on S_2, we would like the information about X in seeing S_1, then S_2 to be the same as that in seeing S_2, then S_1. That is, $I(X,S_1)+I(X,S_2|S_1) = I(X,S_2)+I(X,S_1|S_2)$. It turns out that these simple properties are restrictive enough to leave us with only the Shannon measure of mutual information.

Besides these intuitively appealing properties, the popularity of using mutual information as the measure of information is justified by its central position in communications engineering. In particular, information is thought of as moving through a "channel," in which one enters input data, and output data emerges, possibly error ridden, and the Shannon mutual information of two random variables (e.g., the input data and output data) is equivalent to the expected length in bits of the code needed to generate one from the other. Consider a simple example of drawing a 0 or 1 from a distribution with equal probability on the two

[2] The usual base for log is 2, which implies that the unit of information is "bit," while, in the following chapters, the base used is e, in which case the unit is called "nat." Using a different log base does not qualitatively change our results.

values and sending it through a device such as a telegraph key. Suppose this device can generate the output data from the input data without error, then we can eliminate all uncertainty in the value of the draw by transmitting one bit of information (i.e., "0" or "1"). Note that in this example, the Shannon mutual information $I(X, S) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} - 0 = 1$ (bit).

In addition to its central position in communication engineering, using mutual information as information cost function can also be justified from the perspective of sequential information sampling. Specifically, it is often quite realistic to assume that information is acquired through a sequential sampling process (e,g" Wald 1945). That is, a decision maker sequentially observes signals at a cost and dynamically decides when to stop acquiring information. However, another way to model information acquisition is the static rational inattention approach discussed above, where the information cost is simply a function of the decision maker's prior and posterior belief measured by mutual information. Interestingly, Morris and Strack (2017) show the equivalence of sequential sampling approach and static approach. In particular, they show that mutual information cost function corresponds to a flow cost (in the sequential sampling problem) which vanishes when the agent is close to certain about the state. Furthermore, Hebert and Woodford (2017) find that for a particular family of flow information-cost functions in the sequential sampling problem, the cost function for the equivalent static model is just the mutual information, which provides foundations for rational inattention framework proposed by Sims (1998, 2003, 2006), that do not rely on any analogy in communications engineering.

In terms of empirical support of using Shannon mutual information as information processing cost, a number of recent empirical studies show that using Shannon's mutual information to model the costs of processing information about different choices fits and predicts observed choice data well (e.g., Cheremukhin et al. 2011, Dewan and Neligh 2017). For instance, Cheremukhin et al. (2011) use data from a behavioral experiment to show that people behave according to predictions of rational inattention theory: (1) They behave probabilistically and (2) pay less attention and are thus more prone to error when differences between choice options are smaller. More interestingly, Dewan and Neligh (2017) determine using data from lab experiments that among a variety of cost functions for information acquisition used in theoretical studies (specifically, fixed cost, normal signals with linear precision costs, and the mutual information cost function), the mutual information cost function is the best fit for a large majority (over two-thirds) of their subjects. However, note that although rational inattention theory find some empirical support, some studies also point out that decision makers do not always behave in line with what rational inattention theory predicts and thus some generalization of Shannon's mutual information is proposed and is found to improve the fit of the subject behavior in experiments (e.g., Caplin and Dean, 2013 and Woodford, 2014). In neuro-economics, the entropy function from information theory is used to quantify the amount of information in a neural population (and mutual information is used as a metric of reduction of uncertainty to link stimuli and neural responses; Quiroga and Panzeri 2009, Rolls and Treves 2011.).

Application of Rational Inattention

Rational inattention framework above has been applied to a variety of economic contexts such as consumption-savings problems (e.g., Sims 2006, Luo 2008, Maćkowiak and Wiederholt 2015), rigid pricing (e.g., Maćkowiak and Wiederholt 2009, Matějka 2016), investment decisions (Mondria 2010, Yang 2011), and discrete choice (Matějka and McKay 2015). Specifically, the rational inattention theory is initially applied to explain the inertial reaction of economic agents to external economic information (e.g., Sims 2003, 2006, Luo 2008 and Maćkowiak and Wiederholt 2015). In particular, Maćkowiak and Wiederholt (2015) finds that since aggregate technology shocks are much larger than monetary policy shocks, decision-makers in firms choose to pay more attention to aggregate technology shocks than to monetary policy shocks, and as a result, prices set by firms respond faster to aggregate technology shocks than to monetary policy shocks. Interestingly, their model matches the faster response of the price level to aggregate technology shocks not only qualitatively, but also quantitatively. Similarly, Maćkowiak and Wiederholt (2009) present a rational inattention model in which price setting firms decide what to pay attention to, subject to a constraint on information flow. They find that when idiosyncratic conditions are more variable or more important than aggregate conditions, firms pay more attention to idiosyncratic shocks than to aggregate shocks, and as a result, price responses to idiosyncratic shocks are strong and quick whereas price responses to aggregate shocks are dampened and delayed. Matějka (2016) finds that since misjudging the input cost when it is low is more costly to the seller than when it is high, so the seller pays more attention to shocks leading to low costs, which then implies more flexible low prices and sales-like movements. More interestingly, even when the input cost is continuously distributed, a rationally inattentive seller would choose to price discretely, i.e., he sets up a price plan consisting of a few prices and charges only one of them, in order to economize on his information capacity. This implies that prices are likely to stay fixed when cost shocks are small.

When people are making investment decisions, they also need to allocate attention to various information. For example, Mondria (2010) studies the attention allocation of portfolio investors who choose the composition of their information subject to an information flow constraint. In equilibrium investors choose to observe one linear combination of asset payoffs as a private signal, and as a result, changes in one asset affect both asset prices, leading to comovement of asset prices and the transmission of volatility shocks between two assets. Yang (2011) shows that if players' investment decisions are strategic complementarity, each player is willing to pay attention to information that her opponent pays attention to, which enables players to acquire information that makes efficient coordination possible.

Individuals must often choose among discrete actions with imperfect information about their payoffs. Before choosing, they have an opportunity to incur costs to learn about the payoffs and thus in order to make a good choice, they need to decide how much information to process and which information to pay more attention to and which information to pay less attention to. Interestingly, Matějka and McKay (2015) show that a rationally inattentive decision maker would choose an alternative probabilistically in line with a generalized multinomial logit model, which depends both on the actions' true payoffs as well as on prior belief.

In my dissertation, I apply the rational inattention framework to the context in which the consumer searches product match information and then makes a purchase decision. In particular, I examine how the consumer optimally allocates attention given her prior belief, information processing cost and the credibility of information, and how a seller chooses the marketing strategies taking into account the consumer's attention allocation decision.

Chapter 3: Consumer Attention Allocation

3.1 Model

3.1.1 Attention Allocation as the Choice of Signal Structure

Consider a consumer who wants to determine whether or not she should buy the product, i.e., whether a product matches her needs or not. Denote $X \in \{0, 1\}$ as the binary state for match. If $X = 1$, then the product fits and the consumer obtains utility U_1 from purchasing it (net of price); if $X = 0$, the product does not fit and the consumer obtains utility U_0 from purchasing it (net of price). I assume that the utility of the outside option is 0 and $U_0 < 0 < U_1$. The consumer does not know ex ante the value of state X but has a prior belief about it denoted by $q \equiv P(X = 1)$. I assume that the consumer would know the value of state X after purchasing the product.

By processing information (e.g., inspecting the product, reading product descriptions and reviews, talking to friends, etc.), the consumer receives a signal $S \in \{0, 1\}$ based on which the consumer can update her belief about the match state X. Specifically, if $S = 1$, then the consumer receives a "positive" signal increasing her belief that the product fits (i.e., $P(X = 1|S = 1) > q$, where $P(X = 1|S = 1)$ is the conditional probability that true state X is 1 given the signal S is 1); if $S = 0$, then the consumer receives a "negative" signal decreasing her belief about product fit (i.e., $P(X = 1|S = 0) < q$, where $P(X = 1|S = 0)$ is the conditional probability that true state X is 1 given the signal S is 0). Note that one can interpret the signal S as the aggregation of all the pieces of information processed by a consumer. To be more specific, a consumer may search both some pieces of positive information (e.g., positive product reviews) and some pieces of negative information (e.g., negative reviews) during the whole information search, and if the aggregation of all such processed information increases her belief about product fit, then I say that the consumer receives a positive signal (i.e., $S = 1$); conversely, if the consumer's belief decreases, then I say that she receives a negative signal (i.e., $S = 0$).

Unlike the traditional consumer search literature discussed before, I allow the consumer to flexibly choose the signal accuracy (referred to as signal structure) $\delta_1 \equiv P(S = 1|X = 1)$ and $\delta_0 \equiv P(S = 0|X = 0)$, where δ_1 is the probability of receiving a positive signal given the product indeed fits (i.e., $X = 1$) and δ_0 is the probability of receiving a negative signal given the product does not fit (i.e., $X = 0$). Note that the consumer's attention allocation decision determines the signal accuracy under both "fit" and "not fit" states. Consider again the customer reviews example. If the consumer spends a lot of time and efforts on reading the content of positive reviews but spends little time on negative ones, then she is very likely to find out some good product features that matter to her when the product indeed fits but she may not be able to find out some serious drawbacks when the product does not fit. Therefore, under such attention allocation strategy, the signal is relatively more accurate when the product indeed fits than when it does not fit (i.e., $\delta_1 > \delta_0$). Conversely, when the consumer pays more attention to the negative reviews than positive ones, then the signal would be relatively more accurate when the product does not fit than when it does (i.e., $\delta_0 > \delta_1$). Therefore, I formulate the consumer's attention allocation decision as her choice of the signal structure δ_1 and δ_0. Without loss of generality, I say that the consumer chooses $\delta_1 > \delta_0$ if she pays more attention to positive information than negative information and chooses $\delta_1 < \delta_0$ otherwise. Note that although I do not investigate how the consumer search information step by step, the consumer's choice of δ_1 (i.e., $P(S = 1|X = 1)$) and δ_0 (i.e., $P(S = 0|X = 0)$) uniquely determines her posterior beliefs $P(X = 1|S = 1)$ and $P(X = 1|S = 0)$, which further determines a stopping rule when the consumer searches information sequentially. Specifically, as I will discuss later, the consumer would always buy the product when receiving $S = 1$ and would choose her outside option when receiving $S = 0$. Therefore, when the consumer's posterior belief about product fit is above $P(X = 1|S = 1)$, she would stop searching and buy the product. When it is below $P(X = 1|S = 0)$, she would instead stop searching and choose her outside option. When it is between $P(X = 1|S = 1)$and $P(X = 1|S = 0)$, the consumer would continue searching and keep updating her belief after processing each piece of information. As I mentioned earlier, under some conditions, the static rational inattention approach is equivalent to the sequential sampling approach (e.g., Hebert and Woodford 2017, Morris and Strack 2017)

Now I proceed to answer the following question: How would a consumer choose the optimal signal structure? Intuitively, if the signal has high signal precision, then the consumer can make a better decision based on the signal but more information is needed to be processed, thus leading to higher costs of processing information. Therefore, the consumer chooses the optimal signal structure by trading off the value and cost of processing information. In the following, I first quantify the value of information, then quantify the cost of information, and finally formulate the consumer's optimal attention allocation problem.

3.1.2 Value of Information

Since by definition a positive signal increases a consumer's belief about product fit and a negative signal decreases it, the signal structure must satisfy $\delta_1 + \delta_0 > 1$ so that $P(X = 1|S = 1) > q > P(X = 1|S = 0)$. Otherwise, if $\delta_1 + \delta_0 = 1$, then the posterior belief is always equal to the prior belief, i.e., $P(X = 1|S = 1) = q = P(X = 1|S = 0)$, which implies that the signal is uninformative. If $\delta_1 + \delta_0 < 1$, then $P(X = 1|S = 1) < q < P(X = 1|S = 0)$, which contradicts with the definitions of positive and negative signal. Therefore, the signal received by the consumer must satisfy $\delta_1 + \delta_0 > 1$. I assume if the consumer does not process information, then there is no signal to receive and the consumer's belief does not change.

Furthermore, note that different signal realizations must induce the consumer to take different actions (i.e., the consumer purchases the product when receiving a positive signal and chooses the outside option when receiving a negative signal). Otherwise, if both positive and negative signal lead to the same action, then it must be always better off for the consumer to directly choose that action without incurring any cost to process information.

Accordingly, the value of information can be quantified by the expected utility EV that the consumer can obtain by making decisions based on different signal realizations, as shown in (4).

$$EV \equiv P(S=1)(P(X=1|S=1)U_1 + P(X=0|S=1)U_0) \tag{4}$$

Intuitively, with probability $P(S=1)$, the consumer's information search generates a positive signal, inducing her to purchase the product of which the posterior utility is given by $P(X=1|S=1)U_1 + P(X=0|S=1)U_0$ (with probability $P(S=0)$, the consumer's information search generates a negative signal, inducing her to choose the outside option with utility 0). EV increases with both δ_1 and δ_0 (i.e., $\frac{dEV}{d\delta_1} = qU_1 > 0$ and $\frac{dEV}{d\delta_0} = (1-q)(-U_0) > 0$), implying that the expected utility increases as the signal becomes more accurate. Note that if the consumer processes no information, then the consumer evaluates the product based on her prior belief and purchases it if and only if $qU_1 + (1-q)U_0 \geq 0$.

3.1.3 Cost of Information

Since a particular information search strategy generates a particular signal for the consumer, it is natural to investigate how much information is needed to be processed to obtain such a signal. As discussed in Chapter 2, Shannon's information theory (Shannon 1948, Cover and Thomas 2006) offers a micro-founded metric, called mutual information, that quantifies the total amount of information needed to be processed to obtain a particular signal. According to (3), the amount of information contained in the binary signal in our model is given by

$$
\begin{aligned}
I(X,S) = & -[q\delta_1 + (1-q)(1-\delta_0)]\log[q\delta_1 + (1-q)(1-\delta_0)] \\
& -[q(1-\delta_1) + (1-q)\delta_0]\log[q(1-\delta_1) + (1-q)\delta_0] \\
& +q[\delta_1 \log \delta_1 + (1-\delta_1)\log(1-\delta_1)] \\
& +(1-q)[\delta_0 \log \delta_0 + (1-\delta_0)\log(1-\delta_0)].
\end{aligned} \tag{5}
$$

Given this metric, the cost of obtaining a signal S (i.e., the cost of processing information to change beliefs about states from $p(X)$ to $p(X|S)$) can be measured by $\lambda I(X,S)$, where $\lambda > 0$ is the unit cost of processing one piece of information. If she has high limitations on time and cognitive resources, then the unit cost λ is high; otherwise, it is low.[3] Note that this cost function $\lambda I(X,S)$ is a convex function of δ_1 and δ_0 (i.e., $\lambda \frac{d^2 I}{d\delta_i^2} > 0$, $i \in \{0,1\}$). Furthermore, if the consumer increases attention to one type of information, then the marginal cost of paying attention to the other type of information also increases (i.e., $\lambda \frac{d^2 I}{d\delta_i d\delta_j} > 0$, $i,j \in \{0,1\}$). Finally, if the consumer pays much more attention to one type of information than the other type of information, then it is marginally cheaper to increase attention to the other type of information (i.e., there exists $t \geq 1$ such that if $\delta_i > t\delta_{1-i}$ then $\lambda \frac{dI}{d\delta_i} > \lambda \frac{dI}{d\delta_{1-i}}, i \in \{0,1\}$).

Ignoring some axiomatic properties, one can also use other meaningful metrics to quantify the cost of processing information. For example, processing information also reduces on average the variance of belief distribution. So the amount of information contained in a signal S can also be intuitively quantified as the reduction of variance: $P(X=1)(1-P(X=$

[3] To be more precise, $I(X,S)$ is the minimum information that needs to be processed to obtain the signal S with the given signal structure. In other words, this corresponds to the most efficient processing case. However, consumers may not process information in the most efficient possible manner, perhaps because she is not familiar with the product category. Suppose that when the minimum information needed to be processed is I, the information that consumers actually process is $\xi I, \xi > 1$. Suppose that the cost of processing one unit of information is $\overline{\lambda}$. In this case, the information processing cost becomes $\left(\overline{\lambda}\xi\right) I$. I can then rescale $\overline{\lambda}\xi$ to λ and thus the unit cost becomes larger if the consumer does not process information in the most efficient manner.

1)) $-$ $\{P(S=1)P(X=1|S=1)(1-P(X=1|S=1))+P(S=0)P(X=1|S=0)(1-P(X=1|S=0))\}$, where the first term$P(X=1)(1-P(X=1))$ is the variance of prior belief distribution and the second term is the average variance of posterior belief distribution. Using this variance-based metric does not lead to closed-form solution but numerical analysis shows that it would not change our main results qualitatively in the following chapters.

3.1.4 Optimization Problem for Attention Allocation

The consumer trades off the value and cost of information. She may choose to not process information at all. Alternatively, if she chooses to process information, then the signal received by her must satisfy $\delta_1 + \delta_0 > 1$ (as discussed earlier) and she maximizes the net expected utility of learning as follows:

$$
\begin{aligned}
& EU^* \equiv \sup_{\delta_1,\delta_0} \underbrace{P(S=1)(P(X=1|S=1)U_1 + P(X=0|S=1)U_0)}_{value\, of\, information} - \underbrace{\lambda I(X;S)}_{cost\, of\, information} \\
& \text{s.t. } \delta_1 + \delta_0 > 1 \\
\implies & EU^* \equiv \sup_{\delta_1,\delta_0} \underbrace{q\delta_1 U_1 + (1-q)(1-\delta_0)U_0}_{value\, of\, information} - \underbrace{\lambda I(X;S)}_{cost\, of\, information} \qquad (6) \\
& \text{s.t. } \delta_1 + \delta_0 > 1
\end{aligned}
$$

$I(X,S)$ is given in (5). In the following, I solve for the optimal signal structure chosen by the consumer and analyze its implications on consumer's information search and purchase behavior.

3.2 Analysis: Optimal Attention Allocation

Define $k = \frac{-U_0}{\lambda}, l = \frac{U_1-U_0}{\lambda}, \underline{q} = \frac{1-e^k}{1-e^l}$ and $\overline{q} = \frac{1-e^{-k}}{1-e^{-l}}$ (where $\underline{q} \leq \overline{q}$). The optimal signal structure is given in the following proposition.

Proposition 1 *The following holds:*

1. *If $q \leq \underline{q}$, the consumer processes no information and does not buy the product.*

2. *For $\underline{q} < q < \overline{q}$, δ_1^* and δ_0^* are given as follows:*

$$\delta_1^* = \frac{1 - \frac{1-q}{q}\frac{e^k - 1}{e^l - e^k}}{1 - e^{-l}}, \ \delta_0^* = \frac{1 - \frac{q}{1-q}\frac{e^l - e^k}{e^l\left(e^k - 1\right)}}{1 - e^{-l}}. \tag{7}$$

   The purchase probability $P(S = 1) = \frac{q}{1-e^{-k}} + \frac{1-q}{1-e^{l-k}}$.

3. *If $q \geq \overline{q}$, the consumer processes no information and buys the product.*

Proof: See appendix.

When the prior belief is either very high (i.e., $q \geq \overline{q}$) or very low (i.e., $q \leq \underline{q}$), the consumer has no incentive to process information as changing her beliefs to an extent that her purchase decision changes would require processing a large amount of information, which is too costly. Rather, the consumer makes her purchase decisions based on her prior beliefs — if this belief is high enough then she purchases and if it is low enough then she does not purchase. Only when the prior belief is neither too high nor too low, the consumer chooses to process information to update her belief about product fit, and her optimal information search strategy generates a signal S of which the structure is characterized by δ_1^* and δ_0^* given in (7).

