

The Rationally Inattentive Manager

Rongyu Wang*

Independent Researcher (PhD Economics, University of Edinburgh, 2016), Qingdao, 266072, China

Received October 04, 2019; Revised December 12, 2019; Accepted December 17, 2019

Copyright ©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract We analyze the ordering decisions of a manager who acquires demand information to make inventory decisions. Human beings' cognitive ability in information acquisition is imperfect by nature. We adopt the rational inattention theory to model the information acquisition activity. By reformulating the rational inattention decision problem, we find that rational inattention theory resembles the free energy principle (Friston 2009, 2010) that is used to model brain activities. Therefore, rational inattention theory gets a neurocognitive foundation. We obtain closed-form solutions of the optimal action strategies, which is a discrete distribution although the states are continuous. The process of resulting the optimal action strategies show that the essence of the process, as a neurocognitive activity, is to *think until nothing to think*. The optimal action strategies themselves shows that an effective thinking is hard to obtain unless *think one more step*. We also study the impact on decisions once the primitive information is contaminated. If the primitive information is contaminated by, for example, data sampling errors, the acquired information will be eclipsed by the contamination and exhibit certain features that mistaken the ordering decisions by acquiring information if the primitive information were clean. The specific forms of the well-known one-to-one correspondence between action and signal are also restricted by the contamination. We propose an empirical approach to identify the optimal ordering quantities, and the approach works only if the information contamination is separated from the estimation of the unconditional optimal ordering strategies.

Keywords Rational Inattention Theory, Behavioral Decision Biases

1 Introduction

In this paper, we study a rational inattention decision problem in an inventory decision-making context. The decision maker (manager henceforth) acquires demand information by intensive marketing research and based on the acquired

information he makes the ordering decision. The information acquisition process is costly due to the neuronal limitation of human beings' information acquisition ability. In human brain, the dopaminergic system conduct the information acquisition process (Friston 2009, 2010). As a neuronal activity, the information acquisition process costs energy. Hence, human beings by nature are constrained in their information acquisition ability. Therefore, to make an ordering quantity in a complete information setting is rather ideal. Even if the primitive information is open and clean, human beings are not able to accurately acquire the information. The demand information usually refers to the information of demand, and in this paper we adopt this understanding. The form of our model resembles a newsvendor-type decision making model. However, there are two differences. First, in our model, the managers do not try to forecast any information on demand, but they are trying to acquire information about the existing market demand through an intensive marketing research. Second, technically, in our model, managers update their prior information on demand after acquiring demand information by intensive marketing researches, but in conventional newsvendor models, the managers just rely on prior knowledge to make decisions. We use rational inattention theory to model the managers' information acquisition activity. In order to differentiate our decision-making models from the conventional newsvendor models, we call our decision makers "managers" and not use the jargon "newsboys".

The most recent development of rational inattention theory finds that there exists an optimal consideration set after information acquisition where not all actions are chosen with positive likelihood (Caplin, Dean and Leahy 2019), and the decisions are discrete in the posterior even if the prior distribution follows a Borel measure (Jung, King, Matějka and Sims 2019). Besides following the standard understanding of the rational inattention theory, we reformulate the rational inattention decision problem and find that the reformulated form of the rational inattention decision problem resembles the free energy principle that describes neuronal activity in brain by Friston (2009, 2010), and therefore the rational

inattention theory can get a neurocognitive foundation to understand. This paper is an attempt to understand the rational inattention theory from the neurocognitive perspective that is feasible in the free energy principle. After our paper, we believe more efforts will be needed for further understanding of rational inattention theory from a neurocognitive perspective.

As a contribution to rational inattention theory, we obtain the explicit analytical form of the optimal action strategies (equation 13, Proposition 2). The neuronal process of information acquisition initiates from obtaining the likelihood that any optimal ordering quantity is chosen across all possible demand states, which are expressed by the unconditional probability of choosing certain ordering quantity. The information acquisition process is essentially exhausting all layers of impacts of demand states that could make on the ordering decisions. That means, think (about the impact of demand state on optimal ordering decisions) until nothing (of the impact of demand state on optimal ordering decisions) to think. This stage determines what optimal ordering quantities are possible to choose in the posterior. The acquired information about a specific state weighs the probability that those ordering quantities in the posterior are chosen given the state.

To identify and solve the optimal ordering quantities, additional conditions are needed. We propose the empirical approach to identify the optimal ordering quantities. By applying empirical approach, the unconditional likelihood that certain optimal ordering quantities are chosen are estimated in the first place, and the estimates together with the theoretical formula of the unconditional likelihood are used to identify the optimal ordering quantities. However, the empirical approach may not work well due to the disturbance of technical errors, e.g. the data sampling error, in the estimation. Once the technical errors are involved, the estimation result will deviate and complicate the genuine information acquisition effort without the disturbance of the technical errors. Therefore, technically, separating the technical errors' impact from the estimation of the unconditional likelihood is critical for the validity of the empirical approach. In this paper, we give an example to show how the empirical approach works to identify the posterior distribution once the technical error distribution follows a special form so that analytically the technical error can be separated from the genuine information acquisition effort that is not affected by the technical errors. If and only if the optimal ordering quantities can be identified from our condition, then a binary posterior distribution exists. The pair of optimal ordering quantities, once proven existence, can be multiple.

However, the immediate decisions by information acquisition are hard to be effective. The effectiveness failure means given the information acquired about a certain state, the optimal ordering quantity that is most close to the state is not chosen with highest likelihood with generality. It happens not just for binary posterior distributions, but also for general forms of discrete posterior distributions. Such a result does

not mean information acquisition is useless in decision making. Based on the findings of our model, we suggest that decision makers can refine their information acquisition result into a binary form of posterior distribution as long as they can, and think one more step based on the result that the optimal ordering quantity in the binary posterior that is chosen with lower likelihood is just the one that is closer to the state.

The technical errors are prevalent in our daily decision making by acquiring information. The technical errors include data sampling errors, errors of machines that are used for gathering information or errors caused by ignorance in knowledge, etc. The technical errors contaminate the information collected for information acquisition. Because we have managed to model the brain activity by simply acquiring information from a pure source, it is interesting to further consider once the primitive information is contaminated by the technical errors, how the information acquisition activity and subsequent ordering decisions respond to the contamination. We find that the genuine information acquisition activity is hard to be immune from the contamination of the technical errors. With generality, we specify the impact of technical errors towards the state is additive. A benefit of such specification is that the technical errors' impact is kept at its best in the information acquisition process. That is, the errors in the contaminated primitive information can be completely passed to the acquired information, which is reasonably a worst case of the attack of the technical errors on the information acquisition activity.