To understand the consumer's optimal attention allocation decision, note that a consumer who chooses to process product information not only wants to avoid falsely buying a bad/mismatched product but also wants to avoid falsely rejecting a good/matched product. All else equal, when the consumer has a higher prior belief about product fit (i.e., q is higher) or when the loss of buying a bad product becomes relatively smaller as compared to the gain of buying a good product (i.e., U_1 or U_0 increases), the motivation to avoid falsely rejecting a good/matched product becomes higher. Therefore, the consumer would

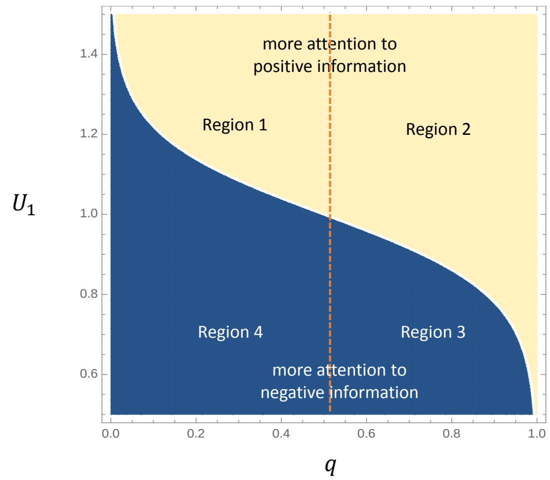


Figure 1: Confirmatory search vs. disconfirmatory search ($\lambda = 0.05$, $U_0 = -1$)

like to pay relatively more attention to positive information and less attention to negative information so that the signal becomes more accurate when the product indeed fits (and the consumer would make less mistakes in this case). Conversely, when the consumer has a lower prior belief or when the loss becomes larger, then the consumer would like to pay more attention to negative information to avoid falsely buying a bad/mismatched product. In short, δ_1^* increases and δ_0^* decreases in q, U_1 and U_0.

Furthermore, I characterize the consumer's attention allocation patterns into two categories: Confirmatory search and disconfirmatory search. Under confirmatory search, the consumer pays more attention to the type of information that favors her prior belief (i.e., $\delta_1^* > \delta_0^*$ if $q > 0.5$ or $\delta_1^* < \delta_0^*$ if $q < 0.5$), as shown in the region 2 and 4 of Figure 1; In contrast, under disconfirmatory search, the consumer pays more attention to the type of information that disfavors her prior belief (i.e., $\delta_1^* < \delta_0^*$ if $q > 0.5$ or $\delta_1^* > \delta_0^*$ if $q < 0.5$), as shown in the region 1 and 3 of Figure 1. Note that the confirmatory search discussed above resonates with the "confirmatory bias" phenomenon widely studied in the psychology literature (e.g., Nickerson 1998) which experimentally show that people tend to seek more evidence that favors their prior beliefs. However, in contrast to the classic psychological explanations that often perceive the confirmatory bias as an inferential error of human rea-

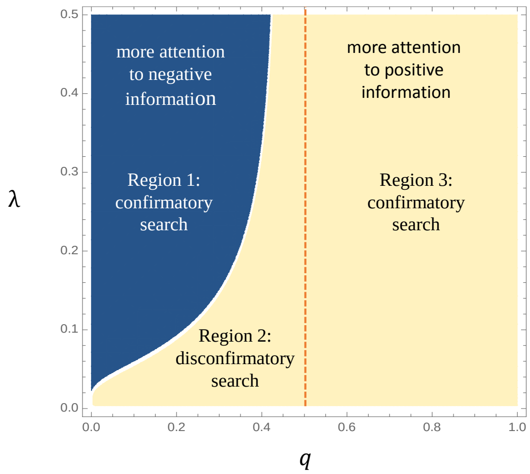


Figure 2: Higher cost leads to more confirmatory search($U_1 = 1.25$, $U_0 = -1$)

soning (e.g., Evans, 1989, p.41), the results in Figure 1 suggest that under some situations the "confirmatory bias" could be an optimal behavior coming out of people optimizing attention to different types of information, while disconfirmatory search would be optimal under other situations.

The following question is: When would the confirmatory search be optimal and when would the disconfirmatory search be optimal? Interestingly, the consumer's information processing cost plays a key role in determining whether the consumer would do confirmatory search or disconfirmatory search. As shown in Figure 2, the consumer would conduct more disconfirmatory search when the information processing cost is low, but she would conduct more confirmatory search when the information processing cost becomes high. More specifically, consider the case where the consumer has a high prior belief but the loss of buying a bad product is larger than the gain of buying a good product (i.e., $U_1 < |U_0|$), as shown in Figure 3a. In this case, when the information processing cost is low, the consumer would pay high attention to both types of information, and since the loss is larger than the gain, the consumer would pay relatively more attention to negative information than positive information in order to reduce the risk of buying a bad product. Therefore, the consumer conducts the disconfirmatory search that disfavors her high prior belief. However, when the information processing cost becomes high enough, the consumer's attention allocation is mainly influenced by her prior belief and thus she would engage in confirmatory search in which she spends more time and efforts on processing positive information than negative information so that the signal received by the consumer is relatively more accurate in the "fit" situation which the consumer believes is more likely to happen than "not fit" situation. This implies that the "confirmatory bias" in psychology literature could be an optimal behavior especially when the consumer has high cognitive limitations and thus high information processing cost. Similarly, one can explain other cases in Figure 3. Note that when the gain is large and prior belief is high (Figure 3c), or when the loss is large and prior belief is low (Figure 3d), the consumer would always do the confirmatory search, with the belief-confirming effect increasing with the cost (the difference between δ_1 and δ_0 always increases in the information processing cost). I summarize the analysis above in the following proposition.

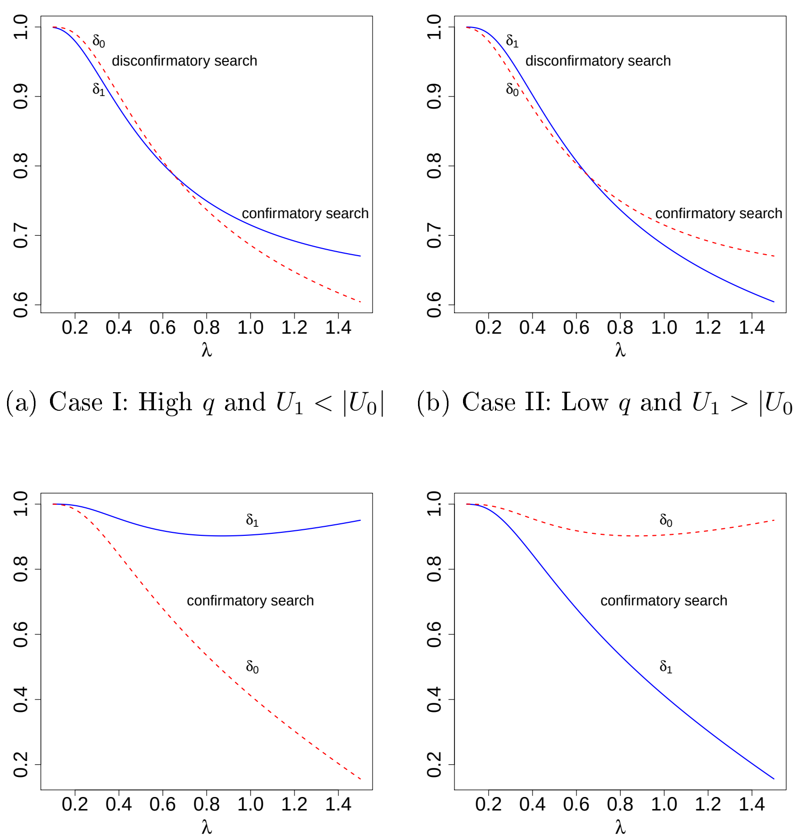
Proposition 2 *Denote λ^* as the unique solution of equation $\frac{1-q}{q}(e^{\frac{-U_0}{\lambda}} - 1)(1 - e^{-\frac{-U_0}{\lambda}}) = \frac{q}{1-q}(e^{\frac{U_1}{\lambda}} - 1)(1 - e^{-\frac{U_1}{\lambda}})$. The following holds:*

1. *If $U_1 > |U_0|$, when $q > \frac{-U_0}{U_1-U_0}$, $\delta_1^* > \delta_0^*$; when $q < \frac{-U_0}{U_1-U_0}$, $\delta_1^* > \delta_0^*$ for $\lambda < \lambda^*$ and $\delta_1^* < \delta_0^*$ for $\lambda > \lambda^*$.*

2. *If $U_1 < |U_0|$, when $q < \frac{-U_0}{U_1-U_0}$, $\delta_1^* < \delta_0^*$; when $q > \frac{-U_0}{U_1-U_0}$, $\delta_1^* < \delta_0^*$ for $\lambda < \lambda^*$ and $\delta_1^* > \delta_0^*$ for $\lambda > \lambda^*$.*

Proof: See appendix.

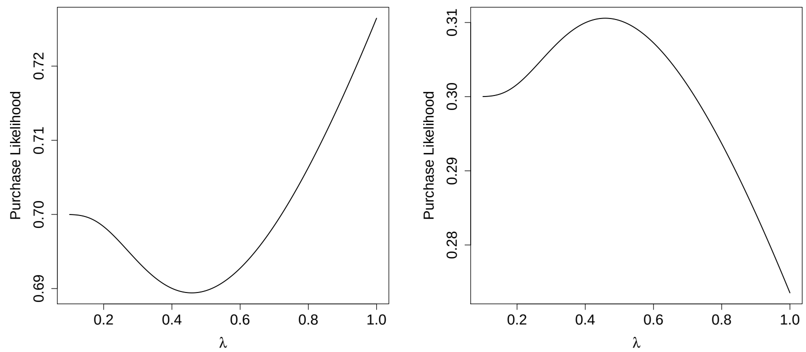
Because of the impact of information cost on attention allocation, the consumer's purchase likelihood may vary with the cost in a non-monotonic way. Consider again the case where the consumer has a high prior belief but the loss of buying a bad product is large (Figure 4a). As discussed earlier, when the information processing cost becomes higher, the consumer would first conduct disconfirmatory search with more attention paid to the



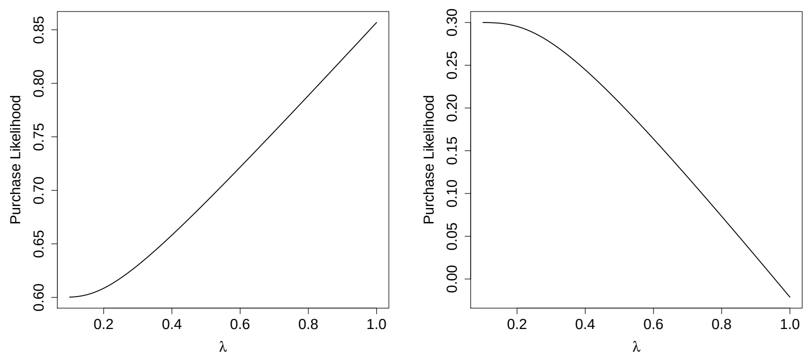
(a) Case I: High q and $U_1 < |U_0|$ (b) Case II: Low q and $U_1 > |U_0|$

(c) Case III: High q and $U_1 > |U_0|$ (d) Case IV: Low q and $U_1 < |U_0|$

Figure 3: Impact of information processing cost on attention allocation



(a) Case I: High q and $U_1 < |U_0|$ (b) Case II: Low q and $U_1 > |U_0|$



(c) Case III: High q and $U_1 > |U_0|$ (d) Case IV: Low q and $U_1 < |U_0|$

Figure 4: Impact of information processing cost on purchase likelihood

negative information and then conduct the confirmatory search with more attention paid to positive information. Therefore, the consumer's purchase likelihood first decreases and then increases with the consumer's information processing cost. Similarly, one can explain other cases in Figure 4. Note that the non-monotonic impact of information processing cost comes out of consumer's attention allocation decision, in contrast to the monotonic results typically found in previous studies on consumer search without attention allocation (e.g., Branco et al. (2012) show that the consumer's purchase likelihood either always increase or always decrease with information processing cost.). Proposition 3 summarizes the impact of information processing cost on purchase likelihood.

Proposition 3 *Denote $\hat{\lambda}$ as the unique solution of equation* $q\frac{e^{\frac{U_0}{\lambda}}\frac{-U_0}{\lambda}}{(e^{\frac{U_0}{\lambda}}-1)^2} = (1-q)\frac{e^{\frac{U_1}{\lambda}}\frac{U_1}{\lambda}}{(e^{\frac{U_1}{\lambda}}-1)^2}$.

The following holds:

1. *If $U_1 > |U_0|$, when $q > \frac{-U_0}{U_1 - U_0}$, the purchase likelihood always increases with information processing cost (i.e., $\frac{dP(S=1)}{d\lambda} > 0$); when $q < \frac{-U_0}{U_1 - U_0}$, the purchase likelihood increases with information processing cost for $\lambda < \hat{\lambda}$, and it decreases with the cost for $\lambda > \hat{\lambda}$.*

2. *If $U_1 < |U_0|$, when $q < \frac{-U_0}{U_1 - U_0}$, the purchase likelihood always decreases with information processing cost (i.e., $\frac{dP(S=1)}{d\lambda} < 0$); when $q > \frac{-U_0}{U_1 - U_0}$, the purchase likelihood decreases with information processing cost for $\lambda < \hat{\lambda}$, and it increases with the cost for $\lambda > \hat{\lambda}$.*

Proof: See appendix.

Comparison with Symmetric Signal Case

The most significant feature of the model above is that the consumer is allowed to flexibly allocate attention to various product information, which is in contrast to the traditional consumer search model ignoring such feature. In this section, I examine whether and how this flexible attention allocation may lead to different implications from those obtained in traditional search model.

Specifically, consider the case where the consumer is not allowed to do the differential learning, i.e., the consumer chooses the signal precision of a symmetric signal. That is, I assume $P(X = 1|S = 1) = P(X = 0|S = 0)$ and thus the consumer's information processing decision is characterized by choosing only one conditional probability: $\delta \equiv P(X = S)$. The higher the signal precision (i.e., $\delta \equiv P(X = S)$) is, the more information is processed. By trading off the value and cost of information, the consumer chooses the optimal signal precision δ^* that maximizes the net expected utility as follows:

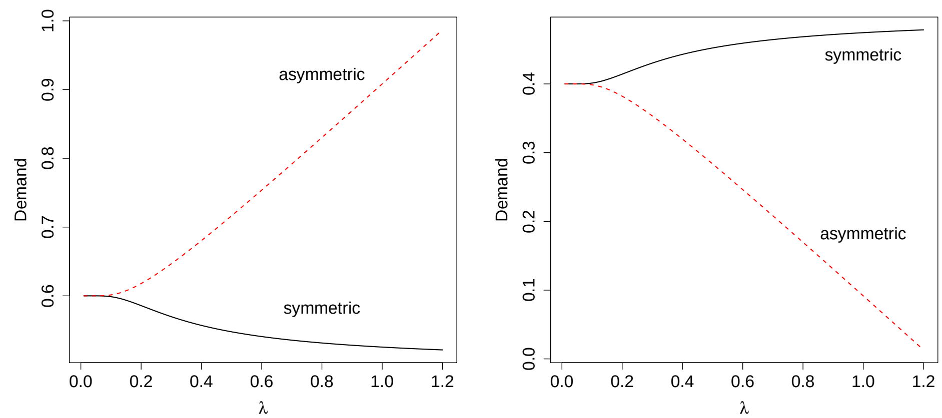
$$EU_S^* \equiv \sup_{\frac{1}{2} < \delta \leq 1} P(S = 1)[P(X = 1|S = 1)U_1 + P(X = 0|S = 1)U_0] - \lambda I(X, S)$$

where $I(X,S) = -[q\delta + (1-q)(1-\delta)] \log[q\delta + (1-q)(1-\delta)] - [1 - q\delta - (1-q)(1-\delta)] \log[1 - q\delta - (1-q)(1-\delta)] + \delta \log \delta + (1-\delta) \log(1-\delta)$.

Figure 5 compares the impact of information processing cost on consumer's purchase likelihood under the asymmetric attention model (main model) and symmetric attention model (benchmark). Clearly, the implications are strikingly different. Specifically, consider the case where the consumer has a high prior belief that the product would fit her. As information processing cost becomes higher, if she is able to flexibly allocate attention, she would process less information but would allocate relatively more attention to positive information than negative information (belief-confirming effect), and thus the purchase likelihood would increase in consumer's information processing cost; however, if she is not able to do the differential learning, higher information processing cost makes the consumer process less information including the positive information and thus the purchase likelihood would instead decrease in information processing cost. Similarly, in the case where the consumer has a low prior belief, the purchase likelihood would decrease in information processing cost under asymmetric attention but would increase under symmetric attention. This comparison highlights the importance of taking into account the consumer's attention allocation when studying the consumer's information processing and purchase behavior.

3.3 Extension: Optimal Attention Allocation under Irreducible Uncertainty

In the previous case, an implicit assumption is that all of the information is available to the consumer and thus the consumer can fully reduce the uncertainty about product fit if she processes all the information. However, there often exists some uncertainty that the consumer can not reduce. For example, the full product information may not always be available to the consumer. To see this, consider the case where a consumer is booking a cruise trip one month ahead. The weather in the destination is an important factor that the consumer takes into account, but it is almost impossible for the consumer to precisely



(a) Case I: Under high prior belief

(b) Case II: Under low prior belief

Figure 5: Comparison between asymmetric attention and symmetric attention

predict the weather one month later in the destination. Therefore, there is irreducible uncertainty about the weather condition in this example. Besides the unavailability of some product information, the credibility of information can also lead to irreducible uncertainty. For instance, online product reviews may be fake and detection of a fake review is often very difficult for both the review sites and the reviewers, thereby leading to some irreducible uncertainty about product fit.

In this section, I am interested in the following question: As more information becomes available or credible, e.g., as more online product reviews become available and can be verified by the review site, how would a consumer change her attention allocation strategy? In particular, would the consumer increase attention to positive information or negative information? In the following, I extend our basic model in the previous section by incorporating the irreducible uncertainty.

First, denote $X \in \{0, 1\}$ as the interim state for fit given the consumer processes all the information available to her. Specifically, if $X = 1$, it implies that the consumer's belief about product fit increases if the consumer processes all the available information; conversely, if $X = 0$, it means that the consumer's belief about product fit decreases if she processes all available product information. Next, denote $Y \in \{0, 1\}$ as the final state for fit, where $Y = 1$ means the product indeed fits and $Y = 0$ means the product does not fit. Furthermore, denote $\alpha \equiv P(X = Y)$ as the probability of interim state X being equal to the final state Y, where $\frac{1}{2} \leq \alpha \leq 1$. In particular, as more uncertainty can be reduced (e.g., more product information becomes available or credible), α increases. Note that if $\alpha = 1$, then it implies that full product information is available to the consumer, and the consumer faces the same problem of information processing as in the previous section; but if $\alpha = \frac{1}{2}$, then it means that no product information is available to the consumer and the consumer makes a purchase decision based on her prior belief. For simplicity, I assume the consumer has an equal prior belief about product fit, i.e., $P(Y = 1) = P(Y = 0) = \frac{1}{2}$. This assumption leads to clear results and our main results would not change qualitatively under an arbitrary prior belief. Note that given this equal prior belief, it is easy to see that $P(X = 1) = P(X = 0) = \frac{1}{2}$.

As in the previous section, the consumer's information processing strategy is characterized her choice of the signal structure $\delta_1 \equiv P(S = 1|X = 1)$ and $\delta_0 \equiv P(S = 0|X = 0)$. Different signal realizations lead to different actions and thus the value of information is quantified by $EV \equiv P(S = 1)(P(Y = 1|S = 1)U_1 + P(Y = 0|S = 1)U_0)$, while the cost of information is quantified by $\lambda I(X, S)$. Note that the consumer can only reduce the uncertainty about the interim state X. In particular, the consumer can perfectly know the state X if she processes all the available information, but she can not further reduce uncertainty between interim state X and final state Y. In other words, α measures the amount of reducible uncertainty, which is exogenous to the consumer (the higher α is, the more uncertainty is reducible). Furthermore, since the consumer can only reduce uncertainty about X, the cost function is given by $\lambda I(X, S)$ rather than $\lambda I(Y, S)$.

Taking into account the irreducible uncertainty, the consumer chooses the optimal signal structure that maximizes the net expected utility as follows:

$$EU^* \equiv \sup_{\delta_1,\delta_0} \underbrace{P(S=1)(P(Y=1|S=1)U_1 + P(Y=0|S=1)U_0)}_{value\ of\ information} - \underbrace{\lambda I(X;S)}_{cost\ of\ information} \quad (8)$$

$$\text{s.t. } \delta_1 + \delta_0 > 1$$

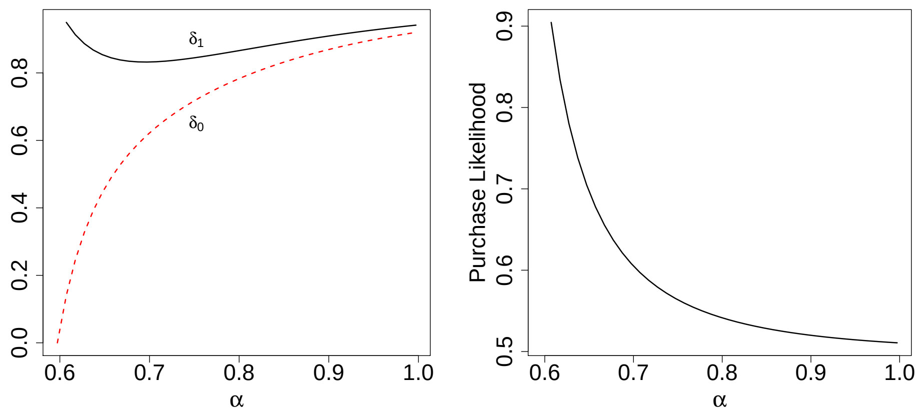
Note that the prior utility $EU' = \frac{1}{2}U_1 + \frac{1}{2}U_0$. If $EU^* \geq \max\{EU', 0\}$, then the consumer chooses to process information with optimal signal structure, δ_1^* and δ_0^*, given by (8), and then she purchases the product only when receiving the positive signal (i.e., $S = 1$). Otherwise, if $\max\{EU', 0\} > EU^*$, then the consumer chooses not to learn and makes a purchase decision based on her prior belief (i.e., the consumer purchases the product if and only if $EU' \geq 0$). The following proposition characterizes the optimal attention allocation under irreducible uncertainty.

Proposition 4 *Define* $k = \frac{U_1+U_0}{\lambda}$, $l = \frac{\alpha U_1+(1-\alpha)U_0}{\lambda}$, $l' = \frac{(1-\alpha)U_1+\alpha U_0}{\lambda}$, $h = 1 + \frac{2(1-e^k)}{2e^k-(e^l+e^{l'})}$, $\alpha^* = \frac{\lambda \log(e^k+\sqrt{e^{2k}-e^k})-U_0}{U_1-U_0}$ *and* $\alpha' = \frac{\lambda \log(1+\sqrt{1-e^k})-U_0}{U_1-U_0}$. *The following holds:*

1. *If* $U_1 + U_0 \geq 0$*: for* $\max\{\alpha^*, \frac{1}{2}\} < \alpha < 1$, $\delta_1^* = \frac{e^l h}{1+e^l h}$ *and* $\delta_0^* = \frac{1}{1+e^{l'} h}$, *and if* $\alpha^* > \frac{1}{2}$ *then for* $\frac{1}{2} < \alpha \leq \alpha^*$, *the consumer purchases without learning.*

2. *If* $U_1 + U_0 < 0$*: for* $\max\{\alpha', \frac{1}{2}\} < \alpha < 1$, $\delta_1^* = \frac{e^l h}{1+e^l h}$ *and* $\delta_0^* = \frac{1}{1+e^{l'} h}$, *and if* $\alpha' > \frac{1}{2}$ *then for* $\frac{1}{2} < \alpha \leq \alpha'$, *the consumer does not learn and does not purchase.*

The above proposition shows that if most of the product information is not available or not credible (i.e., α is small), then paying attention to the available information would not change a consumer's belief very much, and thus the consumer would not start to process any information and would purchase the product if the prior utility $\frac{1}{2}U_1 + \frac{1}{2}U_0 \geq 0$ and choose her outside option if $\frac{1}{2}U_1 + \frac{1}{2}U_0 < 0$. Only when α is high enough, the consumer would process some information, and her subsequent attention allocation decision is influenced by α. In the following, I investigate whether the consumer would increase attention to positive

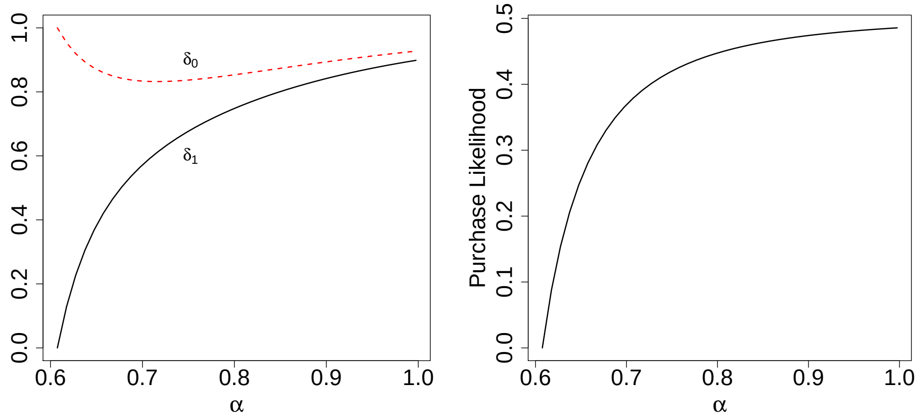


(a) (b)

Figure 6: Impact of irreducible uncertainty on attention allocation under high prior utility

information or negative information as more information becomes available and credible (i.e., α becomes larger).

First, under a relatively high α such that the consumer chooses to process information, if the prior utility $\frac{1}{2}U_1 + \frac{1}{2}U_0 \geq 0$ (note that the prior belief $P(X = 1) = \frac{1}{2}$), then the consumer would allocate relatively more attention to the positive information than negative information because the gain of buying a good product is larger than the loss of buying a bad product. Now, since the consumer initially pays less attention to the negative information, it is marginally cheaper to improve the signal accuracy δ_0 under "not fit" state (i.e., $\lambda \frac{dI(X,S)}{d\delta_0} < \lambda \frac{dI(X,S)}{d\delta_1}$). Therefore, as α becomes higher, the consumer has a higher incentive to learn and would increase attention to the negative information and thus the signal accuracy δ_0 becomes higher. In the meanwhile, since attention is limited, the consumer would decrease attention to the positive information so that the signal accuracy δ_1 under "fit" state decreases, but if α becomes high enough, then the consumer has a very high incentive to process information and thus she would increase attention to both positive and negative information (See Figure 6a). Overall, the consumer's purchase likelihood would decrease in α if the prior utility of the product is high (Figure 6b).



(a) (b)

Figure 7: Impact of irreducible uncertainty on attention allocation under low prior utility

Conversely, if the prior utility $\frac{1}{2}U_1 + \frac{1}{2}U_0 < 0$, the consumer would increase attention to the positive information as α becomes higher, because the consumer initially pays less attention to the positive information due to the low prior utility and thus it is marginally cheaper to improve δ_1 as α becomes higher. In the meanwhile, the consumer would first decrease and then increase attention to positive information as α becomes higher (Figure 7a). Overall, the consumer's purchase likelihood would increase in α (Figure 7b). I summarize the discussion above in the following proposition.

Proposition 5 *The following holds:*

1. *When prior utility* $\frac{1}{2}U_1+\frac{1}{2}U_0 < 0$*,* δ_0^* *increases with* α *for* $\alpha^* < \alpha < 1$*;* δ_1^* *decreases with* α *for* $\alpha^* < \alpha \leq \min\{\alpha^{**}, 1\}$ *and increases with* α *for* $\min\{\alpha^{**}, 1\} < \alpha < 1$*. Purchase probability* $P(S = 1)$ *decreases with* α *for* $\alpha^* < \alpha < 1$*.*

2. *When prior utility* $\frac{1}{2}U_1+\frac{1}{2}U_0 < 0$*,* δ_1^* *increases with* α *for* $\alpha' < \alpha < 1$*;* δ_0^* *decreases with* α *for* $\alpha' < \alpha \leq \min\{\alpha'', 1\}$ *and increases with* α *for* $\min\{\alpha'', 1\} < \alpha < 1$*. Purchase probability* $P(S = 1)$ *increases with* α *for* $\alpha' < \alpha < 1$*.*

Chapter 4: Firm Strategies under Consumer Attention Allocation

Since the consumer's attention allocation decision determines how the consumer would eventually evaluate the product, it is important for the seller to understand the interaction between a certain marketing strategy and the consumer's attention allocation decision. Therefore, in this chapter, I turn my attention to study the firm strategies in response to the consumer's attention allocation decision. In the following, I consider several marketing strategies that closely interact with the consumer's attention allocation: Pricing, return policy and information design (i.e., the seller chooses how much information and which information available for consumers to process).

4.1 Price

Denote $X \in \{0, 1\}$ as the binary state for fit and p as the price chosen by the seller. If the product fits (i.e., $X = 1$), the consumer obtains utility $U_1 = 1 + w - p$, and if the product does not fit (i.e., $X = 0$), then the consumer obtains utility $U_0 = 1 - p$. In this case, the base utility of product is normalized to be 1 and $w > 0$ is the match value. The utility of the outside option is assumed to be 0.

Facing uncertainty about the match, the consumer can process information to reduce the uncertainty. As in Chapter 3, I assume that the consumer has a prior belief $P(X = 1) = q$, and the signal structure chosen by the consumer is characterized by two state-dependent signal accuracy: $\delta_1 \equiv P(S = 1|X = 1)$ and $\delta_0 \equiv P(S = 0|X = 0)$. On the one hand, when the consumer chooses to process information prior to making a purchase decision, she chooses an optimal information processing strategy, characterized by δ_1^* and δ_0^*, that maximizes the net expected utility. Note that when the consumer receives a positive signal (i.e., $S = 1$), she always chooses to purchase the product; but when the consumer receives a negative signal (i.e., $S = 0$), she chooses her outside option. Therefore, when the consumer chooses to process information, she can obtain utility EU^* under optimal attention allocation strategy

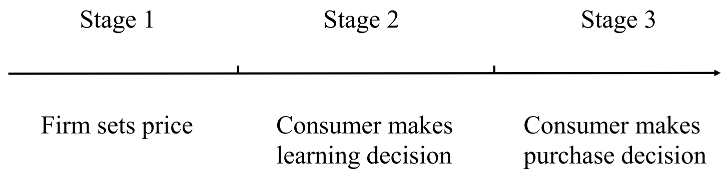


Figure 8: Stages of the game

(i.e., δ_1^* and δ_0^*) that maximizes the net expected utility as follows.

$$EU^* \equiv \sup_{\delta_1,\delta_0} \underbrace{P(S=1)(P(X=1|S=1)(1+w-p)+P(X=0|S=1)(1-p))}_{value\,of\,information} - \underbrace{\lambda I(X;S)}_{cost\,of\,information} \tag{9}$$
$$\text{s.t. } \delta_1+\delta_0>1$$

On the other hand, if the consumer chooses not to process information, then she makes the purchase decision based on her prior belief. In this case, the utility from purchasing based on prior belief is $EU' = qU_1 + (1-q)U_0$, and she purchases the product if and only if $EU' \geq 0$.

First, the seller decides the product price, p. Then, the consumer observes the price and decides whether or not to process information to reduce the uncertainty about product fit. Specifically, If $\max\{EU', 0\} \geq EU^*$, then the consumer chooses not to learn and makes a purchase decision based on her prior belief (i.e., the consumer purchases the product if and only if $EU' \geq 0$); otherwise, if $\max\{EU', 0\} < EU^*$, then the consumer chooses to process information with optimal signal structure, δ_1^* and δ_0^*, and she purchases the product only when receiving the signal $S=1$.

Define $k=\frac{w}{\lambda}$, $l=\frac{p-1}{\lambda}$, $\underline{p}=1-\lambda\log(1-q+qe^{-k})$ and $\overline{p}=1+\lambda\log(1-q+qe^{k})$. The following proposition shows how the price affects the consumer's attention allocation.

Proposition 6 *The following holds:*

1. *If $p \leq \underline{p}$, the consumer processes no information and buys the product.*

2. *If $\underline{p} < p < \bar{p}$, δ_1^* and δ_0^* are given as follows:*

$$\delta_1^* = \frac{1 - \frac{1-q}{q}\frac{e^l - 1}{e^k - e^l}}{1 - e^{-k}}, \quad \delta_0^* = \frac{1 - \frac{q}{1-q}\frac{e^k - e^l}{e^k(e^l - 1)}}{1 - e^{-k}} \tag{10}$$

The purchase probability $P(S = 1) = \frac{q}{1-e^{-l}} + \frac{1-q}{1-e^{k-l}}$. Particularly, when price p becomes higher, δ_1^* decreases and δ_0^* increases.

3. *If $p \geq \bar{p}$, the consumer processes no information and does not buy the product.*

Proof: See appendix.

As illustrated in Figure 9, when the price is low enough (i.e., $p \leq \underline{p}$), there is not much gain from processing information because it would not hurt much even if the product turns out to not fit the consumer. Thus, due to the high prior utility of the product, the consumer processes no information and buys the product. When the price is high enough (i.e., $p \geq \bar{p}$), there is not much gain as well from processing information because the utility of the product is low even if the product turns out to fit the consumer, and thus the consumer processes no information and does not buy the product. Only when the price is medium (i.e., $\underline{p} < p < \bar{p}$), the consumer has an incentive to reduce the uncertainty by processing information. In particular, as price becomes higher but not too high, the utility of buying a matched product decreases, while the disutility of buying a mismatched product increases. Therefore, the consumer increases attention to negative information and decreases attention to positive information as the price increases, i.e., δ_1^* decreases and δ_0^* increases in price p.

Going on to the firm's problem, according to Proposition 6, there exist two pricing regimes under which the consumer may purchase: (1) "learning-prevented" pricing strategy (i.e., $p \leq \underline{p}$) under which the seller charges a relatively low price so that the consumer purchases the product without processing information, and (2) "learning-promoted" pricing strategy (i.e., $\underline{p} < p < \bar{p}$) under which the seller charges a relatively high price so that the consumer always processes information.

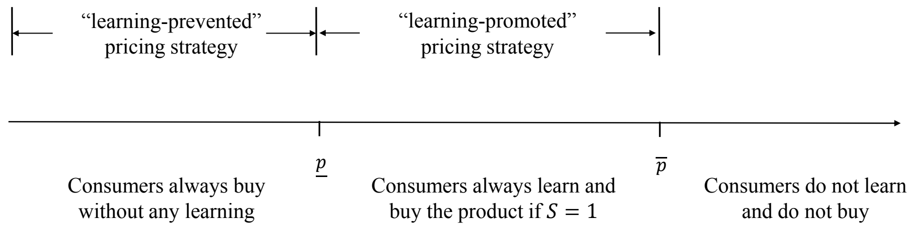


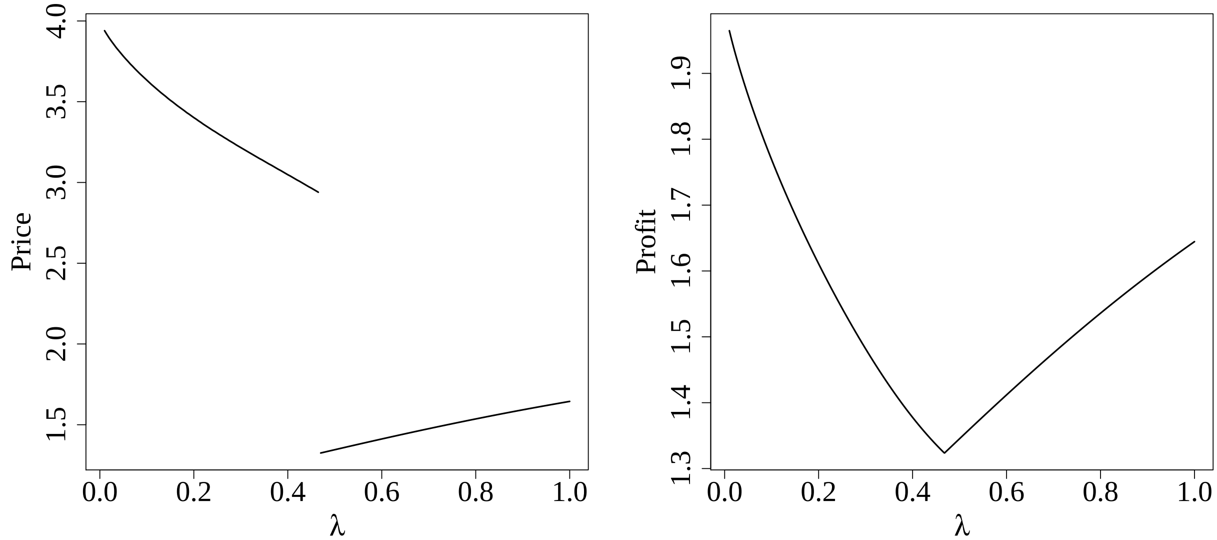
Figure 9: "Learning-Prevented" and "Learning-Promoted" Strategies (horizontal axis denotes price)

Under the "learning-prevented" pricing strategy (i.e., $p \leq \underline{p}$), consumers buy the product without processing information. Therefore, the demand is $D_{pre} = 1$ (I normalize the size of the market to be 1). The optimal price is the maximum possible price at which the consumer does not learn, which is given by $p^*_{pre} = \underline{p}$. Correspondingly, the optimal profit is $\pi^*_{pre} = p^*_{pre} D_{pre} = \underline{p}$. Note that both the optimal price and profit increase in information processing cost λ. The intuition is the following: under the "learning-prevented" pricing strategy, the seller charges a price such that the consumer is indifferent between processing information and buying the product without processing information. Therefore, as information processing cost becomes higher, the consumer has a lower incentive to process information, thereby allowing the seller to charge a higher price without inducing consumers to learn.

Under the "learning-promoted" pricing strategy (i.e., $\underline{p} < p < \bar{p}$), the consumer processes information and purchases the product when receiving $S = 1$. Thus, the demand in this case is $D_{pro} = P(S = 1) = \frac{q}{1-e^{-l}} + \frac{1-q}{1-e^{k-l}}$, and the seller chooses optimal price p^*_{pro} that maximizes profit as below:

$$\pi^*_{pro} \equiv \sup_{p} \left(\frac{q}{1 - e^{-l}} + \frac{1 - q}{1 - e^{k-l}} \right) p \tag{11}$$

$$\text{s.t. } \underline{p} < p < \bar{p}.$$



(a) Price

(b) Profit

Figure 10: Impact of consumer information processing cost on price and profit. For the plots, $q = 0.5$, $w = 3$.