We experiment three technical error distributions (exponential distribution, truncated normal distribution and Weibull distribution) to study the impact of technical errors on the information acquisition activity. We find that the technical errors' impact can be so dominant that overwhelmingly eclipse a genuine information acquisition effort. A typical result in rational inattention theory is that at equilibrium there exists a one-to-one correspondence between signal and action, which represents how the decision maker interprets the observed information. The form of the one-to-one correspondence is not restricted if the primitive information is clean. However, the technical errors, once involved in the information acquisition process, restricts the available form of the one-to-one correspondence. We consider the linear form interpretation and we find that the linear form interpretation cannot exist subject to some technical error distributions and the specific form of the linear interpretation, that is its scale (absolute value) and its sign, is determined by the given form of the technical error distribution. The technical error distribution can mistaken the optimal ordering decisions that were obtained if the primitive information were clean.

This paper is organized as follows. In section 2, we review related literature. In section 3, we describe the model and exhibit the alternative formulation of rational inattention decision problems that can be understood via the free energy principle. In section 4, we exhibit the optimal ordering

strategies, the identification of optimal ordering quantities and implications reflected by the process of obtaining optimal ordering strategies (*think until nothing to think*) and by the optimal ordering strategies themselves (*think one more step*). In section 5, we study how technical errors affect managers' optimal ordering decisions by information acquisition. Section 6 concludes the paper.

2 Literature Review

This paper studies decision biases of a rationally inattentive decision maker given clean primitive information and contaminated primitive information. We first focus on human beings' information acquisition effort on inventory ordering decisions from clean primitive information. It completely reflects the behavioural biases caused by human beings' limited information acquisition ability and is used as benchmark to compare the decision biases by information acquisition from contaminated primitive information. The limited cognitive ability of acquiring information is prevalent in our daily life. People can find that when trying to acquire information from the same information source, the acquired information are usually not even across individuals. Another example is that when reading an academic paper, the extents of understandings of the paper vary across readers by reading it with same rounds. Information acquisition has many concrete forms in human cognitive activities. For example, thinking is a type of cognitive activities. In cognitive psychology, researchers study cognitive activities as information acquisition processes and employ information theory as instruments methodologically to serve the researches (Lachman, Lachman and Butterfield 1979).

We also study cognitive activities as information acquisition processes. Unlike the researches on cognitive activities in cognitive psychology, we adopt rational inattention theory to model the information acquisition process. The rational inattention theory is proposed by Sims (1998, 2003, 2006). The theory endogenizes decision makers' information acquisition activities and it features the flexibility of information acquisition compared with other information-acquisition theories in economics.¹ The construction of rational inattention theory is primarily based on information theory by Claude Shannon (1946). Rational inattention theory has been widely applied in economics and management sciences. There have been many literatures that provide a good overview of the rational inattention theory, e.g. Tutino (2011), Matějka and McKay (2015) and Boyacı and Akçay (2017). For concision, we will not conduct another overview of the theory in this paper but review its most recent development later.

Cognitive psychology has developed to a stage that studies the discipline from the neuronal perspective. With refined understandings of how human brain works to conduct cogni-

tive activities, neuroscientists have proposed new theories to describe cognitive activities of human beings. An emerging one is Friston's (2009, 2010) free energy principle. The free energy principle mathematically describes all human organism's activities from a common ground. That is, to make the states of human organisms as ordered as possible. Such activities include cognition, which is conducted by the dopaminergic system in human brain. Mathematically, human organisms act by reducing the free energy to achieve the objective that make the states of human organism as ordered as possible. The free energy function, which is modelled by Kullback-Leibler divergence between prior and posterior, can be equivalently transformed to the mutual information between the prior and posterior, and vice versa. Therefore, to reduce a free energy and to reduce a mutual information are equivalent. Therefore, given the same motivation and interpretation of the free energy principle, the mathematical formulation by minimizing the typical free energy function can be equivalently transformed to by minimizing a mutual information. Based on the mathematical formulation of the rational inattention decision models, and by the enlightenment of Friston's (2009, 2010) free energy principle, we find that a rational inattention decision problem can be mathematically reformulated to fit Friston's (2009, 2010) free energy principle. Therefore, rational inattention theory has a neurocognitive foundation.

The most recent development of rational inattention theory is by Jung, King, Matějka and Sims (2019) and Caplin, Dean and Leahy (2019). Jung, King, Matějka and Sims (2019) studies a rational inattention decision problem where the states follows Borel measure and actions are continuous. They find that only a countable set of points with no limit points can be considered as ultimate optimal action decisions. Matějka (2016) obtains a similar result in an application of rational inattention theory to study pricing problems by a rationally inattentive seller. In his model the states are continuously distributed and the actions are continuous, the seller is rationally inattentive to sales price. We also study a continuous-state continuous-action model, and we obtain the closed-form solutions of the action strategies. The properties of obtained closed-form action strategies are consistent with the findings by Jung, King, Matějka and Sims (2019) and Matějka (2016). For example, the actions in the posterior are discretely distributed. Beyond the properties that are consistent with the findings by Jung, King, Matějka and Sims (2019) and Matějka (2016), our research further shows that if the primitive information is contaminated, e.g. the primitive data for information acquisition is contaminated by sampling errors, even if the decision maker is rationally inattentive, the resulted actions in the posterior are continuously distributed.

Caplin, Dean and Leahy (2019) studies rational inattention discrete choice problems and find that an optimal consideration set emerges after information acquisition. In the optimal consideration set, only a subset of actions remains as candidate optimal action decisions. By applying

¹Yang (2015) provides a detailed comparison between rational inattention theory and other information-acquisition theories in economics.

Caplin, Dean and Leahy's (2019) approach to derive the optimal consideration set, we derive a similar result in our continuous-state continuous-action model. Our result that actions are discretely distributed in the posterior paves the way for our genuine contributions in the following of this paper.

A contribution of this paper is that we obtain the closed-form solutions of action strategies of a rational inattention decision problem. This is the first closed-form solution of action strategies for continuous-state continuous-action rational inattention decision problems in literature. In our context, we call the obtained action strategies by optimal ordering strategies. By considering information acquisition from clean primitive information, the optimal ordering strategies are the benchmarks that completely reflect the impact of costly information acquisition, a cognitive activity, on ordering decisions.

Schweitzer and Cachon (2000) conduct experiments to study the biases between actual decision making and the fractile solutions of conventional newsvendor models. The conventional newsvendor model calculates a decision maker's ordering quantity without any attempt of information acquisition. However, in practice, it is impossible for a newsvendor manager to make his ordering decision without considering any contingent information of demand. Schweitzer and Cachon (2000) think that the actual inventory ordering decisions will be necessarily affected by behavioural biases, and they want to identify how exactly each type of behavioural biases' impact on inventory ordering decision making such that the decision deviates from the fractile solutions of conventional newsvendor models. In fact, by cognitive psychology, any behavioural phenomenon can be attributed to information acquisition activities. Therefore, the experimental results by Schweitzer and Cachon (2000) can be referred to compare the theoretical results of our model by the decision maker acquiring information from clean primitive information, which abstracts any behavioural phenomena just as information acquisition activities.²

According to Schweitzer and Cachon's (2000) experimental results, there are systematic biases between the fractile solutions of the newsvendor model and the experiment subjects' actual decisions. A regular phenomenon emerges that for making the ordering decisions for low-profit products, overordering happens; for making ordering decisions for high-profit products, underordering happens.³ By its exhibition, this systematic bias is also called "pull-to-center" bias. Schweitzer and Cachon think the systematic biases are caused by some anomalies on preferences, such as the tendency of minimizing ex-post inventory error, and

²The demand state is designed by the experimenters in Schweitzer and Cachon (2000) and hence accurately known by the experimenters. Therefore, the primitive information in Schweitzer and Cachon (2000) is clean.

³In Schweitzer and Cachon (2000), overordering is defined that actual ordering quantity is more than the fractile solution of the conventional newsvendor model, and underordering defined that actual ordering quantity is less than the fractile solution of the conventional newsvendor model.

anchoring and insufficient adjustment.

For our model, the optimal ordering strategy by information acquisition from clean primitive information is discrete. It could exhibit biases similar to the "pull-to-center" bias. Suppose the optimal ordering strategy follows a binary distribution. Therefore, if a systematic biases towards the fractile solution happens, the fractile solution should be smaller than the small optimal ordering quantity obtained by information acquisition to exhibit overordering, while bigger than the large optimal ordering quantity obtained by information acquisition to exhibit underordering. However, it also happens that the fractile solution is valued between the small ordering quantity and the large optimal ordering quantity. Therefore, our model cannot conclude that the "pull-to-center" bias surely happens for ordering decision making by information acquisition from clean primitive information. If the decision maker just acquires information from clean primitive information and makes ordering decisions, their actions should exhibit a more complicated pattern than the "pull-to-center" bias, even if the actions are presented in their simplest form, a binary distribution.

In our paper, based on the insights derived from the study of behavioral decision biases, we further study how the contamination to primitive information by technical errors affect the inventory decision making by information acquisition. In reality, any primitive information is hard to expect to be completely clean, and humans have to acquire the information they intend to get by interpreting such contaminated primitive information. Considering how a typical human being with conventional information acquisition ability acquires the underlying information and make decisions by interpreting the contaminated primitive information is not studied before. We believe that the topic is interesting and worthwhile for comprehensive researches in future.

3 The Model

A manager acquires demand information by doing an marketing research before making an inventory decision. The actual demand can only be known when sales are over. In order to benchmark the effect of cognitive errors in the acquired information and ultimate decisions, we assume the source of information where the manager acquires information from is clean, i.e. the primitive information containing no technical errors. The manager acquires information on demand from interpreting the primitive information, i.e. acquiring information from primitive information. The knowledge obtained from analyzing and interpreting the primitive information will be summarized during the information acquisition process and used to update the existing knowledge about demand, which is probabilistic since the uncertainty around demand still remains as a consequence of the managers' limited cognitive ability in

acquiring information. The event sequence of the rational inattention newsvendor problem is described as follows:

1. The manager acquires demand information from the primitive information on demand by intensive marketing researches;
2. Updating demand information by the acquired information from marketing research, the manager makes an ordering decision that maximizes his expected profit based on the posterior information;
3. Sales start and the ordered goods are sold.

Signal, which is denoted by s , refers to the information obtained by interpreting the primitive information. Accordingly, S denotes the signal set, which contains all signals that represent all possible interpretation results. If contingent information about demand cannot affect the decision maker to make his ordering decisions, then their decisions can only rely on prior information on demand. By making use of historical data, the manager is able to get a prior distribution describing demand. A demand state is denoted by $\theta \in \mathbb{R}$. The distribution of θ is given by $G(\theta)$, of which density function is denoted by $g(\theta)$. We specify the prior distribution featuring by the following properties: $g(\theta) \geq 0$ for all $\theta \geq 0$, and $g(\theta) = 0$ for all $\theta < 0$. An inventory decision based only on prior information on demand is the fractile solution of the conventional newsvendor model. In an integrated setting, suppose the production cost of the product is c and sold at price p , the optimal ordering quantity (fractile solution) α^* equals $G^{-1}(\frac{p-c}{p})$. We use α to represent ordering quantities, and α^* to represent the optimal ordering quantities.

The rationally inattentive manager makes his ordering decisions by acquiring demand information. The information acquisition activity in our context is a neuronal activity. Human brain has a natural tendency that resist disorder and can be regarded as an optimization machine. Based on the presumptions, brain theories such as free energy principle (Friston 2009, 2010) has been proposed. Our understanding of information acquisition activity is underpinned by Friston's (2009, 2010) theory, although we model our decision maker's problem according to rational inattention theory. In fact, rational inattention theory, as said to model information acquisition process as a cognitive activity, needs a foundation to support it. The foundation, by the purpose of the theory, should be neuropsychology or neurocognition science, which specifically studies information acquisition as a neuropsychological or neurocognitive phenomenon. The logic of the free energy principle can explain the rational inattention decision problems.

Given the acquired information on demand, the manager chooses the optimal ordering quantity α^* that leads to the highest expected profit. This expectation is completely grounded on noises involved in the acquired demand information due to humans' limited cognitive ability. It is

completely the neuronal noises.

Based on the understanding that information acquisition activity is a neurocognitive process, we use Matějka and McKay's (2015) two-stage approach to construct our manager's decision problem. Matějka and McKay (2015) used the two-stage approach to model the rational inattention discrete choice problem. It separates the decision problem by a stage of determining an information strategy and a stage of determining an action strategy. In the first stage, an information strategy that determines the joint distribution of s and θ and the set S is obtained, which represents how the decision maker updates his belief about demand and what information he will acquire by interpreting primitive information by considering all demand states respectively. In the second stage, he chooses an ordering quantity to maximize his expected profit by obtaining a realized signal (knowledge from interpreting the primitive information) and updating his belief (renewed perception on potential demand).