Note that since a higher price induces consumer to pay more attention to negative information of a product and this effect becomes stronger under a higher information processing cost, the demand sensitivity to price becomes higher as information processing cost becomes higher, forcing the seller to charge a lower price (i.e., p^*_{pro} decreases in λ). This is shown in the following proposition.

Proposition 7 *Given $\underline{p} < p < \bar{p}$, as information processing cost λ becomes higher, the demand sensitivity to price becomes higher, i.e.,* $\frac{d^2 P(S=1)}{d\lambda dp} < 0$.

Proof: See appendix.

Overall, when the consumer has a low information processing cost, she has a high incentive to process information and her willingness to pay becomes higher when receiving a positive signal $S = 1$, and thus the seller prefers the "learning-promoted" pricing strategy to encourage the consumer to learn. However, when the consumer has a high information processing cost, she has a low incentive to process information and thus the seller prefers the "learning-prevented" pricing strategy to induce the consumer to buy the product without any learning. Because of the switch of pricing regime, the optimal price and profit may first decrease (under "learning-promoted" pricing strategy) and then increase (under "learning-prevented" pricing strategy) in the consumer's information processing cost. This is illustrated in Figure 10. The analysis above is summarized in the following proposition.

Proposition 8 *The optimal price and profit may first decrease and then increase in the consumer's information processing cost* λ*.*

Proof: The proof is clear from the arguments above.

Pricing without considering attention allocation In the pricing model above, I consider the firm's pricing decision when the consumer can flexibly choose the signal structure, i.e, choose both how much information to process and how to allocate attention to different types of information. In order to further examine how the consumer's attention allocation decision influences the firm's pricing decision, I compare the results above with those obtained in the following case where the consumer is not allowed to do the differential learning, i.e., the consumer chooses the signal precision of a symmetric signal. Specifically, I assume $P(X = 1|S = 1) = P(X = 0|S = 0)$ and thus the consumer's information processing decision is characterized by choosing only one conditional probability: $\delta \equiv P(X = S)$. This model serves as benchmark with which I can compare the results of the full model to understand how seller strategies may be different under the more general learning formulation.

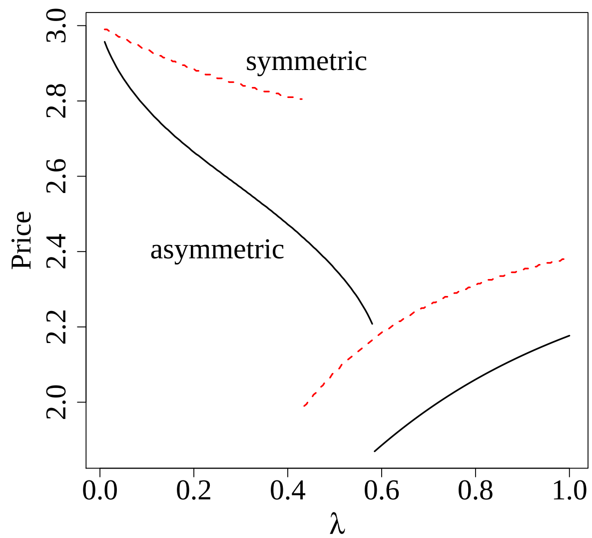
In the benchmark model, the higher the signal precision (i.e., $\delta \equiv P(X = S)$) is, the more information is processed. By trading off the value and cost of information, the consumer chooses the optimal signal precision δ^* that maximizes the net expected utility as follows:

$$EU_S^* \equiv \sup_{\frac{1}{2}<\delta\leq 1} P(S=1)[P(X=1|S=1)(1+w-p)+P(X=0|S=1)(1-p)]-\lambda I(X,S)$$

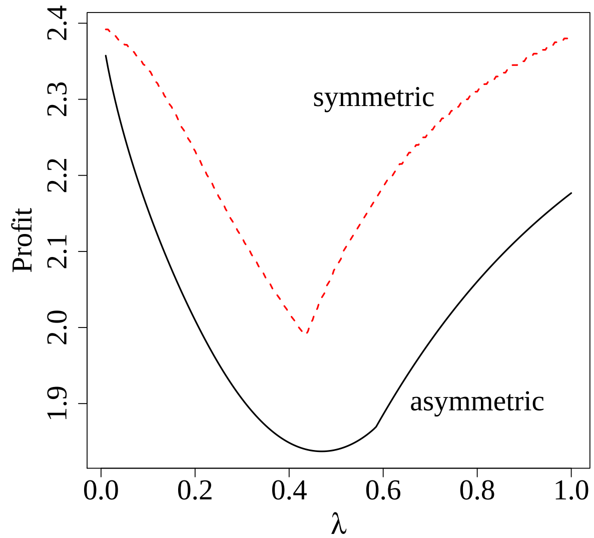
where $I(X,S) = -[q\delta + (1-q)(1-\delta)]\log[q\delta + (1-q)(1-\delta)] - [1 - q\delta - (1-q)(1-\delta)]\log[1 - q\delta - (1-q)(1-\delta)] + \delta\log\delta + (1-\delta)\log(1-\delta)$.

I find that the consumer's asymmetric attention allocation restricts the firm's pricing capabilities and thus reduces the profit (see Figure 11). Specifically, in the asymmetric attention model, the consumer can obtain higher utility by allocating differential attention to different types of information. Therefore, when the information processing cost is very high and the seller carries out the "learning-prevented" pricing strategy, the seller has to charge a lower price in the asymmetric case to prevent the consumer from learning than it does in the symmetric case. On the other hand, when the information processing cost is very low and the consumer chooses to learn, the demand sensitivity to price is higher in the asymmetric case than in the symmetric case. Therefore, under "learning-promoted" pricing strategy, the seller has to charge a lower price as well in the asymmetric case than in the symmetric case. Note that if the prior belief that the product will fit is high, then the consumer will allocate relatively more attention to positive information and less attention to negative information as the information processing cost becomes higher. This effect makes the "learning-promoted" pricing strategy generally more attractive to the seller in the asymmetric case than in the symmetric case. Therefore, when the information processing cost is medium, the seller may undercharge the price without considering the consumer's asymmetric attention allocation, because the seller would carry out "learning-promoted" pricing strategy in the asymmetric case but carry out "learning-prevented" pricing strategy in the symmetric case (Figure 11(a)). Since the consumer's asymmetric attention allocation restricts the firm's pricing capabilities, the seller obtains a lower profit in asymmetric attention case than in symmetric case (Figure 11(b)). I summarize the analysis above in the following proposition.

Proposition 9 *Asymmetric attention allocation restricts the firm's pricing capabilities and thus reduces the seller's profit.*



(a) Price



(b) Profit

Figure 11: Comparison between Asymmetric Model and Symmetric (Benchmark) Model (for the plots, $q = 0.8$, $w = 2$).

Proof: The proof is clear from the arguments above.

4.2 Return Policy

It is common that a seller offers return policy that allows consumers to return the product if it turns out not to fit. How does this return policy influence the consumer's attention allocation decision? And when should the seller offer the return policy taking into account its impact on consumer's attention allocation? In this section, I answer these questions. Specifically, as before, if $X = 1$ (i.e., the product fits), the utility from purchasing is $U_{X=1} = 1 + w - p$; However, if $X = 0$ (i.e., the product does not fit) and if the consumer returns the product, the utility is $U_{X=0} = -c_r$, where $c_r > 0$ is the returning cost for the consumer, while if he chooses not to return, then the utility is $U_{X=0} = 1 - p$. Clearly, the consumer will return the mismatched product if and only if $p \geq 1 + c_r$. Note that if the seller offers the return policy but charges a price p such that $p < 1 + c_r$, then the consumer does not return the mismatched product and thus the seller faces the same profit optimization problem as in

Section except with an additional constraint on price (i.e., $p < 1 + c_r$), which implies that offering return policy would never increase the profit in this case. Therefore, whenever the seller chooses to offer return policy, the price p must be larger or equal to $1 + c_r$.

I analyze this model in the lines of our previous analyses. The timing is the same as in Section except that at the beginning, the seller decides whether or not to offer a return policy in addition to choosing the price. Define $k = \frac{w}{\lambda}$, $l = \frac{p-1}{\lambda}$ and $h = \frac{c_r}{\lambda}$. The optimal signal structure, $\delta^*_{1,r}$ and $\delta^*_{0,r}$, chosen by the consumer with a return policy is given as follows:

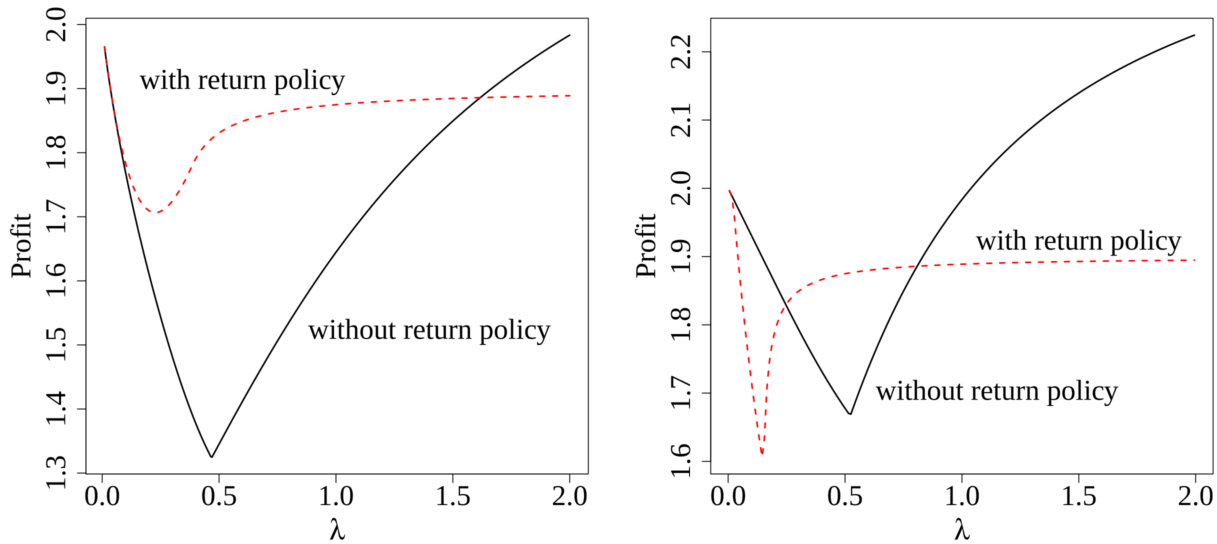
$$\delta^*_{1,r} = \frac{1 - \frac{1-q}{q}\frac{e^h-1}{e^{k+l+h}-e^h}}{1-e^{-(k+l+h)}}, \quad \delta^*_{0,r} = \frac{1 - \frac{q}{1-q}\frac{e^{k+l+h}-e^h}{e^{k+l+h}(e^h-1)}}{1-e^{-(k+l+h)}}. \tag{12}$$

By comparing (10) with (12), I find that offering return policy induces the consumer to allocate relatively more attention to positive information and less attention to negative information (i.e., $\delta^*_{1,r} > \delta^*_1$ and $\delta^*_{0,r} < \delta^*_0$). Intuitively, offering return policy decreases the loss of buying a mismatched product because the consumer can always return the mismatched product. Therefore, the consumer has a lower incentive to pay attention to negative information and correspondingly increases attention to positive information.

Proceeding to the firm's decisions, I find that under return policy, there also exist two pricing regimes under which the consumer may purchase. Define $\underline{p}_r = 1 + w + c_r + \lambda \log(1 - \frac{1-e^{-h}}{q})$ and $\bar{p}^r = 1 + w + c_r - \lambda \log(1 + \frac{e^h-1}{q})$. For $1 + c_r \leq p \leq \underline{p}_r$, the seller follows a "learning-prevented" pricing strategy and for $\underline{p}_r < p < \bar{p}^r$, the seller follows a "learning-promoted" pricing strategy (for $p \geq \bar{p}^r$, consumers do not learn and do not purchase.). Under the "learning-prevented" pricing strategy (i.e., $1 + c_r \leq p \leq \underline{p}_r$), the consumer processes no information and buys the product. In this case, with probability $P(X = 0) = 1 - q$, the product purchased by the consumer does not fit and thus the consumer returns it back to the seller. Consequently, the final demand for the product is always q and the seller charges optimal price $p^*_{pre,r} = \underline{p}_r$ and obtains profit $\pi^*_{pre,r} = q\underline{p}_r$. Under the "learning-promoted" pricing strategy (i.e., $\underline{p}_r < p < \bar{p}^r$), the consumer chooses to learn and chooses the optimal signal structure given in (12). In this case, the demand for the product is $D_{pro,r} = P(S=1)P(X=1|S=1) = P(X=1)P(S=1|X=1) = q\delta^*_{1,r}$ and the seller chooses the optimal price $p^*_{pro,r}$ that maximizes the profit given as follows:

$$\pi^*_{pro,r} \equiv \sup_p q\delta^*_{1,r}p$$
$$\text{s.t. } \underline{p}_r < p < \bar{p}^r$$

Given information processing cost λ, the seller chooses to offer return policy if and only if $\max\{\pi^*_{pre,r}, \pi^*_{pro,r}\} \geq \max\{\pi^*_{pre}, \pi^*_{pro}\}$, where π^*_{pre} and π^*_{pro} are the optimal profits without return policy under "learning-prevented" and "learning-promoted" pricing strategy respectively. I find that offering return policy is not profitable when the information processing cost is very high. This is illustrated through an example in Figure 12(a). When the information processing cost is high, the seller prefers "learning-prevented" pricing strategy under which the seller discourages the consumer from learning. In this case, offering a return policy further reduces the consumer's learning incentive, which allows the seller to charge a higher price. However, when the information processing cost is very high, this price increasing effect of return policy is small and cannot offset the demand decreasing effect, and offering a return policy in this case would reduce the firm's profit. On the other hand, when the information processing cost is low, the seller prefers "learning-promoted" pricing strategy under which the consumer always processes information. In this case, offering return policy induces the consumer to allocate relatively more attention to positive information and less attention to negative information. This asymmetric attention allocation effect of return policy alleviates the demand decreasing effect of return policy and promotes the price increasing effect. Overall, the seller prefers offering return policy when the information processing cost is not very high.



(a) Asymmetric attention model

(b) Symmetric attention model (benchmark)

Figure 12: Profitability of Return Policy

As before, I compare the asymmetric attention model with the symmetric attention (benchmark) model in which the consumer chooses the signal precision of a symmetric signal. I find that the consumer's flexible attention allocation improves the profitability of return policy. Specifically, offering a return policy induces the consumer to allocate relatively more attention to positive information and less attention to negative information, thus alleviating the demand decreasing effect of return policy but promoting the price increasing effect, and failure of considering this asymmetric attention allocation effect of return policy would make return policy appear less attractive. In particular, when the learning cost is low and the seller carries out the "learning-promoted" pricing strategy, the symmetric attention model would suggest that the seller should not offer a return policy (Figure 12(b)), which is not suggested by the asymmetric attention model (Figure 12(a)). I summarize the analysis above in the following proposition.

Proposition 10 *The seller prefers offering return policy unless the information processing cost is very high. Furthermore, the consumer's flexible attention allocation improves the* *profitability of return policy.*

Proof: The proof is clear from the arguments above.

4.3 Firm Information Design

A seller often can decide how much information and which information is available for consumers to process. For example, a software company can decide how many and which features are available for consumers to try in its free trial, and a seller may also have an incentive to suppress the negative product information, e.g., the seller can pay the third party to remove some negative online product reviews about its product. In general, a seller can influence the information environment in various ways, which is referred to as "information design". In the following, I start to investigate the firm's information design taking into account the consumer's attention allocation behavior. In particular, I examine when a seller has a high incentive vs. low incentive to suppress the negative information in the environment.

As before, consider a consumer who wants to reduce the uncertainty about whether the product matches her needs or not. Denote $X \in \{0, 1\}$ as the binary state for fit. If $X = 1$, then the product fits and the consumer obtains utility $U_1 = 1 - p$ from purchasing, where p is the price of the product; if $X = 0$, the product does not fit and the consumer obtains utility $U_0 = w - p$ from purchasing, where $w < 1$. I assume that the utility of the outside option is 0. The consumer does not know ex ante the value of state X but has a prior belief about it denoted by $q \equiv P(X = 1)$ (and I assume that the seller has the same prior belief as the consumer). By processing information, the consumer receives a signal $S \in \{0, 1\}$ about the state of X based on which the consumer can update her belief about product fit, and she can choose the signal structure $\delta_1 \equiv P(S = 1|X = 1)$ and $\delta_0 \equiv P(S = 0|X = 0)$ before processing any information. The timing of the game is discussed as follows (Figure 13).

At the beginning, the seller not only chooses the price but also chooses how much and

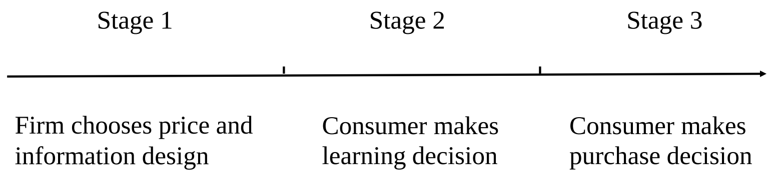


Figure 13: Stages of the game with information design

which information to be available for consumers to process (referred to as "information design"). In particular, the seller's information design imposes the upper bounds Ω_1 and Ω_0 on signal structure such that $P(S = 1|X = 1) \leq \Omega_1$ and $P(S = 0|X = 0) \leq \Omega_0$. For example, when the seller suppresses the negative information in the environment (e.g., removing some negative product reviews on some reviews sites), consumers are less likely to find out the serious drawbacks of the product that matter to them and thus they may be more likely to mistakenly buy a mismatched product, which corresponds to a lower Ω_0. Consider another information design example where a seller decides whether or not to allow his customers who purchased the product to write public reviews on the product page of the seller's own official website. Allowing customers to do this can enrich product information in the environment, which increases both Ω_1 and Ω_0. In particular, if dissatisfied customers are more likely than satisfied customers to leave reviews, then the seller actually chooses an information environment in which Ω_0 may be higher than Ω_1. In addition to the examples above, a seller may have other ways to influence the information environment. For instance, a seller can decide how many and which features to be available in the free trial of his product or choose the public testing environment for his product, which also determines the information structure Ω_1 and Ω_0. Note that throughout this paper I focus on the case where a seller can fully control the information environment (that is, the seller can flexibly choose Ω_0 and Ω_1.). This allows me to obtain clear results about the seller's incentive of information design in response to the consumer's attention allocation behavior, and the main results in the following would not change qualitatively if we restricts the seller's capability of influencing the information environment, e.g., the consumer may obtain product information from sources that can not be influenced by the seller.

Next, observing the price and information design by the seller, the consumer decides whether or not to process information to reduce uncertainty about product fit. Note that in many cases, consumers indeed have the chance to observe the information design before processing information. For example, in the free trial case, a consumer may first observe what features are included in the free trial and then decides whether or not to incur costs to try it. Besides, in the public testing example, a consumer may also first observe the testing environment and then chooses whether or not to pay attention to the results coming out of the testing. This observability assumption is similar to the assumption that a sender commits to the signal structure, which has been widely used in Bayesian Persuasion literature (e.g., Kamenica and Gentzkow 2011, Gentzkow and Kamenica 2016).[4] As discussed earlier, when she decides to process information, she chooses the signal structure $\delta_1 \equiv P(S=1|X=1)$ and $\delta_0 \equiv P(S=0|X=0)$ by optimizing attention to different types of information available to her. That is, she chooses an optimal signal structure, δ_1^* and δ_0^*, that maximizes the net expected utility as follows:

$$EU^* \equiv \sup_{\delta_1,\delta_0} \underbrace{q\delta_1(1-p)+(1-q)(1-\delta_0)(w-p)}_{value\,of\,information} - \underbrace{\lambda I(X,S)}_{cost\,of\,information} \tag{13}$$

$$\text{s.t. } 0<\delta_1 \leq \Omega_1,\ 0<\delta_0 \leq \Omega_0 \text{ and } \delta_1+\delta_0>1.$$

Denote $EU' \equiv qU_1 + (1-q)U_0$ as the expected utility based on the consumer's prior belief (referred to as "prior utility"). If $EU^* \geq \max\{EU', 0\}$, then the consumer chooses to process information with optimal signal structure, δ_1^* and δ_0^*, given by (13), and then she purchases the product only when receiving the positive signal (i.e., $S=1$). Otherwise, if

[4] A seller may also influence the information environment in an unobservable way, and how consumer's attention allocation affects this kind of information design is left for future research.

$\max\{EU', 0\} > EU^*$, then the consumer chooses not to learn and makes a purchase decision based on her prior belief (i.e., the consumer purchases the product if and only if $EU' \geq 0$).

On the one hand, if the seller charges a relatively low price p such that $p \leq q + (1-q)w$ (and thus prior utility $EU' \equiv q(1-p) + (1-q)(w-p) \geq 0$), then the seller has no incentive to provide information, because the consumer would always purchase the product when no information is available. On the other hand, if the seller charges a relatively high price p such that $p > q + (1-q)w$ (and thus prior utility $EU' < 0$), then the seller has to provide some information in order to invite the consumer to learn the match information. Otherwise, if no information is provided, then the consumer would not buy the product because of the low prior utility (i.e., $EU' < 0$). Therefore, the seller would provide information only when $p > q + (1-q)w$. In the following, I first discuss the seller's optimal information design to maximize the consumer's purchase likelihood given $p > q + (1-q)w$, and then I consider both optimal information design and optimal pricing decision.

Given $p > q + (1-q)w$, suppose the seller provides full information (i.e., $\Omega_1 = 1$ and $\Omega_0 = 1$). Under this full-information case, the consumer faces the same information processing problem as before. Specifically, one can show that given $\underline{q} < q < \bar{q}$ where $\underline{q} = \frac{1-e^{\frac{p-w}{\lambda}}}{1-e^{\frac{1-w}{\lambda}}}$ and $\overline{q} = \frac{p-w}{1-w}$, the consumer chooses to process information with optimal signal structure δ_1^* and δ_0^* that maximizes (6), and the seller can obtain profit $\hat{\pi} = P(S=1)p = (q\delta_1^* + (1-q)(1-\delta_0^*))p$. Now, all else equal, suppose the seller imposes a constraint $\Omega_0 = \delta_0^* - \epsilon$ on the signal structure $P(S=0|X=0)$, where $\epsilon > 0$ is an arbitrary small positive value. Given this constraint, the consumer's optimal attention allocation in (13) is given by δ_1^{**} and δ_0^{**}, where δ_0^{**} is bounded at Ω_0 (i.e., $\delta_0^{**} = \Omega_0 = \delta_0^* - \epsilon$) and δ_1^{**} satisfies $\delta_1^* < \delta_1^{**} < 1$. In other words, under the information constraint $\Omega_0 = \delta_0^* - \epsilon$, the consumer would process relatively less negative information but would process relatively more positive information as compared to the full information case. This increases the consumer's purchase likelihood and thus the seller obtains higher profit $\tilde{\pi}$ (i.e., $\tilde{\pi} = (q\delta_1^{**} + (1-q)(1-(\delta_0^* - \epsilon)))p > \hat{\pi} = (q\delta_1^* + (1-q)(1-\delta_0^*))p$). Therefore, a seller would further reduce Ω_0 until the consumer is indifferent between processing information and not processing information (and not buy the product)[5]. Denote Ω_0^* as the optimal upper bound that the seller can impose on $P(S = 0|X = 0)$, and denote EU^*_{design} as the utility of optimal learning under the constraint Ω_0^*. Since $EU^*_{design} = P(S = 1)(P(X = 1|S = 1)U_1 + P(X = 0|S = 1)U_0) - \lambda I(X, S)$, one can obtain the following relationship

$$EU^*_{design} = q\delta_1^{**}(1-p) + (1-q)(1-\Omega_0^*)(w-p) - \lambda I(X,S)_{\delta_1=\delta_1^{**},\delta_0=\Omega^*_{0,}} = 0 \qquad (14)$$

$I(X,S)_{\delta_1=\delta_1^{**},\delta_0=\Omega^*_{0,exo}}$ is given in (5) with $\delta_1 = \delta_1^{**}$ and $\delta_0 = \Omega_0^*$. Furthermore, δ_1^{**} satisfies $\delta_1^* < \delta_1^{**} < 1$ and is given by the first order condition of consumer maximizing the utility of learning with respect to δ_1, as shown in (15).

$$e^{\frac{1-p}{\lambda}} = \frac{q + (1-q)\frac{\Omega_0^*}{1-\delta_1^{**}}}{q + (1-q)\frac{1-\Omega_0^*}{\delta_1^{**}}} \qquad (15)$$

Note that equation (14) and (15) jointly determine Ω_0^* and δ_1^{**}. Next, it is easy to see that the seller has no incentive to impose constraint Ω_1 on $P(S = 1|X = 1)$, because any Ω_1 satisfying $\delta_1^{**} \leq \Omega_1 \leq 1$ would not change the consumer's information processing strategy and thus would not change her purchase likelihood (if $\Omega_1 < \delta_1^{**}$, then δ_1^{**} is bounded at Ω_1 and the consumer's purchase likelihood always decreases). However, for ease of exposition, I always assume that the seller sets the constraint $\Omega_1^* = \delta_1^{**}$, i.e., the seller has no incentive to provide information more than the amount that the consumer will process. This assumption can be justified when the seller has an arbitrarily small cost to provide information.

Since the seller's main incentive of information design is to impose the constraint Ω_0 on $P(S = 0|X = 0)$, I am interested in how this incentive is influenced by the consumer's

[5]Note that $EU' < 0$ under $p > q + (1-q)w$. So if the consumer does not process information, she would choose her outside option. I also assume that the consumer would always process information when she is indifferent between processing information and not processing information.

attention allocation behavior. First, when the consumer has a lower prior belief about product fit, our intuition may suggest that a seller would have a higher incentive to suppress negative information. However, this intuition has a flaw, because it does not consider how the consumer's prior belief influences her attention allocation to different types of information. As discussed in Chapter 3, a lower prior belief would motivate the consumer to pay relatively more attention to negative information and less attention to positive information. Therefore, in order to invite the consumer with a lower prior belief to process information, the seller has to relax the constraint Ω_0, allowing relatively more negative information available for the consumer to process. Otherwise, the utility of learning would be too low for the consumer and thus she would not process information and not buy the product (note that the prior utility$EU' < 0$ given $p > q + (1 - q)w$). In short, if the consumer has a lower prior belief, the seller would make relatively more negative information available for the consumer to process, i.e., optimal information constraint Ω_0^* *decreases* with consumer's prior belief q. Moreover, when the seller relaxes Ω_0 and thus more negative information is available for the consumer to process, the consumer would decrease attention to positive information (i.e., δ_1^{**} decreases), inducing the seller to decrease Ω_1 (under the assumption that the seller has no incentive to provide information more than the amount that the consumer will process).Therefore, the optimal constraint Ω_1^* *increases* with q (see Figure 14a).

Similarly, as the consumer's information processing cost λ becomes higher, the utility of processing information decreases. Thus, to invite the consumer to learn, the seller has to relax the constraint Ω_0, allowing relatively more negative information available for consumers to process. Thus, the optimal constraint Ω_0^* *increases* in consumer's information processing cost λ (and the optimal constraint Ω_1^* *decreases* in λ because the consumer would decrease attention to positive information, see Figure 14b). The analysis above is summarized in the following proposition.

Proposition 11 ***(Optimal information design under exogenous price)*** *Denote* δ_1^*

and δ_0^* as the optimal signal structure chosen by the consumer under full information case. Suppose price p is exogenously given,

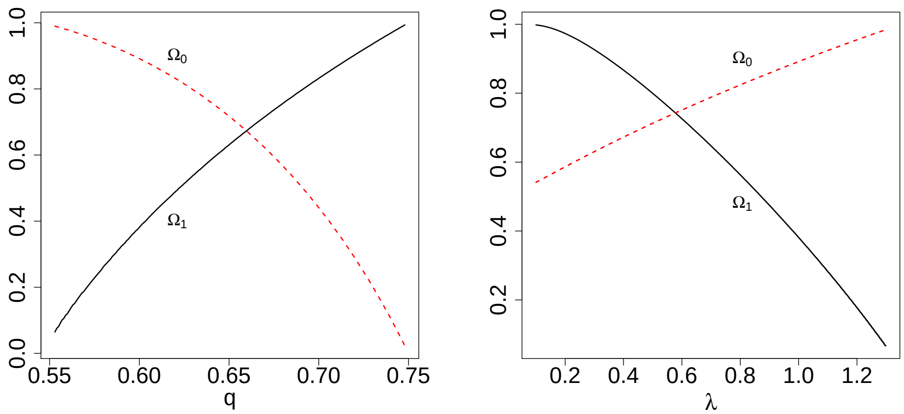
1. If $\underline{q} \equiv \frac{1-e^{\frac{p-w}{\lambda}}}{1-e^{\frac{1-w}{\lambda}}} < q < \bar{q} \equiv \frac{p-w}{1-w}$, then the seller would design an information environment characterized by Ω_0^* and Ω_1^* to invite the consumer to learn. Under this information design, the optimal signal structure δ_1^{**} and δ_0^{**} chosen by the consumer satisfies $\delta_0^{**} = \Omega_0^* < \delta_0^*$ and $\delta_1^* < \delta_1^{**} < 1$,, i.e., the consumer processes relatively more positive information than negative information as compared to the full information case. The optimal information design Ω_0^* and Ω_1^* are jointly determined by equation (14) and (15) (given the assumption that the seller always sets Ω_1^* equal to δ_1^{**}). In particular, lower prior belief q or higher information processing cost λ motivates the seller to increase Ω_0^* but decrease Ω_1^*;

2. If $q \leq \underline{q}$, the seller provides no information and the consumer always chooses outside option;

3. If $q \geq \bar{q}$, the seller provides no information and the consumer always purchases the product

Proof: See appendix.

Now I turn attention to the optimal pricing under information design. As discussed previously, a seller would provide information only when it charges a relatively high price p such that $p > q + (1-q)w$ (i.e., the prior utility $EU' < 0$). Therefore, if the seller provides no information, then he charges a price $p^* = q + (1-q)w$ such that the consumer always purchases the product without any learning. If the seller instead chooses to provide some information, then he charges a price $p > q + (1-q)w$ and designs the information environment such that the consumer is indifferent between processing information and not processing information (and not buy the product), as shown earlier in equation (14). In this case, the consumer always processes information and the optimal price p^* satisfies the following necessary condition, the first order condition of consumer maximizing the utility



(a) Impact of prior belief on information design ($p = 0.5, \lambda = 1, w = -1$)

(b) Impact of learning cost on information design ($p = 0.5, q = 0.6, w = -1$)

Figure 14: Optimal information design under exogenous price $p > q + (1-q)w$

of learning with respect to δ_1,

$$p^* = 1 - \lambda \log \frac{q + (1-q)\frac{\Omega_0}{1-\delta_1^{**}}}{q + (1-q)\frac{1-\Omega_0}{\delta_1^{**}}} \tag{16}$$

Now, according to equation (14) and (16), one can obtain the following equation (17), which implicitly determines δ_1^{**} as a function of Ω_0, denoted by $\delta_1^{**}(\Omega_0)$.

$$(q\delta_1^{**} + (1-q)(1-\Omega_0))\underbrace{(1 - \lambda \log \frac{q + (1-q)\frac{\Omega_0}{1-\delta_1^{**}}}{q + (1-q)\frac{1-\Omega_0}{\delta_1^{**}}})}_{p^*} = q\delta_1^{**} + (1-q)(1-\Omega_0)w - \lambda I(X,S)_{\delta_1^{**},\Omega_0} \tag{17}$$

As before, I assume the seller always sets $\Omega_1 = \delta_1^{**}$, i.e., the seller has no incentive to provide information more than the amount that the consumer will process. Now, based on equation (17), one can derive the seller's profit π as a function of Ω_0, which is given by (18)

$$\pi = q\delta_1^{**}(\Omega_0) + (1-q)(1-\Omega_0)w - \lambda I(X,S)_{\delta_1^{**}(\Omega_0),\Omega_0} \tag{18}$$

.

Therefore, the optimal information design Ω_0^* can be obtained by solving the following profit optimization problem

$$\pi^* = \sup_{0<\Omega_0<1} q\delta_1^{**}(\Omega_0) + (1-q)(1-\Omega_0)w - \lambda I(X,S)_{\delta_1^{**}(\Omega_0),\Omega_0} \tag{19}$$

After obtaining the optimal Ω_0^*, one can recover the optimal Ω_1^* from (17) and then optimal price p^* from (16). Although there is no closed-form solution for optimal pricing and optimal information design, numerical analysis shows a number of interesting insights about the pricing and information design (Figure 15). I find that on the one hand, when the consumer's information processing cost is low, a seller is willing to invite consumers to learn about the product match, and thus it would make some product information available for consumers to learn. Moreover, if the consumer chooses to learn, then the seller can charge a relatively high price, because the consumer's willingness to pay becomes higher when receiving a positive signal. In particular, a lower information processing cost increases the consumer's incentive to learn and thus the seller can charge a higher price. Therefore, the optimal price decreases with the consumer's information processing cost. However, when the consumer's information processing cost is high enough, the seller does not want the consumer to learn, because otherwise he has to charge a very low price to invite consumers to process information. Thus, when the consumer has a high information processing cost, the seller would not provide any product information and instead would charge a relatively low price $p^* = q + (1-q)w$ (i.e., $EU' = 0$) such that the consumer would always purchase the product.

Furthermore, I find that a seller may charge a lower price when he can fully control the information environment (i.e., the seller can choose Ω_1 and Ω_0 flexibly) than when he can not (and full information is available to consumers), as shown in Figure 15. This is because a lower price can increase the consumer's incentive of processing information, allowing the

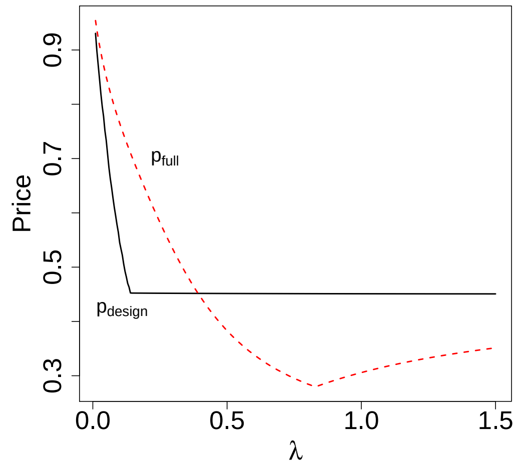
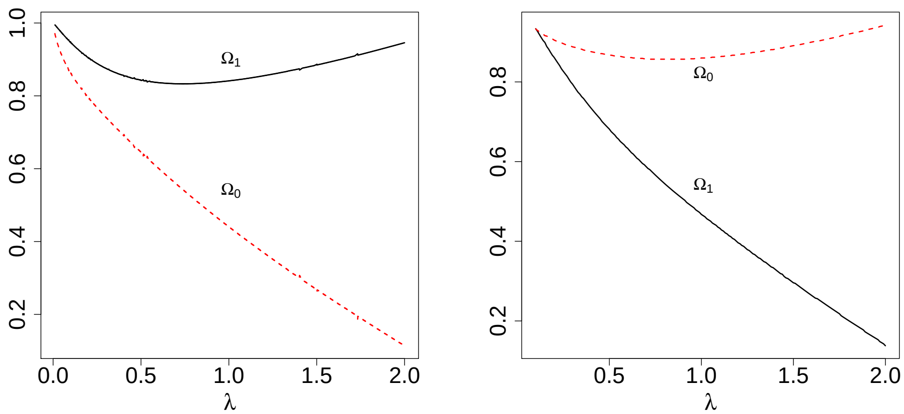


Figure 15: Price comparison between full information and information design. p_{full} is the optimal price under full information and p_{design} is the optimal price under information design

seller to further bias the information environment to increase the consumer's purchase likelihood. However, when the consumer has a very high information processing cost, the seller wants the consumer to buy the product without processing any information. Therefore, if the seller can design the information environment, he can eliminate the consumer's search by simply making no information available and charging a price $p^* = q + (1 - q)w$; but if the seller has no control on the information environment and full information is always available for the consumer, then the seller has to charge a relatively lower price to eliminate the consumer's search, because the consumer would have a relatively higher incentive to process information in the full information case as compared to the information design case. Overall, when the consumer's information processing cost is low, a seller may charge a lower price with information design than without information design, but he may charge a higher price instead when the consumer's information processing cost is high.

Finally, in terms of optimal information design under endogenous pricing, I find that the seller, similar to the exogenous price case, has a lower incentive to suppress negative information in the environment, as the consumer's prior belief becomes lower. However, I find



(a) Under high prior belief ($q = 0.6, w = -1$)

(b) Under low prior belief ($q = 0.4, w = -1$)

Figure 16: Optimal information design under endogenous pricing

that the consumer's information processing cost influences a seller's information design in a more sophisticated way under the endogenous price case. Specifically, given the consumer has a high prior belief, a higher information processing cost motivates a seller to further decrease the proportion of negative information in the environment (i.e., the difference between Ω_1 and Ω_0 becomes larger, see Figure 16a), which is in contrast to the exogenous price case above; but if the consumer has a low prior belief, then a higher information processing cost instead forces a seller to increase the proportion of negative information (i.e., the difference between Ω_0 and Ω_1 becomes larger, see Figure 16b). To understand this, note that as discussed earlier, the consumer's attention allocation is mainly affected by her prior belief if the information processing cost is high. Therefore, as information processing cost becomes higher, a consumer with a high prior belief would have a lower interest in processing negative information, allowing the seller to decrease the proportion of negative information in the environment; However, if the consumer has a low prior belief, then the consumer would like to process more negative information, forcing the seller to increase the proportion of negative information. I summarize the results from numerical analysis in the following proposition.

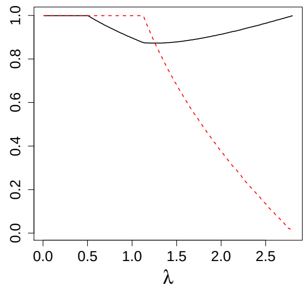
Proposition 12 *Under endogenous pricing, when the consumer has a lower prior belief, the seller has a lower incentive to suppress negative information. Furthermore, the incentive to suppress negative information increases with consumer's information processing cost when the consumer has a high prior belief, but the incentive would decrease with the cost when the consumer has a low prior belief. Finally, if the information processing cost is low, the seller may charge a lower price under information design than under no information design; if the information processing cost is high, the seller may instead charge a higher price under information design than under no information design.*

Chapter 5: Discussions and Conclusions

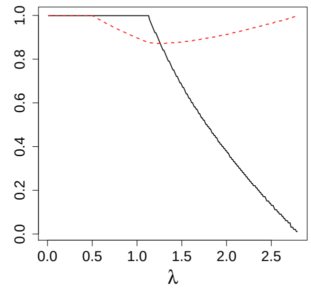
In this chapter, I first discuss the potential alternative information cost functions that can be used in our context, and then I discuss the practical implications of this research and how to test the proposed theory in an experimental setting. Finally, I make conclusions about my dissertation.