The posterior belief is $f(\theta|s)$, which is obtained according to the obtained information strategy $f(\theta, s)$ and Bayes theorem $\frac{f(\theta, s)}{f(s)}$, and $f(s)$ is the unconditional density function of signals. Following the two-stage approach, we solve this problem backward. First, we look at the second stage. Given the updated belief, the manager chooses the optimal ordering quantity α^* to maximize his expected profit:

$$\Pi(f(\theta|s)) \equiv \max_{\alpha} \int_{-\infty}^{+\infty} [p \min(\theta, \alpha) - c\alpha] f(\theta|s) d\theta \quad (1)$$

where $p \min(\theta, \alpha) - c\alpha$ is the actual profit that the decision maker will get by choosing an ordering quantity α . The specification of the actual profit function is featured by that under the complete information setting where the decision maker has the perfect ability to acquire information, the optimal ordering quantity α^* that maximizes the profit is θ . Therefore, by knowing the fact, a rule of thumb for managers to choose their optimal ordering quantity is choosing the demand they perceive by acquiring state information. In a costly information acquisition setting, for example in our model, any deviation from the ultimate determination of optimal ordering quantities from θ is the reflection of the imperfectness of information acquisition.

And hence the optimal ordering decision is

$$\alpha^* \equiv \arg \max_{\alpha} \int_{-\infty}^{+\infty} [p \min(\theta, \alpha) - c\alpha] f(\theta|s) d\theta \quad (2)$$

Next, we proceed to the first stage. In this stage, the information strategy is determined endogenously. The information strategy is the resulted way of thinking that describes how human brain works when acquiring information. The information strategy $f(\theta, s)$ and the related posterior belief $f(\theta|s)$ epitomize the information acquisition process. The information strategy also depends on the signal set S , which formation reflects the information acquisition activities of all states. Therefore, by considering all states, the determination of the information strategy can be formulated by

IA-1:

$$\max_{f(s,\theta)} \int_{-\infty}^{+\infty} \int_S \Pi(f(\theta|s)) f(s|\theta) g(\theta) ds d\theta$$

s.t.

$$I(f(\theta|s)) \geq 0$$

$$\int_S f(s, \theta) ds = g(\theta) \quad \theta \in \Theta \subset \mathbb{R}$$

where $\Pi(\cdot)$ is determined by (1) and S as defined is a signal set. Given that S is the signal set obtained at the optimum, i.e. the optimal signal set, the obtained $f(s, \theta)$ is the optimal information strategy.

In rational inattention theory, the information acquisition process is modelled by the reduction of expected entropy. It is a standard setting in rational inattention literature, e.g. Sims (1998, 2003, 2006) and Matějka and McKay (2015). In our context, the entropy based information acquisition process is given by

$$I(f(\theta|s)) \equiv H(g(\theta)) - E_s[H(f(\theta|s))] \quad (3)$$

where $H(\cdot)$ represents the entropy function. The entropy function measures the informational content of a distribution, or the average surprise of a state. In information theory, function $I(\cdot)$ is also called mutual information. The above mutual function measures how much information about θ is refined expectedly after signal s is obtained. Even if mutual information $I(f(\theta|s))$ is nonnegative with generality, the amount of information to process is still required to be nonnegative such that the information acquisition activity is restricted to be a process that refines the knowledge of the demand state. For a complete overview of concepts including Shannon entropy and mutual information in information theory, refer to Cover and Thomas (1991). Therefore, the informational content of demand state θ and the informational content of θ after an update by signal s are

$$H(g(\theta)) = - \int g(\theta) \log g(\theta) d\theta$$

and

$$H(f(\theta|s)) = - \int f(\theta|s) \log f(\theta|s) d\theta$$

respectively.

In rational inattention theory, a rationally inattentive agent at equilibrium acquires information that is most relevant to his objective. In plain language, the manager always tries his best to acquire the information he thinks most useful. In our context, the manager acquires demand information that is most relevant to his profit maximization purpose. The formulation of **IA-1** makes the feature salient. Decision maker's objective is to find an information strategy that maximizes his expected profit, and he needs additional information (s) to assist him doing so. The constraint $I(f(\theta|s)) \geq 0$ represents the information acquisition process which determines what information the decision maker can obtain to serve the profit optimization purpose. Therefore, under **IA-1**, information strategy reflects profit-maximization related information at optimum.

In practice, analysts naturally has the tendency to acquire information that is most relevant to companies' operational objectives like stocking, promotion, etc. Compared with acquiring information from a bunch of materials comprehensively, to acquire information specifically is more efficient. After getting the signal by interpreting the primitive information, the manager updates and renews his knowledge on demand. The information acquisition process completes as the update finishes.

IA-1 focuses on profit maximization, even if it is constrained by an information acquisition activity. However, if the decision maker focuses on information acquisition, then the rational-inattention decision problem can be formulated by treating information acquisition as the decision maker's objective. **IA-2** is the formulation of the decision making problem by focusing on information acquisition:

IA-2:

$$\min_{f(s,\theta)} I(f(\theta|s))$$

s.t.

$$- \int_{-\infty}^{+\infty} \int_S \Pi(f(\theta|s)) f(s|\theta) g(\theta) ds d\theta \leq 0$$

$$\int_S f(s, \theta) ds = g(\theta) \quad \theta \in \Theta \subset \mathbb{R}$$

The constraint $-\int_{-\infty}^{+\infty} \int_S \Pi(f(\theta|s)) f(s|\theta) g(\theta) ds d\theta \leq 0$ can be understood as the disutility or unhappiness of the information acquisition activity beyond the information acquisition cost. Regardless of the constraint, the decision maker tries to minimize the mutual information between the prior knowledge of state and the posterior knowledge of state. That is, the optimal information strategy should make the acquired information as much consistent as with

the prior knowledge on demand. Therefore, the acquired information makes the knowledge of state as ordered as possible. It resembles the free energy principle (Friston 2009, 2010) that minimizes a free energy functional, i.e. the Kullback-Leibler divergence between the distributions $g(\theta)$ and $f(\theta|s)$. According to the free energy principle, in order to make the brain state in order, the brain minimizes a free energy functional to achieve a steady state, which can be non-equilibrium. Mutual information and free energy functional have definite relationship with each other. Therefore, mathematically, one function can be transformed to another, and the free energy principle also interprets **IA-2**. That is, the result of **IA-2** is non-equilibrium, which differs from the interpretation of the result of **IA-1**, in which the information acquisition result is at equilibrium. In **IA-2**, the aim of information acquisition is just for making the knowledge of state in order as much as possible, not finding the best order of the knowledge of state. If such a status is achieved, then the decision maker halts his information acquisition activity. However, in **IA-1**, decision maker aims to maximize his utility, a status he would not deviate if he is in the status, and information acquisition assists him to achieve this objective.

However, the process of minimizing the difference between the entropy of the two distributions is an unhappy process. The extra disutility the information acquisition process brings, i.e. the disutility from the constraint $-\int_{-\infty}^{+\infty} \int_S \Pi(f(\theta|s))f(s|\theta)g(\theta)dsd\theta \leq 0$, critically weighs how consistent the acquired information is with the prior knowledge on demand. Therefore, the acquired information on demand inevitably is biased due to the extra disutility of information acquisition. According to the free energy principle, minimizing the Kullback-Leibler distance between prior and posterior models brain's natural tendency to resist the disorder of the organism after an information transmission or equivalently information acquisition in our context. To minimize the mutual information between the prior and posterior in our context, by its mathematical similarity with the specification of free energy principle, can also be understood as a modeling of reversing the disordering tendency by human brain. Therefore, **IA-2** emphasizes the information acquisition activity as a neuronal activity by additionally considering extra unhappiness the process incurs.

A well known result in rational inattention theory is that there exists a one-to-one correspondence between action and signal at equilibrium. Matějka and McKay (2015) prove the result in rational-inattention discrete choice problems. The one-to-one correspondence relationship can be used to reformulate a rational inattention decision problem from obtaining an information strategy to obtaining an action strategy.

Such a one-to-one correspondence between signal and action at optimum also exists for decision problem **IA-1**, which is a continuous-state continuous-action rational inattention decision problem. We provide a proof in an additional

appendix (Appendix for Additional Proofs) to show the existence of the one-to-one correspondence in our rational inattention decision problem. The additional appendix is available upon request. The one-to-one correspondence allows us to transform our rational inattention decision problem from obtaining the optimal information strategy to obtaining the optimal action strategy (optimal ordering strategy). In the additional appendix, we also provide a proof to exhibit how the one-to-one correspondence relationship is used to reformulate the expression of the decision problem from based on information strategy to based on action strategy.

However, the one-to-one correspondence relationship is not necessarily held for **IA-2**. As we have stated, in **IA-2**, the brain minimizes a free energy in order to achieve a steady state. The steady state is not necessarily at equilibrium. Back to our decision problem, for utility maximization purpose (**IA-1**), an additional signal to describe a given action is inefficient; however, for minimizing free energy (**IA-2**), an additional signal describing a given action can be necessary in order to achieve a steady state of the information acquisition organ. In this paper, we will not delve into what relationship between action and signal exhibit at steady state in **IA-2** and instead we tentatively deduce the action strategies of **IA-2** once a one-to-one correspondence between action and signal is held at steady state. The exact relationship between action and signal at steady state in **IA-2** is expected to study in our future work.

The one-to-one correspondence between state and signal is applied to **IA-1** as well as **IA-2** to reformulate the decision problem. Therefore, by reformulation according to the one-to-one correspondence relationship, the rationally inattentive manager's ordering decision problem is transformed to:

DP-1 (obtained from reformulating IA-1):

$$\max_{f(\alpha^*|\theta)} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi(f(\theta|\alpha^*))f(\alpha^*|\theta)g(\theta)d\theta d\alpha^* \quad (4)$$

s.t.

$$I(f(\alpha^*|\theta)) \geq 0 \quad (5)$$

$$f(\alpha^*|\theta) \geq 0$$

$$\int_{-\infty}^{+\infty} f(\alpha^*|\theta)d\alpha^* = 1$$

or

DP-2 (obtained from reformulating IA-2 once signal and action exhibits a one-to-one correspondence at steady state):

$$\min_{f(\alpha^*|\theta)} I(f(\alpha^*|\theta))$$

s.t.

$$-\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi(f(\theta|\alpha^*)) f(\alpha^*|\theta) g(\theta) d\theta d\alpha^* \leq 0 \quad (6)$$

$$f(\alpha^*|\theta) \geq 0 \quad (7)$$

$$\int_{-\infty}^{+\infty} f(\alpha^*|\theta) d\alpha^* = 1 \quad (8)$$

where $I(f(\alpha^*|\theta)) = -\int_{-\infty}^{+\infty} f(\alpha^*) \log f(\alpha^*) d\alpha^* + \int_{-\infty}^{+\infty} (\int_{-\infty}^{+\infty} f(\alpha^*|\theta) \log f(\alpha^*|\theta) d\alpha^*) g(\theta) d\theta$ and let $f(\alpha^*) = \int_{-\infty}^{+\infty} f(\alpha^*|\theta) g(\theta) d\theta$ be the unconditional density function of α^* .

Once again, we emphasize that given the one-to-one correspondence relationship between state and signal, **DP-1** and **DP-2** are equivalent; the one-to-one correspondence between state and signal is not necessarily held in **IA-2**.

In the rest of the paper, we name $f(\alpha^*)$ by the unconditional optimal ordering strategy. In the following, we will first discuss the manager's ordering decisions by acquiring information and exhibit our findings on this subject. After that, we will study how technical errors, an external source of noises, affect the manager's ordering decisions and present our findings on this subject.

4 The Optimal Ordering Strategy

By solving **DP-1**, we obtain an implicit form of the optimal ordering strategy, which describes the relationship between the optimal ordering strategy and unconditional optimal ordering strategy. The implicit form is given by

$$f(\alpha^*|\theta) = \frac{f(\alpha^*) \exp(\frac{p \min(\theta, \alpha^*) - c\alpha^*}{\lambda})}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{p \min(\theta, \alpha^*) - c\alpha^*}{\lambda}) d\alpha^*} \quad (9)$$

where

$$\int_{-\infty}^{+\infty} \frac{\exp(\frac{p \min(\theta, \alpha^*) - c\alpha^*}{\lambda})}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{p \min(\theta, \alpha^*) - c\alpha^*}{\lambda}) d\alpha^*} g(\theta) d\theta = 1 \quad (10)$$

If $\lambda = 0$, then $\alpha^* = \theta$. λ is the shadow price of the constraint (5). It represents the price the decision maker is willing to pay for acquiring one more unit of information at optimum. The situation where λ equals 0 indicates that the decision maker has a perfect ability in information acquisition.