5.1 Discussion about alternative information cost functions

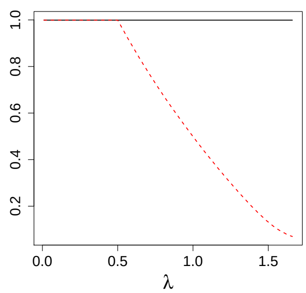
Although our use of mutual information as the measure of information costs has found both theoretical and empirical support as mentioned earlier, it is worth applying other meaningful cost functions to check robustness of results obtained using the entropy based cost function. One alternative cost function is based on the intuition that processing information also reduces on average the variance of belief distribution. Therefore, the amount of information contained in a signal can plausibly measured by the reduction of the variance of belief distribution. Specifically, suppose the consumer has a prior belief that the product would match with probability $P(X = 1)$, then the variance of the prior belief distribution is given by $P(X = 1)(1 - P(X = 1))$. By processing information, the consumer receives a signal S based on which she updates her prior belief to some posterior beliefs. If $S = 1$, then the variance of her posterior belief distribution is given by $P(X = 1|S = 1)(1-P(X = 1|S = 1))$; if $S = 0$, then the variance of her posterior belief distribution is given by $P(X = 1|S = 0)(1 - P(X = 1|S = 0))$. Note that the expected variance of posterior belief distributions is always weakly smaller than the variance of prior belief distribution. That is, $P(S = 1)P(X = 1|S = 1)(1-P(X = 1|S = 1))+P(S = 0)P(X = 1|S = 0)(1-P(X = 1|S = 0)) \leq P(X = 1)(1-P(X = 1))$. Now, we can quantify the cost of obtaining a particular signal by the reduction of variance of belief distribution $P(X = 1)(1-P(X = 1))-\{P(S = 1)P(X = 1|S = 1)(1-P(X = 1|S = 1))+P(S = 0)P(X = 1|S = 0)(1-P(X = 1|S = 0))\}$. Using this variance-based metric does not lead to closed-form solution but numerical analysis shows



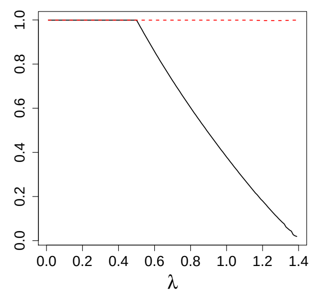
(a) Case I: High q and $U_1 < |U_0|$



(b) Case II: Low q and $U_1 > |U_0|$



(c) Case III: High q and $U_1 > |U_0|$



(d) Case IV: Low q and $U_1 < |U_0|$

Figure 17: Impact of information processing cost on attention allocation under alternative cost information (red dotted line is for δ_0 and black solid line is for δ_1)

that it would not change our main results qualitatively. Figure 17 shows the consumer's attention allocation behavior under this alternative cost function, which is qualitatively similar to the results obtained using entropy cost function (See Figure 3 in Section 3.2).

Finally, one may also come up with other appropriate ad-hoc cost functions. However, it is important to note that the potential ad-hoc cost functions at least should satisfy two basic properties: (1) The cost should be 0 when the signal is uninformative (i.e., $\delta_1 + \delta_0 = 1$); (2) the cost function should be a function of both the signal structure and the consumer's prior belief (that is, if the consumer is a priori more certain about the product match, then the amount of information contained in a signal should be smaller). Given these two basic properties, it turns out not to be an easy task to come up with an appropriate ad-hoc cost function. On the contrary, the entropy-based cost function that I have used is derived from some intuitive axioms and has been widely accepted and adopted in multiple disciplines where information processing is considered.

5.2 Discussion about practical implications and experimental test

Practical implications

In marketing, researchers have found in a variety of contexts that consumers often pay asymmetric attention to different product information. For example, Hoch and Ha (1986) find that advertising induces consumers to conduct confirmatory information search for the advertised brand. The explanation given by the authors is that advertising forms a tentative hypothesis that the product is good for the consumer, and the consumer tends to confirm that hypothesis by searching more positive information that supports this hypothesis. In addition, John, Scott and Bettman (1986) find that consumers tend to conduct confirmatory search for covariation assessment. That is, those consumers who believed that price and quality are positively related elected to sample higher-priced products than consumers who believed that there is little relationship between price and quality. In contrast to the previous research, my research shows that the confirmatory search could be optimal behavior coming out of the consumer optimizing attention to different types of information. Therefore, it offers a rational explanation for consumer confirmatory search, which is complementary to the psychological explanations proposed in the previous research. More importantly, I further characterize the conditions under which confirmatory search and disconfirmatory search are optimal respectively. This deepens our understanding and provides some new predictions on consumer information search behavior that can be tested empirically or experimentally (In the next section, I propose a basic idea about experimental testing of these theoretical predictions.).

In addition to consumer information search, this research also sheds light on firm strategies. For example, I show that when the consumer wants to search information about product match, lowering the search cost could decrease the consumer's price sensitivity, allowing the seller to charge a higher price. This may explain why some online retailers tend to make the product information search on their websites as easy as possible. For instance, when a consumer is searching hotels on Booking.com and is browsing the product page for a particular hotel, the website automatically analyzes all of the previous customer reviews written for this hotel and offers an overview of the hotel features frequently mentioned in the customer reviews. Plausibly, this practice may decrease the consumer's search cost by allowing the consumer to obtain the most important product information in an easier way. Therefore, according to the discussion in Chapter 4, this practice can decrease the consumer's price sensitivity and allows the sellers to charge a higher price, which also benefits the online retailer.

Furthermore, my research also shows that when the consumer has a high information processing cost, the seller may want to charge a relatively low price, inducing the consumer to buy the product without any learning. Some observations suggest that this type of pricing strategy may also be used in practice. For example, when booking a flight on a website such as Expedia.com, consumers are often pitched at the stage of payment with some add-on products, such as insurance, pick-up services or hotels. Plausibly, the consumer may not want to further search the product information for these add-on products at the stage of payment, perhaps because searching the flights has already consumed a lot of the consumer's mental resources and time. In other words, the consumer has a high information processing cost for searching information on these add-on products. Given this, my research may explain the cursory observation that the website tends to pitch the consumers with relatively *cheap* add-on products among the alternatives. The reason is that pitching with a cheap add-on product is more likely to induce the consumers who have a high information processing cost to buy the add-on immediately without further learning.

As mentioned in the previous chapter, firms can also design the information environment to increase the consumer's purchase likelihood. For example, a seller can choose how much information and which information to be available in the advertising or on the product page. Intuitively, a seller has an inherent motivation to induce the consumer to pay more attention to positive information of the product and pay less attention to negative information. Therefore, the seller may want to suppress the negative product reviews on the reviews sites or does not disclose the information of some certain product attributes which may dissatisfy consumers with high probability. Furthermore, when designing the information environment, the seller may take into account the consumer attention allocation behavior because consumers often pay asymmetric attention to different types of information with some information being totally ignored. In my research, I find that when the information processing cost is high for consumers, the information environment designed by the seller is quite asymmetric (i.e., disproportionately more positive information under high prior belief and disproportionately more negative information under low prior belief). One may observe this phenomenon or test this prediction in the context where an online seller designs both the mobile store and the desktop store. In this context, if on average consumers have a higher information processing cost when visiting the mobile store than when visiting the desktop store, then we may observe that the seller provides less information in a more asymmetric way in his mobile store than in the desktop store.

Moreover, I also find that the seller may charge a lower price when he can control the information environment than when he can not. This may offer a new explanation for the observation that some new products (e.g., a new restaurant, a new education program, etc.) tend to charge a lower price (e.g., offer price discount) in the beginning than in the later periods.[6] Specifically, in the beginning, it is relatively easy for the seller to control the information environment by advertising or by the salesman's efforts to persuade the consumers to purchase the product. In later periods, the seller may lose the control of the information environment. For instance, more customer reviews may become available in

[6] An alternative explanation is that charging a lower price can invite more consumers to try the product in the early periods, which can grow the brand awareness more quickly.

reviews sites and thus the potential consumers can obtain information from the reviews sites, decreasing the seller's capability of influencing the information environment. Therefore, according to the discussion in Section 4.3, the seller has the incentive to charge a lower price in the early periods to increase the consumer's learning incentive. This increases the seller's capability of persuading the consumer to purchase the product by information design, which increases the consumer's purchase likelihood.

To further examine the theoretical implications made in my research, I propose a basic idea in the following about testing consumer attention allocation behavior and the seller behavior in an experimental environemnt.

Experimental testing

One way to test the theoretical implications in my research is to simulate an online shopping environment in the experiment where a typical product page includes both overall rating for the product and specific product reviews. Specifically, in the experiment, the subjects are asked to make a purchase decision for a certain product which is actually sold on an online retailer such as Amazon (e.g., a mug or a computer mouse). Before making a purchase decision, the subject can read the overall rating and specific product reviews about this product on the product page. If she decides to purchase, then she will get this product with some probability; if she decides not to buy, then she can participate in a lottery and win a certain amount of money with some probability, which serves as an outside option for the subject.[7] Note that we can elicit two conditions for product utilities: (1) Large expected gain from the lottery (i.e., the utility of outside option is high), and (2) small expected gain from the lottery (i.e., the utility of outside option is low). Under the first condition, the gain of buying a matched product is relatively small as compared to the loss of buying an unmatched product; Under the second condition, the gain of buying a matched product is

[7]Depending on the research budget and the product category, we can make the outside option to be either stochastic or deterministic.

relatively large as compared to the loss of buying an unmatched product.

Next, in terms of prior belief, it is reasonable to assume that the consumer's prior belief is mainly influenced by the overall rating of the product. That is, a high overall rating elicits a high prior belief , while a low overall rating elicits a low prior belief. To better elicit the prior belief, the section for specific product reviews can be initially folded on the product page and the subject needs to click to unfold the product reviews section. In other words, the product page initially contains only the overall rating and relevant product description, which is assumed to form the subject's prior belief. Then, if the subject chooses to unfold the product reviews section, the specific product reviews would appear on the product page.

Furthermore, to influence the consumer's information processing cost, we can design at least two conditions: (1) The subject is given enough time to make a purchase decision; (2) the subject must make a purchase decision in a short period of time. The idea is that the consumer's information processing cost would be influenced by the time pressure. Plausibly, under the first condition, the subject's information processing cost would be low; however, under the second condition, the subject's information processing cost would be high. Another way to influence the consumer's information processing cost is to manipulate the subject's mental resources by asking the subjects to memorize and recall a sequence of numbers before and after the main task, which is typically used in psychology literature.

Given the manipulation of the subject's prior belief, product utilities, and the information processing cost, we can track the subject's attention on those product reviews by monitoring his or her eye-movement using eye-tracking devices or monitoring the mouse-clicking activities. Therefore, the amount of attention paid to positive reviews and negative reviews can be measured by the amount of time spent on each type of reviews respectively.

The basic experiment roughly discussed above can be used to test the theoretical implications derived in my dissertation. First, as I mentioned earlier, when the consumer has a higher prior belief about product fit, she would have a higher incentive to avoid the mistake of not buying a good product when the product indeed matches her needs, and thus she would pay relatively more attention to positive information than negative information. Therefore, we have the following testable hypothesis:

Hypothesis 1. *If the product has a higher overall rating, the subjects would pay relatively more attention to positive reviews than negative ones, and vice versa.*

Furthermore, one of the main results on consumer attention is that the consumer would do more confirmatory search under a high information processing cost and may do more disconfirmatory search under a low information processing cost. This leads to the following hypotheses:

Hypothesis 2a. *If the product has a high overall rating, then the subjects in the high information processing cost condition would pay relatively more attention to positive reviews and less attention to negative reviews than those subjects in the low information processing cost condition;*

Hypothesis 2b. *If the product has a low overall rating, then the subjects in the high information processing cost condition would pay relatively more attention to negative reviews and less attention to positive reviews than those subjects in the low information processing cost condition.*

Hypothesis 3a. *If the product has a high overall rating and the utility of outside option is large, then the subjects in the high information processing cost condition would pay relatively more attention to positive reviews than negative reviews (i.e., confirmatory search), while the subjects in the low information processing cost condition would pay relatively more attention to negative reviews than positive reviews (i.e., disconfirmatory search).*

Hypothesis 3b. *If the product has a low overall rating and the utility of outside option is low, then the subjects in the high information processing cost condition would pay relatively more attention to negative reviews than positive reviews (i.e., confirmatory search), while the subjects in the low information processing cost condition would pay relatively more attention to positive reviews than negative reviews (i.e., disconfirmatory search).*

Regarding the firm strategies, to introduce the price factor into the purchase decision, we can endow the subjects with a certain amount of virtual money at the beginning which can be used to purchase the product or to participate in the lottery (outside option). The idea is that if the subject spends some portion of virtual money on purchasing the product, then the remaining money can be used to play the lottery (more remaining money means higher stake in the lottery). By doing so, we can elicit the impact of product price on consumer's purchase decision. Note that in the experiment, the subjects would see the price on the product page all the time since the very beginning. According to the discussion in Chapter 4, we have the following hypotheses to test on firm strategies.

Hypothesis 4. *A higher product price would induce subjects to pay more attention to negative reviews and less attention to positive reviews. This impact of a higher price on attention is stronger for the subjects under high information processing cost condition than for the subjects under low information processing cost condition.*

Note that if the experimental results support the above Hypothesis 4, then it would further lend support to the implication that a higher information processing cost can increase the consumer's price sensitivity and thus force the firm to charge a lower price. Next, to examine the implications on firm information design, we can extend our above experiment by dividing the subjects into two groups: Consumers and sellers. The basic idea is the following: we first allow the "sellers" to decide on the product price and also choose which product reviews to show on the product page (the researcher first selects the real product reviews for the product on Amazon and then asks the "sellers" in the experiment to choose which product reviews to show on the simulated product page.). As in the previous experiment, the "consumers" in the experiment are endowed with a certain amount of virtual money at the beginning which can be used to purchase a product or to participate in the lottery (outside option) if he or she decides not to purchase a product. A seller's goal is to obtain as much virtual money as possible (i.e., maximize the profit) from the consumers, and then the seller can use such money to play lottery afterwards.). Note that a seller is randomly assigned with a product from the sample and the seller understands the consumer's situation and knows that there would be a bunch of consumers visiting the product page and there is no competition between sellers. With this proposed experiment, we can test the following hypotheses on firm information design and pricing.

Hypothesis 5. *If the seller is assigned to a product with a lower overall rating, then the seller would increase the proportion of negative product reviews in the environment.*

Hypothesis 6. *The seller would charge a lower price if the seller can choose which product reviews are available for consumers to process than if the seller can not.*

Note that to test Hypothesis 6, the sellers in the experiment are randomly assigned to either the condition where they can choose the product reviews on the product page or the condition where they can not. The discussions above only give a basic idea about the experimental tests of theoretical implications obtained in my model and more details about the experiment are needed to be figured out in the future.

5.3 Conclusions

Facing massive amounts of product information, consumers often have to incur nontrivial costs including time and mental efforts to process information. Therefore, they often need to make choices regarding the subjects to which they pay more or less attention. In this paper, I have studied several questions related to consumer attention allocation: How would a consumer optimize attention to various product information before making a purchase decision? How is this attention allocation decision influenced by some interesting factors such as the consumer's prior belief, information processing cost, and the credibility of information? How does the seller design the marketing strategies taking into account the consumer's attention allocation? To answer these questions, I build an analytical model in which a consumer's attention allocation decision is captured by her choice of signal structure and she chooses the optimal signal structure by trading off the value and cost of information, where the cost of information is measured using information theory (Shannon 1948).

My research sheds new lights on consumer information search behavior and provides interesting implications for marketing strategies. Specifically, I find that a consumer conducts either confirmatory search or disconfirmatory search. In particular, a consumer may engage in more disconfirmatory search when she has a low information processing cost but engage in more confirmatory search when she has a high cost. This result suggests that the "confirmatory bias" behavior widely studied in psychology literature could be an optimal behavior coming out of people optimizing attention to different types of information, especially when people has high cognitive limitations and thus high information processing costs. Next, I find that the consumer's purchase likelihood may vary with her information processing cost in a non-monotonic way, depending on the consumer's prior belief and the utilities of buying a matched and mismatched product. Moreover, I show that when more information becomes available or credible, the consumer would increase attention to negative information when the prior utility of the product is high but she would increase attention to positive information when the prior utility is low.

In terms of firm strategies, I find that under a low information processing cost, the seller would charge a relatively high price such that consumers always process information; but under a high information processing cost, the seller would charge a relatively low price such that consumers purchase the product without any learning. The optimal price and profit would first decrease and then increase in consumer's information processing cost. Furthermore, offering the return policy induces the consumer to pay more attention to positive information and less attention to negative information, and the seller would offer such return policy except when the consumer has a very high information processing cost. Finally, when a seller can influence the information environment, he would have a lower incentive to suppress the negative information when the consumer has a lower prior belief about product fit. Besides, a higher information processing cost for a consumer would increase or decrease a seller's incentive to suppress the negative information in the environment, depending on whether the seller can adjust the product price and whether the consumer has a high or low prior belief. Interestingly, the seller may charge a lower price when he can fully control the information environment than when he can not.

Future research could extend current work by examining how the consumer's attention allocation influences the competition among many sellers. Intuitively, suppose a consumer is deciding which one among several alternatives better fits her. When the information processing cost is low, the consumer has a high incentive to process information and there may exist an equilibrium where the firms charge prices such that the consumer always processes information. In particular, as the information processing cost becomes higher but not too high, the demand sensitivity to price becomes higher as discussed earlier and thus the price competition may become more severe.

Furthermore, although Shannon mutual information has been widely adopted theoretically and empirically in the literature as the measure of information processing cost, it would be interesting to understand and account for systematic deviations in consumers' information processing cost from the mutual information based formulation that I have used in the dissertation.

References

Anderson, S. P., & Renault, R. (1999) "Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model," *RAND Journal of Economics*, 30(4), 719–735.

Bakos, J. Y. (1997) "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," *Management Science*, 43(12), 1676–1692.

Bergemann, D., & Morris, S. (2016) "Information design, bayesian persuasion and bayes correlated equilibrium," *The American Economic Review*,106 (5), 586–591.

Bergemann, D., & Morris, S. (2017) "Information design: A unified perspective," *Working Paper*, Yale University.

Branco, F., Sun, M., & Villas-Boas, J. M. (2012) "Optimal Search for Product Information," *Management Science*, 58(11), 2037–2056.

Branco, F., Sun, M., & Villas-Boas, J. M. (2016) "Too Much Information? Information Provision and Search Costs," *Marketing Science*, 35(4), 605–618.

Caplin, A., & Dean, M. (2013) "Rational Inattention and State Dependent Stochastic Choice," Working paper, New York University. http://stanford.io/2ofXrM3.

Cheremukhin, A., Popova A., & Tutino A. (2011)"Experimental Evidence of Rational Inattention," Working Paper, Federal Reserve Bank of Dallas.

Cover, T. M. & Thomas, J. A. (2006) Elements of Information Theory. Wiley, Hoboken, NJ.

Csiszár, I. (2008). "Axiomatic characterizations of information measures," *Entropy*, 10(3), 261–273.

Dessein, W., Galeotti, A. & Santos, T. (2016) "Rational Inattention and Organizational Focus," *American Economic Review*, 106(6), 1522–1536.