Solving **DP-2**, the optimal ordering strategy is the conditional probability density function in the following form

$$f(\alpha^*|\theta) = \frac{f(\alpha^*) \exp(\lambda(p \min(\theta, \alpha^*) - c\alpha^*))}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp(\lambda(p \min(\theta, \alpha^*) - c\alpha^*)) d\alpha^*} \quad (11)$$

where

$$\int_{-\infty}^{+\infty} \frac{\exp(\lambda(p \min(\theta, \alpha^*) - c\alpha^*))}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp(\lambda(p \min(\theta, \alpha^*) - c\alpha^*)) d\alpha^*} g(\theta) d\theta = 1 \quad (12)$$

The derivation of equation (9) or (11) is shown in Appendix A. Equation (10) or (12) comes from the probability property that $f(\alpha^*) = \int_{-\infty}^{+\infty} f(\alpha^*|\theta) g(\theta) d\theta$. If $\lambda = +\infty$, then $\alpha^* = \theta$. In equations (11) and (12), λ is the shadow price of the constraint (6). It represents the price the decision maker is willing to pay reduce one more unit of extra unhappiness incurred during information acquisition at steady state. The situation where λ equals $+\infty$ indicates that the decision maker can suffer infinite unhappiness for acquiring information. If $\lambda = 0$, the optimal ordering decision α^* is independent from demand state θ . It indicates that the decision maker is unwilling to pay anything for reducing the extra disutility for the information acquisition activity. Because the extra disutility is a necessary part of information acquisition, the refusal to pay means the decision maker refuse the entire information acquisition activity. Therefore, his decision is irrelevant with the state.

Because **DP-2** is obtained by assuming that at steady state a one-to-one correspondence between action and signal is held, and the actual correspondence relationship at steady state may be beyond a one-to-one mapping. **DP-2** is an experiment to tentatively discuss what properties **IA-2** may have if a steady state is achieved. Therefore, in this paper, we will not continue discussing **DP-2**. In the following, we will focus on **DP-1**.

In existing literature on rational inattention, the formulation of rational inattention decision problems follow the specification of **DP-1** and action strategies similar to equation (9) are what researchers can obtain at best. No more explicit solutions for action strategies are found in the existing literature on rational inattention. Without a more explicit solution for action strategies, profound analytical or computational analyses are hard to implement. This paper would also end here without our findings. In this paper, we will derive and exhibit the closed form of action strategies $f(\alpha^*|\theta)$ and $f(\alpha^*)$. The insights in this paper on how limited cognitive ability in information acquisition affects ordering decisions benefits from these closed form formulas.

The optimal ordering strategy $f(\alpha^*|\theta)$ describes the randomness in the manager's ordering decisions with respect to the accurate demand information. If we follow the free energy principle to understand the information acquisition activity in our model, the randomness reflects the neuronal

noise. λ can reflect the manager's ability to interpret primitive information to acquire demand information. If λ equals 0 at optimum, the manager's information acquisition ability is perfect. In this situation, he can accurately obtain the underlying demand information, which also equals ex-post optimal ordering quantity. For $\lambda > 0$, it indicates that the decision maker's information acquisition ability is limited. The information acquisition process reduce the neuronal noises that covers the accurate knowledge of the underlying state θ . Some neuronal noises are left in the acquired information and hence managers' ordering decisions exhibit a probabilistic form. Therefore, the action response α^* towards the accurate demand information θ is random.

Before we exhibit the closed-form solutions of $f(\alpha^*|\theta)$ and $f(\alpha^*)$, which are contributions of this paper, we have to note that after the information acquisition process, not all optimal ordering quantities α^* can be chosen with a positive likelihood. By studying rational inattention discrete choice problems, Caplin, Dean and Leahy (2019) find that an optimal consideration set of choices emerges after the decision makers' information acquisition processes. The optimal consideration set is a subset of all choices, and in the optimal consideration set, all choices can be chosen with a positive likelihood given any state. For rational inattention decision models where states variables follows a Borel measure and decision variables are continuous, Jung, King, Matějka and Sims (2019) prove that after information acquisition, only a countable set of points with no limit points can be considered as the candidates of ultimate decisions. Prior to Jung, King, Matějka and Sims' (2019) paper, Matějka (2016) obtained a similar results in an application where the seller is rationally inattentive to sales prices. In Matějka's (2016) model, the state follows a continuous distribution and the decision variable is also continuous. By following the proof approach of Caplin, Dean and Leahy (2019), we obtain a necessary and sufficient condition that confirms Jung, King, Matějka and Sims' (2019) result of the emergence of a countable set of points as potential optimal decisions in the continuous-state continuous-action setting, but also the mathematical formulation of our condition looks similar to its counterpart of rational inattention discrete choice models in Caplin, Dean and Leahy (2019):

Proposition 1: Define set D such that $D = \{\alpha^* | f(\alpha^*) \geq 0\}$ and $F(\alpha^*) = \int_{-\infty}^{+\infty} \frac{\exp(\frac{\pi(\theta, \alpha^*)}{\lambda})}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{\pi(\theta, \alpha^*)}{\lambda}) d\alpha^*} g(\theta) d\theta$. For $\alpha^* \notin D$, $F(\alpha^*) \leq 1$. For $\alpha^* \in D$, $F(\alpha^*) = 1$. $\pi(\theta, \alpha^*)$ can be equal to $\pi(\theta, \alpha^*) = p \min(\theta, \alpha^*) - c\alpha^*$.

In Proposition 1, the result is not changed by the form of $\pi(\theta, \alpha^*)$. It can be the profit function considered in our paper, or any other types of profit or utility functions. Because $F(\alpha^*)$ is analytic, therefore the solutions of $F(\alpha^*) = 1$ are scattered points on real line. Therefore,

Corollary 1: Supposing equation $F(\alpha^*) = 1$ have solutions, $f(\alpha^*)$ is a discrete distribution. Accordingly,

$f(\alpha^*|\theta)$ is discrete as well.

Suppose $f(\alpha^*)$ follows a binary distribution. That is set D only contains two elements and we denote them by α_1^* and α_2^* . Given that $\pi(\theta, \alpha^*) = p \min(\theta, \alpha^*) - c\alpha^*$, we obtain the closed-form solution of $f(\alpha^*)$:

Proposition 2: Suppose $F(\alpha^*) = 1$ have two solutions α_1^* and α_2^* , and hence $f(\alpha^*)$ follows a binary distribution, i.e. $\alpha_1^*, \alpha_2^* \in D$ and $f(\alpha_1^*) + f(\alpha_2^*) = 1$. The closed-form of $f(\alpha^*)$ is

$$f(\alpha_1^*) = \frac{T^{(n)}(\theta) - \frac{\partial^n \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^n}}{\frac{\partial^n \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda})}{\partial \theta^n} - \frac{\partial^n \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^n}} \quad (13)$$

where

$$T(\theta) \equiv \frac{g(\theta)}{\xi \exp(-\frac{p-c}{\lambda}\theta) + ((\eta(\theta) - \varepsilon(\theta))\frac{p-c}{p} - (\eta'(\theta) - \varepsilon'(\theta))\frac{\lambda}{p})\theta \exp(-\frac{p-c}{\lambda}\theta)}$$

$$\varepsilon(\alpha^*) = \frac{c}{\lambda} \frac{\partial F(\alpha^*)}{\partial \alpha^*} \text{ and } \eta(\alpha^*) = \frac{\partial^2 F(\alpha^*)}{\partial \alpha^{*2}}. \xi \text{ is a parameter of the equation } \frac{g(\theta)}{A(\theta)} = \xi \exp(-\frac{p-c}{\lambda}\theta) + ((\eta(\theta) - \varepsilon(\theta))\frac{p-c}{p} - (\eta'(\theta) - \varepsilon'(\theta))\frac{\lambda}{p})\theta \exp(-\frac{p-c}{\lambda}\theta), \text{ where } A(\theta) = \int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{p \min(\theta, \alpha^*) - c\alpha^*}{\lambda}) d\alpha^*. n \text{ is the minimum rounds of differentiation such that } \frac{\partial T^{(n)}(\theta)}{\partial \theta} = 0, \frac{\partial \frac{\partial^n \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^n}}{\partial \theta} = 0 \text{ and } \frac{\partial \frac{\partial^n \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda})}{\partial \theta^n}}{\partial \theta} = 0.$$

Determining the exact value of ξ requires additional conditions. For example, as stated in Appendix, a specific value of ξ is known at a particular θ_0 such that $\xi = \frac{g(\theta_0)}{A(\theta_0)} \exp(\frac{p-c}{\lambda}\theta_0) - ((\eta - \varepsilon)\frac{p-c}{p} - (\eta' - \varepsilon')\frac{\lambda}{p})\theta_0$. From Proposition 2, we can see that whether or not a posterior exists after the information acquisition activity is not only up to whether $F(\alpha^*) = 1$ have any solution, but also depends on the form of the utility functions such that with enough rounds of differentiation, $\frac{\partial T^{(n)}(\theta)}{\partial \theta} = 0, \frac{\partial \frac{\partial^n \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^n}}{\partial \theta} = 0$ and $\frac{\partial \frac{\partial^n \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda})}{\partial \theta^n}}{\partial \theta} = 0$.

From Proposition 1 and Proposition 2, we can find that the determination of $f(\alpha^*)$ and D is invariant with any contingent state θ , but as a neurocognitive process, the information acquisition from θ varies the likelihood of any point in D chosen with the presence of θ , i.e. $f(\alpha^*|\theta)$. Therefore, the neurocognitive process' impact is evident on determining the likelihood conditional on the underlying state. Is it relevant with the determination of D and $f(\alpha^*)$, which is invariant with any contingent state θ ? From Proposition 2, we can see that there is relevance between $f(\alpha^*)$ with θ , which is present in the process of obtaining $f(\alpha^*)$, even if the ultimate form of $f(\alpha^*)$ is irrelevant with θ . The n rounds of differentiation are for eliminating the impact of θ on $f(\alpha^*)$. From Proposition 3, which is shown below, it can be known that given $f(\alpha^*)$, the value of points in D can be determined. Therefore, the neuronal process of information acquisition is relevant with the determination of $f(\alpha^*)$ and D . The neuronal process of determining $f(\alpha^*)$ and D is eliminating the impact of θ in the two objects. According to the definition that $f(\alpha^*) = \int_{-\infty}^{+\infty} f(\alpha^*|\theta)g(\theta)d\theta$, to eliminate

the impact of θ in $f(\alpha^*)$, integration is used. That is, determining $f(\alpha^*)$ by considering the likelihood that a given α^* is chosen across all demand states. But in the implementation of information acquisition as a neuronal activity, the brain will constantly differentiate the impact of θ in $f(\alpha^*)$ to eliminate its impact in it. Differentiation is a process of exploring the underlying properties of a function or an object in deeper layers. The elimination of the impact of θ by differentiation is therefore a process to exhausting all its impact on $f(\alpha^*)$. In the rational inattention decision model, $f(\alpha^*)$ determines both $f(\alpha^*|\theta)$ and D . Equation (13) reveals the mechanism of the neuronal implementation of the information acquisition activity.

For identifying the value of α_1^* and α_2^* , because both α_1^* and α_2^* satisfy that $F(\alpha^*) = \int_{-\infty}^{+\infty} \frac{\exp(\frac{\pi(\theta, \alpha^*)}{\lambda})g(\theta)}{f(\alpha_1^*) \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda}) + f(\alpha_2^*) \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})} d\theta = 1$, therefore the equation $F(\alpha^*) = 1$ can be used as a condition to identify α_1^* and α_2^* . Obviously, we need additional conditions to identify α_1^* and α_2^* .

The next condition can come from the following approach. That is sampling data to estimate $f(\alpha_1^*)$ or $f(\alpha_2^*)$. Based on the estimates of $f(\alpha_1^*)$ or $f(\alpha_2^*)$, equation (13) then becomes an equation with two unknown variables α_1^* and α_2^* . Such modified equation (13), together with $F(\alpha^*) = 1$, can be used to identify $f(\alpha_1^*)$ and $f(\alpha_2^*)$. We call this approach by *empirical approach*.

However, although the empirical approach is in theory feasible, implementing the empirical approach may not be that promising. A salient problem comes from the technical errors, for example the data sampling errors. It poses a concern in particular when we try to estimate $f(\alpha_1^*)$ and $f(\alpha_2^*)$. We will discuss the impact of technical errors on decision maker's information acquisition effort in detail in next section. As we will exhibit in the three examples with specific technical error distributions, the information acquisition efforts are very easy to be eclipsed by the impact of technical errors. If α_1^* and α_2^* are the optimal ordering quantities determined by the decision maker's information acquisition effort without the influence of the technical errors, then once the technical errors take impact, the result can be shifted so that other value of optimal ordering quantities are more likely to choose than α_1^* and α_2^* . Some other values of optimal ordering quantities are obtained in the estimated posterior distribution and there may be more than two discrete points of optimal ordering quantities in the estimated posterior distribution. Therefore, if we cannot separate the impact of technical errors in the estimation of posterior distribution, the empirical approach cannot work well. Although we will not delve into this topic comprehensively, which is not the purpose of this paper, but we think instrumental variable method can be applied to separate the impact of technical errors from the estimation of posterior distribution.