Dewan, A. & Nathaniel N. (2017) "Estimating Information Cost Functions in Models of Rational Inattention," Working paper, Columbia University.

Diamond, P. A. (1971) "A Model of Price Adjustment," *Journal of Economic Theory*, 3(2), 156–168.

Ellison, G. & Wolitzky A. (2012) "A Search Cost Model of Obfuscation," *RAND Journal of Economics*, 43(3), 417–441.

Evans, J. S. B. (1989). Bias in human reasoning: Causes and consequences. Lawrence Erlbaum Associates, Inc.

Gardete, P. M. (2013) "Cheap-Talk Advertising and Misrepresentation in Vertically Differentiated Markets," *Marketing Science*, 32(4), 609–621.

Gentzkow, M. & Kamenica, E. (2016)"Competition in persuasion," *The Review of Economic Studies*, 84(1), 300–322.

Guo, L., & Zhang, J. (2012) "Consumer Deliberation and Product Line Design," *Marketing Science*, 31(6), 995–1007.

Hébert, B., & Woodford, M. (2017) "Rational Inattention and Sequential Information Sampling," Working Paper, National Bureau of Economic Research.

Hoch, S. J., & Ha, Y. W. (1986) "Consumer learning: Advertising and the ambiguity of product experience," *Journal of consumer research*, 13(2), 221-233.

John, D. R., Scott, C. A., & Bettman, J. R. (1986) "Sampling data for covariation assessment: The effect of prior beliefs on search patterns," *Journal of Consumer Research*, 13(1), 38-47.

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

Ke, T. T., Shen, Z. J. M., & Villas-Boas, J. M. (2016) "Search for Information on Multiple Products," *Management Science*, 62(12), 3576–3603.

Kuksov, D. (2004) "Buyer Search Costs and Endogenous Product Design," *Marketing Science*, 23(4), 490–499.

Kuksov, D., & Villas-Boas, J. M. (2010) "When More Alternatives Lead to Less Choice," *Marketing Science*, 29(3), 507–524.

Lal, R., & Sarvary, M. (1999) "When and How is the Internet Likely to Decrease Price Competition?" *Marketing Science*, 18(4), 485–503.

Liu, L., & Dukes, A. (2016) "Consumer Search with Limited Product Evaluation," *Journal of Economics & Management Strategy*, 25(1), 32–55.

Luo, Y. (2008) "Consumption Dynamics Under Information Processing Constraints," *Review of Economic Dynamics*, 11(2), 366–385.

Luo, Y., & Young, E. R. (2009) "Rational Inattention and Aggregate Fluctuations," *The BE Journal of Macroeconomics*, 9(1), 14.

Maćkowiak, B., & Wiederholt, M. (2009) "Optimal Sticky Prices Under Rational Inattention," *American Economic Review*, 99(3), 769–803.

Maćkowiak, B., & Wiederholt, M. (2015) "Business Cycle Dynamics Under Rational Inattention," *Review of Economic Studies*, 82(4), 1502–1532.

Martin, D. (2017) "Strategic Pricing and Rational Inattention to Quality," *Games and Economic Behavior*, 104, 131–145.

Matějka, F., & McKay, A. (2015) "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105(1), 272–298.

Matějka, F. (2016) "Rationally Inattentive Seller: Sales and Discrete Pricing," *Review of Economic Studies*, 83(3), 1125–1155.

Mayzlin, D., & Shin, J. (2011) "Uninformative Advertising as an Invitation to Search," *Marketing Science*, 30(4), 666–685.

Mondria, J. (2010) "Portfolio choice, attention allocation, and price comovement," *Journal of Economic Theory*, 145(5), 1837–1864.

Morris, S., & Strack, P. (2017) "The Wald problem and the equivalence of sequential sampling and static information costs," Working paper.

Nickerson, R. S. (1998) "Conselleration Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology*, 2(2), 175–220.

Paciello, L., & Wiederholt, M. (2013) "Exogenous Information, Endogenous Information, and Optimal Monetary Policy," *Review of Economic Studies 81(1)*, 356–388.

Quiroga, R. Q., & Panzeri, S. (2009) "Extracting Information from Neuronal Populations: Information Theory and Decoding Approaches," *Nature Reviews Neuroscience*, 10(3), 173–185.

Rolls, E. T., & Treves, A. (2011) "The Neuronal Encoding of Information in the Brain," *Progress in Neurobiology*, 95(3), 448–490.

Shannon, C. E. (1948) "A Mathematical Theory of Communication," *Bell System Technical Journal*, 379–423.

Sims, C. A. (1998) "Stickiness," In Carnegie-Rochester Conference Series on Public Policy, 49, 317–356).

Sims, C. A. (2003) "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665–690.

Sims, C. A. (2006) "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review*, 96(2), 158–163.

Stahl, D. O. (1989) "Oligopolistic Pricing with Sequential Consumer Search," *American Economic Review*, 79(4), 700–712.

Steiner, J., Stewart, C., & Matějka, F. (2017) "Rational Inattention Dynamics: Inertia and Delay in Decision Making," *Econometrica*, 85(2), 521–553.

Villas-Boas, J. M. (2009) "Product Variety and Endogenous with Evaluation Costs," *Management Science*, 55(8), 1338–1346.

Wald, A. (1945) "Sequential method of sampling for deciding between two courses of action," *Journal of the American Statistical Association*, 40(231), 277-306.

Weitzman, M. L. (1979) "Optimal search for the best alternative," *Econometrica: Journal of the Econometric Society*, 641-654.

Woodford, M. (2014) "An Optimizing Neuroeconomic Model of Discrete Choice," Working paper, Columbia University. http://www.nber.org/papers/w19897.

Yang, M. (2011) "Coordination with Rational Inattention," Working Paper, Economic Theory Center. https://ssrn.com/abstract=1858728.

APPENDIX A

Proof of Proposition 1

The consumer solves the following optimization problem.

$$\begin{aligned} EU^* \equiv & \sup_{0 \leq \delta_1 \leq 1, 0 \leq \delta_0 \leq 1} P(S=1)[P(X=1|S=1)U_1 + P(X=0|S=1)U_0] - \lambda I(X,S) \\ \text{s.t.} \quad & \delta_1 + \delta_0 > 1 \end{aligned}$$

where

$$\begin{aligned} I(X,S) &= -[q\delta_1 + (1-q)(1-\delta_0)] \log[q\delta_1 + (1-q)(1-\delta_0)] \\ & \quad -[q(1-\delta_1) + (1-q)\delta_0] \log[q(1-\delta_1) + (1-q)\delta_0] \\ & \quad +q[\delta_1 \log \delta_1 + (1-\delta_1) \log(1-\delta_1)] \\ & \quad +(1-q)[\delta_0 \log \delta_0 + (1-\delta_0) \log(1-\delta_0)] \end{aligned}$$

Note that the objective function is strictly concave function and thus the first order condition is sufficient for finding the unique global optimal solution.I first ignore the constraints (i.e., $0 \leq \delta_1 \leq 1,\ 0 \leq \delta_0 \leq 1,\ \delta_1 + \delta_0 > 1$). The first-order conditions with respect to δ_1 and δ_0 are given as follows

$$e^{\frac{U_1}{\lambda}} = \frac{q + (1-q)\frac{\delta_0}{1-\delta_1}}{q + (1-q)\frac{1-\delta_0}{\delta_1}} \text{ and } e^{\frac{-U_0}{\lambda}} = \frac{1-q+q\frac{\delta_1}{1-\delta_0}}{1-q+q\frac{1-\delta_1}{\delta_0}}. \tag{20}$$

Solving this, I obtain:

$$\delta_1^* = \frac{1 - \frac{1-q}{q} \frac{e^{\frac{-U_0}{\lambda}} - 1}{e^{\frac{U_1 - U_0}{\lambda}} - e^{\frac{-U_0}{\lambda}}}}{1 - e^{-\frac{U_1 - U_0}{\lambda}}} \text{ and } \delta_0^* = \frac{1 - \frac{q}{1-q} \frac{e^{\frac{U_1-U_0}{\lambda}} - e^{\frac{-U_0}{\lambda}}}{e^{\frac{U_1-U_0}{\lambda}} (e^{\frac{-U_0}{\lambda}} - 1)}}{1 - e^{-\frac{U_1-U_0}{\lambda}}}$$

Now, one can easily check that when $\frac{e^{\frac{-U_0}{\lambda}}-1}{e^{\frac{U_1-U_0}{\lambda}}-1} < q < \frac{1-e^{\frac{U_0}{\lambda}}}{1-e^{-\frac{U_1-U_0}{\lambda}}}$, $0 < \delta_1^* < 1$, $0 < \delta_0^* < 1$, $\delta_1^* + \delta_0^* > 1$ and $P(S=1) = \frac{q}{1-e^{-\frac{-U_0}{\lambda}}} - \frac{1-q}{e^{\frac{U_1}{\lambda}}-1}$. When $q \leq \frac{e^{\frac{-U_0}{\lambda}}-1}{e^{\frac{U_1-U_0}{\lambda}}-1}$, the utility of outside option, 0, is larger than EU^* and prior utility $EU_0 \equiv qU_1 + (1-q)U_0$. Therefore, the consumer processes no information and chooses the outside option. When $q \geq \frac{1-e^{\frac{U_0}{\lambda}}}{1-e^{-\frac{U_1-U_0}{\lambda}}}$, the prior utility $EU_0 \equiv qU_1 + (1-q)U_0$ is larger than EU^* and the utility of outside option. Therefore, the consumer processes no information and purchases the product. *Q.E.D.*

Proof of Proposition 2

According to proposition 1, given $\underline{q} < q < \overline{q}$, $\delta_1^* \geq \delta_0^*$ if and only if

$$\frac{1-q}{q}\frac{e^{\frac{-U_0}{\lambda}}-1}{1-e^{-\frac{U_1}{\lambda}}} \leq \frac{q}{1-q}\frac{e^{\frac{U_1}{\lambda}}(1-e^{-\frac{U_1}{\lambda}})}{1-e^{\frac{U_0}{\lambda}}}$$

Note that $\frac{1-q}{q}\frac{e^{\frac{-U_0}{\lambda}}-1}{1-e^{-\frac{U_1}{\lambda}}} \leq \frac{q}{1-q}\frac{e^{\frac{U_1}{\lambda}}(1-e^{-\frac{U_1}{\lambda}})}{1-e^{\frac{U_0}{\lambda}}}$ is equivalent to $\frac{1-q}{q}(e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) \leq \frac{q}{1-q}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}})$. Now denote $H_1 \equiv \frac{1-q}{q}(e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}})$ and $H_0 \equiv \frac{q}{1-q}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}})$. It is not hard to check that both H_1 and H_0 decrease with λ. In particular, $H_1 \to 0$ and $H_0 \to 0$ as $\lambda \to \infty$; $H_1 \to \infty$ and $H_0 \to \infty$ as $\lambda \to 0$. Furthermore, one can prove that $\frac{dH_1}{d\lambda} > \frac{dH_0}{d\lambda}$ if $\frac{U_1}{\lambda} > \frac{-U_0}{\lambda}$, $\frac{dH_1}{d\lambda} < \frac{dH_0}{d\lambda}$ if $\frac{U_1}{\lambda} < \frac{-U_0}{\lambda}$, and $\frac{dH_1}{d\lambda} = \frac{dH_0}{d\lambda}$ if $\frac{U_1}{\lambda} = \frac{-U_0}{\lambda}$.

First, consider the case $\frac{U_1}{\lambda} > \frac{-U_0}{\lambda}$ (i.e., $U_1 > -U_0$). According to the above properties of H_1 and H_0, it is not hard to prove that there must exist a unique solution, λ^*, of equation $H_1 = H_0$, and for $\lambda \in (0, \lambda^*)$, $H_1 < H_0$ and thus $\delta_1^* > \delta_0^*$; for $\lambda \in (\lambda^*, \infty)$, $H_1 > H_0$ and thus $\delta_1^* < \delta_0^*$. From proposition 1, I know that $\bar{q} = \frac{1-e^{\frac{U_0}{\lambda}}}{1-e^{-\frac{U_1-U_0}{\lambda}}}$ and $\underline{q} = \frac{e^{\frac{-U_0}{\lambda}}-1}{e^{\frac{U_1-U_0}{\lambda}}-1}$. It is easy to check that $\bar{q}$ decreases with λ and $\underline{q}$ increases with λ. In particular, as $\lambda \to 0$, $\underline{q} \to 0$ and $\bar{q} \to 1$; as $\lambda \to \infty$, $\underline{q} \to \frac{-U_0}{U_1-U_0}$ and $\bar{q} \to \frac{-U_0}{U_1-U_0}$. Now if $q > \frac{-U_0}{U_1-U_0}$, then denote $\bar{\lambda}$ as the unique solution of equation $q = \frac{1-e^{\frac{U_0}{\lambda}}}{1-e^{-\frac{U_1-U_0}{\lambda}}}$. Note that if $0 < \lambda < \bar{\lambda}$, then the consumer would process information and the attention allocation strategy is given by δ_1^* and δ_0^* in proposition 1; if $\lambda > \bar{\lambda}$, then the consumer would not process information and would always buy the product. In the following, I will prove that $\bar{\lambda} < \lambda^*$ for $q > \frac{-U_0}{U_1-U_0}$ and $\bar{\lambda} > \lambda^*$ for $q < \frac{-U_0}{U_1-U_0}$. To see this, I first consider the scenario where $q > \frac{-U_0}{U_1-U_0}$. In this scenario, I can prove that $\frac{1-\bar{q}}{\bar{q}}(e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) \leq \frac{\bar{q}}{1-\bar{q}}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}})$. To see this, note that

$$
\begin{aligned}
\frac{1-\bar{q}}{\bar{q}}(e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) &\leq \frac{\bar{q}}{1-\bar{q}}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}}) \\
\leftrightarrow (e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) &\leq (\frac{\bar{q}}{1-\bar{q}})^2(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}}) \\
\leftrightarrow (e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) &\leq \frac{(\frac{1-e^{\frac{U_0}{\lambda}}}{1-e^{-\frac{U_1-U_0}{\lambda}}})^2}{(\frac{e^{\frac{U_0}{\lambda}}-e^{-\frac{U_1-U_0}{\lambda}}}{1-e^{-\frac{U_1-U_0}{\lambda}}})^2}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}}) \\
\leftrightarrow (e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) &\leq \frac{(e^{\frac{-U_0}{\lambda}}-1)^2}{(1-e^{\frac{-U_1}{\lambda}})^2}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}}) \\
\leftrightarrow \frac{1-e^{\frac{U_0}{\lambda}}}{e^{\frac{-U_0}{\lambda}}-1} &\leq \frac{1-e^{\frac{U_1}{\lambda}}}{e^{\frac{-U_1}{\lambda}}-1}
\end{aligned}
$$

Since $\frac{-U_1}{\lambda} < 0 < \frac{-U_0}{\lambda}$ and $-U_1 < U_0$, it is easy to see that $\frac{1-e^{\frac{U_0}{\lambda}}}{e^{\frac{-U_0}{\lambda}}-1} \leq \frac{1-e^{\frac{U_1}{\lambda}}}{e^{\frac{-U_1}{\lambda}}-1}$ always. Therefore, it must always be that $\frac{1-\bar{q}}{\bar{q}}(e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) \leq \frac{\bar{q}}{1-\bar{q}}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}})$. This implies that at $\lambda = \bar{\lambda}$, $H_1 < H_0$ and according to the properties of H_1 and H_0, it must be that $\bar{\lambda} < \lambda^*$ and thus $H_1 < H_0$ (*i.e.*, $\delta_1^* > \delta_0^*$) for any $\lambda < \bar{\lambda}$. Next, consider the scenario where $q < \frac{-U_0}{U_1-U_0}$. In this scenario, I can prove that $\frac{1-\underline{q}}{\underline{q}}(e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) > \frac{\underline{q}}{1-\underline{q}}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}})$. To see this, note that

$$
\begin{aligned}
\frac{1-\underline{q}}{\underline{q}}(e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) &> \frac{\underline{q}}{1-\underline{q}}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}}) \\
\leftrightarrow (e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) &> (\frac{\underline{q}}{1-\underline{q}})^2(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}}) \\
\leftrightarrow (e^{\frac{-U_0}{\lambda}}-1)(1-e^{\frac{U_0}{\lambda}}) &> \frac{(\frac{e^{\frac{-U_0}{\lambda}}-1}{e^{\frac{U_1-U_0}{\lambda}}-1})^2}{(\frac{e^{\frac{U_1-U_0}{\lambda}}-e^{\frac{-U_0}{\lambda}}}{e^{\frac{U_1-U_0}{\lambda}}-1})^2}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}}) \\
\leftrightarrow e^{\frac{-U_0}{\lambda}} &> e^{-\frac{U_1}{\lambda}}
\end{aligned}
$$

Note that since $\frac{-U_1}{\lambda} < 0 < \frac{-U_0}{\lambda}$, it must always be that $e^{\frac{-U_0}{\lambda}} > e^{-\frac{U_1}{\lambda}}$ and thus $\frac{1-\underline{q}}{\underline{q}}(e^{\frac{-U_0}{\lambda}} - 1)(1-e^{\frac{U_0}{\lambda}}) > \frac{\underline{q}}{1-\underline{q}}(e^{\frac{U_1}{\lambda}}-1)(1-e^{-\frac{U_1}{\lambda}})$. This implies that at $\lambda = \bar{\lambda}$, $H_1 > H_0$ (i.e., $\delta_1^* < \delta_0^*$), and according to the properties of H_1 and H_0, it must be that $\bar{\lambda} > \lambda^*$. Therefore, for $0 < \lambda < \lambda^*$, $H_1 < H_0$ (i.e., $\delta_1^* > \delta_0^*$); for $\lambda^* < \lambda < \bar{\lambda}$, $H_1 > H_0$ (i.e., $\delta_1^* < \delta_0^*$).

Similarly, one can prove the results for the case where $\frac{U_1}{\lambda} < \frac{-U_0}{\lambda}$ (i.e., $U_1 < -U_0$). Q.E.D.

Proof Proposition 3

According to Proposition 1, $\frac{dP(S=1)}{d\lambda} = (q\frac{-e^{\frac{U_0}{\lambda}}\frac{U_0}{\lambda}}{(e^{\frac{U_0}{\lambda}}-1)^2} - (1-q)\frac{e^{\frac{U_1}{\lambda}}\frac{U_1}{\lambda}}{(e^{\frac{U_1}{\lambda}}-1)^2})\frac{1}{\lambda}$. Therefore, denote $\Delta = \frac{\frac{e^{\frac{U_1}{\lambda}}\frac{U_1}{\lambda}}{(e^{\frac{U_1}{\lambda}}-1)^2}}{\frac{-e^{\frac{U_0}{\lambda}}\frac{U_0}{\lambda}}{(e^{\frac{U_0}{\lambda}}-1)^2}}$, if $\frac{q}{1-q} > \Delta$, then $\frac{dP(S=1)}{d\lambda} > 0$; otherwise, if $\frac{q}{1-q} < (=)\Delta$, then $\frac{dP(S=1)}{d\lambda} < (=)0$. Now, it is not hard to see that $\Delta \to \frac{-U_0}{U_1}$ as $\lambda \to \infty$. Furthermore, if $U_1 > (<)|U_0|$, then Δ increases (decreases) with λ. Thus, given $U_1 > |U_0|$, if $\frac{q}{1-q} > \frac{-U_0}{U_1}$ (i.e., $q > \frac{-U_0}{U_1-U_0}$), then it must always be that $\frac{q}{1-q} > \Delta$, which implies $\frac{dP(S=1)}{d\lambda} > 0$; but if $\frac{q}{1-q} < \frac{-U_0}{U_1}$ (i.e., $q < \frac{-U_0}{U_1-U_0}$), there exists $\hat{\lambda}$ such that for $\lambda < \hat{\lambda}$, $\frac{dP(S=1)}{d\lambda} > 0$ and for $\lambda > (=)\hat{\lambda}$, $\frac{dP(S=1)}{d\lambda} < (=)0$ ($\hat{\lambda}$ is the unique solution of $\frac{q}{1-q} = \Delta$). On the other hand, given $U_1 < |U_0|$, if $\frac{q}{1-q} < \frac{-U_0}{U_1}$ (i.e., $q < \frac{-U_0}{U_1-U_0}$), then it must always be that $\frac{q}{1-q} < \Delta$, which implies $\frac{dP(S=1)}{d\lambda} < 0$; but if

$\frac{q}{1-q} > \frac{-U_0}{U_1}$ (i.e., $q > \frac{-U_0}{U_1-U_0}$), then there exists $\hat{\lambda}$ such that for $\lambda < \hat{\lambda}$, $\frac{dP(S=1)}{d\lambda} < 0$ and for $\lambda > (=)\hat{\lambda}$, $\frac{dP(S=1)}{d\lambda} > (=)0$ ($\hat{\lambda}$ is the unique solution of $\frac{q}{1-q} = \Delta$.). Q.E.D.

Proof for Proposition 4

Since$P(X=1) = P(X=0) = \frac{1}{2}$, $P(Y=1) = P(Y=0) = \frac{1}{2}\alpha + \frac{1}{2}(1-\alpha) = \frac{1}{2}$.

$$H(S) = -\{\Delta \log \Delta + (1-\Delta)\log(1-\Delta)\} \text{ where } \Delta \equiv P(S=1) = \tfrac{1}{2}\delta_1 + \tfrac{1}{2}(1-\delta_0).$$

$$H(S|X) = -\{\tfrac{1}{2}[\delta_1 \log \delta_1 + (1-\delta_1)\log(1-\delta_1)] + \tfrac{1}{2}[\delta_0 \log \delta_0 + (1-\delta_0)\log(1-\delta_0)]\}.$$

The optimal information processing for consumer is given as follows

$$EU^* \equiv \sup_{\delta_1,\delta_0} P(S=1)[P(Y=1|S=1)U_1 + P(Y=0|S=1)U_0] - \lambda[H(S) - H(S|X)]$$

$$\text{s.t. } \delta_1 + \delta_0 > 1$$

$$\Longrightarrow EU^* \equiv \sup_{\delta_1,\delta_0} \tfrac{1}{2}[\alpha\delta_1 + (1-\alpha)(1-\delta_0)]U_1 + \tfrac{1}{2}[(1-\alpha)\delta_1 + \alpha(1-\delta_0)]U_0 - \lambda[H(S) - H(S|X)]$$

$$\text{s.t. } \delta_1 + \delta_0 > 1$$

I first ignore the constraint $\delta_1 + \delta_0 > 1$. The first-order conditions with respect to δ_1 and δ_0 are given as follows

$$\delta_1 = \frac{e^{\frac{\alpha U_1 + (1-\alpha)U_0}{\lambda}}\frac{\Delta}{1-\Delta}}{1+e^{\frac{\alpha U_1 + (1-\alpha)U_0}{\lambda}}\frac{\Delta}{1-\Delta}} \text{ and } \delta_0 = \frac{1}{1+e^{\frac{(1-\alpha)U_1 + \alpha U_0}{\lambda}}\frac{\Delta}{1-\Delta}},$$

Where $\Delta = \frac{1}{2}\delta_1 + \frac{1}{2}(1-\delta_0)$.

Denote $h \equiv \frac{\Delta}{1-\Delta}$. Note that $0 < \delta_1^* < 1$ and $0 < \delta_0^* < 1$ if and only if $0 < h < \infty$. Now suppose $0 < h < \infty$. From the first order conditions above, I obtain

$$\begin{aligned} \frac{h}{1+h} &= \frac{1}{2}\left[\frac{e^{\frac{\alpha U_1 + (1-\alpha)U_0}{\lambda}}h}{1+e^{\frac{\alpha U_1 + (1-\alpha)U_0}{\lambda}}h} + \frac{e^{\frac{(1-\alpha)U_1 + \alpha U_0}{\lambda}}h}{1+e^{\frac{(1-\alpha)U_1 + \alpha U_0}{\lambda}}h}\right] \\ \to h &= 1 + \frac{2(1-e^k)}{2e^k - (e^l + e^{l'})}, \end{aligned} \tag{21}$$

Where $k \equiv \frac{U_1+U_0}{\lambda}$, $l \equiv \frac{\alpha U_1+(1-\alpha)U_0}{\lambda}$ and $l' \equiv \frac{(1-\alpha)U_1+\alpha U_0}{\lambda}$.

Note that because of our assumptions (i.e., $1 \geq \alpha > \frac{1}{2}$, $U_1 > 0$ and $U_0 < 0$) , e^l is an increasing function of α and $e^l \geq e^{\frac{k}{2}}$.

(1) When $U_1 + U_0 \geq 0$, it is easy to see that if $\infty > h > 0$, then h is a decreasing function of e^l and thus a decreasing function of α. Furthermore, according to (21), the necessary and sufficient condition for $\infty > h > 0$ is that $e^l > e^k + \sqrt{e^{2k} - e^k}$. Therefore, for $1 \geq \alpha > \alpha^* = \frac{\lambda log(e^k+\sqrt{e^{2k}-e^k})-U_0}{U_1-U_0}$, $\infty > h > 0$ and I have interior solutions as follows

$$\delta_1^* = \frac{e^l h}{1+e^l h} \text{ and } \delta_0^* = \frac{1}{1+e^{l'} h},$$

where h is given in (21).

Note that for $\frac{1}{2} < \alpha \leq \alpha^*$, the prior utility $EU_0 = \frac{1}{2}U_1 + \frac{1}{2}U_0$ is larger than $\max\{EU^*, 0\}$. Therefore, the consumer processes no information and purchases the product.

(2) When $U_1 + U_0 < 0$, it is easy to see that if $\infty > h > 0$, then h is an increasing function of e^l and thus an increasing function of α. Furthermore, according to (21), the necessary and sufficient condition for $\infty > h > 0$ is that $e^l > 1 + \sqrt{1 - e^k}$. Therefore, for $1 \geq \alpha > \alpha' = \frac{\lambda log(1+\sqrt{1-e^k})-U_0}{U_1-U_0}$, $\infty > h > 0$ and I have the same interior solutions as before

$$\delta_1^* = \frac{e^l h}{1+e^l h} \text{ and } \delta_0^* = \frac{1}{1+e^{l'} h},$$

where h is given in (21).

Note that for $\frac{1}{2} < \alpha \leq \alpha'$, the utility of outside option, 0, is larger than $\max\{EU_0, EU^*\}$. Therefore, the consumer processes no information and chooses the outside option. *Q.E.D.*

Proof of Proposition 5

According to proof of Proposition 2,

(1) When $U_1 + U_0 > 0$, for $1 \geq \alpha > \alpha^*$, $\infty > h > 0$ and h is a decreasing function of e^l and thus a decreasing function of α. Therefore, the probability of purchasing, $P(S = 1) = \Delta = \frac{h}{1+h}$, is decreasing with α. Since l' is also decreasing with α, $\delta_0^* = \frac{1}{1+e^{l'} h}$ is

increasing with α. To see the effect of α on δ_1^*, first denote $n = e^l$ and $m = e^k$, and thus $\delta_1^* = \frac{n[1+\frac{2(1-m)}{[2m-(n+\frac{m}{n})]}]}{1+n[1+\frac{2(1-m)}{[2m-(n+\frac{m}{n})]}]}$. Further denote $K(n) \equiv n[1 + \frac{2(1-m)}{[2m-(n+\frac{m}{n})]}]$ and it is not hard to see that

$$\frac{dK(n)}{dn} = \frac{[2-(n+\frac{m}{n})][2m-(n+\frac{m}{n})] + 2(1-m)(n-\frac{m}{n})}{(2m-(n+\frac{m}{n}))^2}$$

Note that n is a function of α and denote function $G(n(\alpha)) \equiv [2-(n+\frac{m}{n})][2m-(n+\frac{m}{n})] + 2(1-m)(n-\frac{m}{n})$. One can show that if $G(n(\alpha = 1)) > 0$, then δ_1^* first decreases with α for $\alpha^* < \alpha < \alpha^{**}$ and then increases with α for $\alpha^{**} < \alpha \leq 1$, where α^{**}satisfies $G(n(\alpha = \alpha^{**})) = 0$; if $G(n(\alpha = 1)) \leq 0$, then δ_1^* always decreases with α for $\alpha^* < \alpha \leq 1$.

(2) When $U_1 + U_0 < 0$, for $1 \geq \alpha > \alpha'$, $\infty > h > 0$ and h is an increasing function of e^l and thus an increasing function of α. Therefore, the probability of purchasing, $P(S = 1) = \Delta = \frac{h}{1+h}$, is increasing with α. It is also easy to see $\delta_1^* = \frac{e^l h}{1+e^l h}$ is increasing with α. For the effect of α on δ_0^*, note that $\delta_0^* = \frac{1}{1+\frac{m}{n}[1+\frac{2(1-m)}{[2m-(n+\frac{m}{n})]}]}$, and denote $R(n) = \frac{[1+\frac{2(1-m)}{[2m-(n+\frac{m}{n})]}]}{n}$ and it is easy to see that

$$\frac{dR(n)}{dn} = \frac{1}{n^2}\frac{[2-(n+\frac{m}{n})][(n+\frac{m}{n})-2m] + 2(1-m)(n-\frac{m}{n})}{(2m-(n+\frac{m}{n}))^2}$$

Further denote $G^*(n(\alpha)) \equiv [2-(n+\frac{m}{n})][(n+\frac{m}{n})-2m] + 2(1-m)(n-\frac{m}{n})$. One can show that if $G^*(n(\alpha = 1)) < 0$, then δ_0^* decreases with α for $\alpha' < \alpha < \alpha''$and increases with α for $\alpha'' < \alpha \leq 1$, where α'' satisfies $G^*(n(\alpha = \alpha'')) = 0$; if $G^*(n(\alpha = 1)) \geq 0$, then δ_0^* decreases with α for $\alpha^* < \alpha \leq 1$.

(3) When $U_1 + U_0 = 0$. $\delta_1^* = \delta_0^* = \frac{e^l}{1+e^l}$ and $P(S = 1) = \frac{1}{2}$. So both δ_1^* and δ_0^* increase with α, while the purchase probability does not vary with α. *Q.E.D.*

Proof of Proposition 6

The consumer solves the following optimization problem:

$$EU^* \equiv \sup_{\delta_1, \delta_0} P(S=1)[P(X=1|S=1)(1+w-p)+P(X=0|S=1)(1-p)] - \lambda I(X,S)$$

$$\text{s.t. } \delta_1 + \delta_0 > 1$$

where $I(X,S)$ is given in (5). We first ignore the constraint $\delta_1 + \delta_0 > 1$. The first-order conditions with respect to δ_1 and δ_0 are given as follows

$$e^{\frac{1+w-p}{\lambda}} = \frac{q+(1-q)\frac{\delta_0}{1-\delta_1}}{q+(1-q)\frac{1-\delta_0}{\delta_1}} \text{ and } e^{\frac{p-1}{\lambda}} = \frac{1-q+q\frac{\delta_1}{1-\delta_0}}{1-q+q\frac{1-\delta_1}{\delta_0}}$$

Solving this, we obtain δ_1^* and δ_0^* as in (10). With the definitions $l = \frac{p-1}{\lambda}$ and $k = \frac{w}{\lambda}$ (as in the main text), one can easily check that when $1-\lambda\log(1-q+qe^{-k}) < p < 1+\lambda\log(1-q+qe^{k})$, $0 < \delta_1^* < 1$, $0 < \delta_0^* < 1$, and $P(S=1) = \frac{q}{1-e^{-l}} + \frac{1-q}{1-e^{k-l}}$. Since l increases with p, it is easy to see that δ_1^* decreases with p, while δ_0^* increases with p. When $p \geq 1+\lambda\log(1-q+qe^{k})$, the utility of outside option, 0, is larger than both EU^* and prior utility $EU_0 \equiv qU_1 + (1-q)U_0$. Therefore, the consumer processes no information and chooses the outside option. When $p \leq 1-\lambda\log(1-q+qe^{-k})$, the prior utility EU_0 is larger than $EU^*$and K. Therefore, the consumer processes no information and purchases the product. *Q.E.D.*

Proof of Proposition 7

According to Proposition 1, $\frac{dP(S=1)}{d\lambda} = (q\frac{-e^{\frac{U_0}{\lambda}}\frac{U_0}{\lambda}}{(e^{\frac{U_0}{\lambda}}-1)^2} - (1-q)\frac{e^{\frac{U_1}{\lambda}}\frac{U_1}{\lambda}}{(e^{\frac{U_1}{\lambda}}-1)^2})\frac{1}{\lambda}$. Note that the function $\frac{e^x x}{(e^x-1)^2}$ is decreasing in x. Therefore, as p increases, $\frac{e^{\frac{U_0}{\lambda}}\frac{U_0}{\lambda}}{(e^{\frac{U_0}{\lambda}}-1)^2}$ and $\frac{e^{\frac{U_1}{\lambda}}\frac{U_1}{\lambda}}{(e^{\frac{U_1}{\lambda}}-1)^2}$ increase, and thus $\frac{d^2P(S=1)}{d\lambda dp} < 0$.

Proof of Proposition 11

According to Proposition 1, if $q \leq \underline{q} \equiv \frac{1-e^{\frac{p-w}{\lambda}}}{1-e^{\frac{1-w}{\lambda}}}$, then the consumer would not process information and would always choose her outside option, even when full information is available. Therefore, the seller has no incentive to provide information. Next, given $p > q + (1-q)w$, only when $\underline{q} < q < \bar{q} \equiv \frac{p-w}{1-w}$ (note that $p > q + (1-q)w$ leads to $q < \frac{p-w}{1-w}$), the consumer would process information, because$EU^* > 0$ according to Proposition 1. In this case, as discussed earlier, the seller would choose the information design such that the optimal signal structure δ_1^{**} and δ_0^{**} chosen by the consumer satisfy $\delta_0^{**} = \Omega_0^* < \delta_0^*$ and $\delta_1^* < \delta_1^{**} < 1$, and the consumer is indifferent between processing information and choosing her outside option without processing any information, given by equation (14).

Differentiating equation (14) with respect to q, I have the following relationship:

$$\frac{dEU^*_{design}}{d\lambda} = \frac{\partial EU^*_{design}}{\partial \delta_1}\frac{d\delta_1^{**}}{dq} + \frac{\partial EU^*_{design}}{\partial \delta_0}\frac{d\Omega_0^*}{dq} + \frac{\partial EU^*_{design}}{\partial q} = 0$$

Note that $\frac{\partial EU^*_{design}}{\partial \delta_1} = 0$ at $\delta_1 = \delta_1^{**}$and $\frac{\partial EU^*_{design}}{\partial \delta_0} > 0$ at $\delta_0 = \Omega_0^*$. So if $\frac{\partial EU^*_{design}}{\partial q} > 0$, then $\frac{d\Omega_0^*}{dq} < 0$; otherwise, if $\frac{\partial EU^*_{design}}{\partial q} < (=)0$, then $\frac{d\Omega_0^*}{dq} > (=)0$. Now, I show $\frac{\partial EU^*_{design}}{\partial q} > 0$. Note that

$$\begin{aligned}\frac{\partial EU^*_{design}}{\partial q} &= \delta_1^{**}(1-p) - (1-\Omega_0^*)(w-p) \\ &\quad - \lambda\{-(\delta_1^{**} + \Omega_0^* - 1)\log(q\delta_1^{**} + (1-q)(1-\Omega_0^*)) \\ &\quad + (\delta_1^{**} + \Omega_0^* - 1)\log(q(1-\delta_1^{**}) + (1-q)\Omega_0^*)\}\end{aligned}$$

Thus, $\frac{\partial^2 EU^*_{design}}{\partial q^2} = \lambda(\delta_1^{**} + \Omega_0^* - 1)^2(\frac{1}{q(1-\delta_1^{**})+(1-q)\Omega_0^*} + \frac{1}{q\delta_1^{**}+(1-q)(1-\Omega_0^*)}) > 0$. This implies if $\frac{\partial EU^*_{design}}{\partial q} \leq 0$ at q satisfying $EU^*_{design}(\delta_1^{**}(q), \Omega_0^*(q), q) = 0$, then for any $\hat{q}$ satisfying $\underline{q} < \hat{q} < q$, I have $EU^*_{design}(\delta_1^{**}(\hat{q}), \Omega_0^*(\hat{q}), \hat{q}) > 0$. Next, for any q satisfying $\underline{q} < q < \bar{q} \equiv$

$\frac{p-w}{1-w}$, denote the maximal utility of learning under full information as $EU^*(q)$. It must be that $EU^*(q) > EU^*_{design}(\delta_1^{**}(q), \Omega_0^*(q), q)$, because information design decreases the utility of learning. Furthermore, from Proposition 1, I know that if $q \to \underline{q}$, then the maximal utility of learning under full information $EU^*(q) \to 0$, which implies that $EU^*_{design}(\delta_1^{**}(q), \Omega_0^*(q), q) \to 0$ as $q \to \underline{q}$. However, this contradicts with $EU^*_{design}(\delta_1^{**}(\hat{q}), \Omega_0^*(\hat{q}), \hat{q}) > 0$ for any $\hat{q}$ satisfying $\underline{q} < \hat{q} < q$. So it must be that $\frac{\partial EU^*_{design}}{\partial q} > 0$ at q satisfying $EU^*_{design}(\delta_1^{**}(q), \Omega_0^*(q), q) = 0$. Therefore, $\frac{d\Omega_0^*}{dq} < 0$. Finally, according to equation (15), the right hand side of equation (15) decreases with q and increases with Ω_0^*. Thus, I must have $\frac{d\delta_1^{**}}{dq} > 0$. In a word, lowerprior belief q motivates the seller to increase Ω_0^* but decrease Ω_1^*.

Similarly, in terms of the impact of λ on information design. Differentiating equation (14) with respect to λ, I have the following relationship:

$$\frac{dEU^*_{design}}{d\lambda} = \frac{\partial EU^*_{design}}{\partial \delta_1}\frac{d\delta_1^{**}}{d\lambda} + \frac{\partial EU^*_{design}}{\partial \delta_0}\frac{d\Omega_0^*}{d\lambda} - I(X,S)_{\delta_1=\delta_1^{**},\delta_0=\Omega^*_{0,}} = 0 \tag{22}$$

Note that $\frac{\partial EU^*_{design}}{\partial \delta_1} = 0$ at $\delta_1 = \delta_1^{**}$and $\frac{\partial EU^*_{design}}{\partial \delta_0} > 0$ at $\delta_0 = \Omega_0^*$. Therefore, (22) implies that $\frac{d\Omega_0^*}{d\lambda} > 0$. Next, according to equation (15), as λ increases, the left hand side of equation (15) decreases, and since $\frac{d\Omega_0^*}{d\lambda} > 0$, it must be that $\frac{d\delta_1^{**}}{d\lambda} < 0$ (and thus $\frac{d\Omega_1^*}{d\lambda} < 0$ under the assumption that a seller always sets $\Omega_1^* = \delta_1^{**}$). Therefore, higher information processing cost λ motivates the seller to increase Ω_0^* but decrease Ω_1^*. Q.E.D.