In the following, we give an example to explain how empirical approach works to identify α_1^* and α_2^* . It adopts the indirect approach to obtain $f(\alpha_1^*)$ or $f(\alpha_2^*)$ from an alternative estimate. The technical error distribution has a particular feature that helps us analytically obtain $f(\alpha^*)$ from the estimate. The derivation of $f(\alpha^*)$ also needs the one-to-one correspondence relationship between signal and action. We describe the relationship by function $\alpha^* = m(s)$, which we will discuss in detail in next section as well.

Proposition 3: Suppose the technical error distribution $f_\varepsilon(\varepsilon)$ satisfies that $f_\varepsilon(\varepsilon_1 + \varepsilon_2) = f_\varepsilon(\varepsilon_1)f_\varepsilon(\varepsilon_2)$. Then, $\alpha_1^*, \alpha_2^* \in D$ are

the solutions of the following equation group:

$$\begin{cases} \int_{-\infty}^{+\infty} \frac{\exp(\frac{\pi(\theta, \alpha^*)}{\lambda})g(\theta)}{f(\alpha_1^*) \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda}) + f(\alpha_2^*) \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})} d\theta = 1 \\ f(\alpha_1^*) = \frac{\gamma - f_\varepsilon(-m^{-1}(\alpha_2^*))}{f_\varepsilon(-m^{-1}(\alpha_1^*)) - f_\varepsilon(-m^{-1}(\alpha_2^*))} \end{cases} \quad (14)$$

where $f(\alpha_1^*)$ is given by equation (13) and $f(\alpha_2^*) = 1 - f(\alpha_1^*)$. γ is obtained by estimating the linear relationship between $f(\alpha^{*'})$ and $f_\varepsilon(m^{-1}(\alpha^{*'}))$, i.e. $f(\alpha^{*'}) = \gamma f_\varepsilon(m^{-1}(\alpha^{*'})) + v$, where v is the estimation error. $m(\cdot)$ is the one-to-one correspondence relationship between signal and action. If the above equation group has solutions, then $\alpha_1^*, \alpha_2^* \in D$ exist and can be obtained by solving the equation group. Otherwise, there does not exist a binary form of $f(\alpha^*)$. \square

There are some distributions that fit the feature that $f_\varepsilon(\varepsilon_1 + \varepsilon_2) = f_\varepsilon(\varepsilon_1)f_\varepsilon(\varepsilon_2)$. For example, the following Weibull distribution satisfies the feature:

$$f_\varepsilon(\varepsilon) = \begin{cases} \frac{1}{b} \exp(-\frac{\varepsilon}{b}) & \varepsilon \geq 0 \\ 0 & \varepsilon < 0 \end{cases}$$

where $b > 0$. The empirical approach can be applied to identify more general forms of posterior distributions beyond just a binary distribution, but apparently it requires the theoretical form of $f(\alpha^*)$. Just like in our binary posterior example, we have equation (13), and therefore we can apply the empirical approach together with the theoretical condition to identify α_1^* and α_2^* . Both conditions obtained theoretically and empirically are necessary for the identification. Under what conditions do the identification conditions work? For example, in some circumstances, equation group (14) may not have solutions. In the following, we give a sufficient condition for the existence of α_1^* and α_2^* in equation group (14) to make the identification work.

Corollary 2: If there exists a finite interval $[a, b] \subset (0, +\infty)$ of θ such that $g(\theta) \geq 0$ for $\theta \in [a, b]$ and $g(\theta) = 0$ otherwise, then according to Brouwer's fixed point theorem, equation group (14) has at least one solution of $(\alpha_1^*, \alpha_2^*) \in [a, b] \times [a, b]$. \square

From above results and discussions, we can see that the determination of the set D is irrelevant with the contingent state θ . It is because when the managers think about what candidate optimal ordering quantities they may choose, they think about the problem across all states. The contingent state only matters with the likelihood that the optimal ordering quantities concluded after information acquisition are chosen given a state. Therefore, an effective thinking can only be that the optimal ordering quantity that is most close to the state chosen with the highest likelihood. In a binary distribution, it means if θ is closer to α_2^* than to α_1^* , i.e. $|\alpha_2^* - \theta| < |\alpha_1^* - \theta|$, in the manager's mind $f(\alpha_2^*|\theta) > f(\alpha_1^*|\theta)$. Likewise, if θ is closer to α_1^* than to α_2^* , i.e. $|\alpha_1^* - \theta| < |\alpha_2^* - \theta|$, in the manager's mind $f(\alpha_1^*|\theta) > f(\alpha_2^*|\theta)$. However, we find that such an effective thinking in the form of a binary posterior is hard to achieve directly.

Corollary 3: Given a binary posterior such that $D = \{\alpha_1^*, \alpha_2^* | \alpha_1^* < \alpha_2^*\}$, an effective thinking such that if

$|\alpha_2^* - \theta| < |\alpha_1^* - \theta|$, $f(\alpha_2^*|\theta) > f(\alpha_1^*|\theta)$ or if $|\alpha_1^* - \theta| < |\alpha_2^* - \theta|$, $f(\alpha_1^*|\theta) > f(\alpha_2^*|\theta)$ can be obtained if and only if the parameter specifications can lead to the following result:

$$\left\{ \begin{aligned} \frac{\alpha_1^* + \alpha_2^*}{2} &= \frac{\lambda}{p} \ln \left[\frac{f(\alpha_1^*)}{f(\alpha_2^*)} \exp\left(\frac{p\alpha_1^* - c(\alpha_1^* - \alpha_2^*)}{\lambda}\right) \right] \\ \frac{\exp(\frac{c\alpha_2^*}{\lambda})}{\exp(\frac{c\alpha_1^*}{\lambda})} &> \frac{f(\alpha_2^*)}{f(\alpha_1^*)} > \frac{\exp(\frac{(p-c)\alpha_1^*}{\lambda})}{\exp(\frac{(p-c)\alpha_2^*}{\lambda})} \end{aligned} \right\}$$

However, almost surely, parameters are hard to satisfy the above conditions.

Therefore, according to Corollary 3, a binary posterior is hard to directly deliver an effective thinking. Our result in Corollary 3 is determined by the rational inattention theory as well as the specification of the profit function. It can be expected that if the formulation of the profit function is changed, the result in Corollary 3 is changed such that an effective thinking can be directly achieved by a binary posterior.

A Suggestion to Correct the Cognitive Biases: However, an effective thinking can be achieved indirectly under a binary posterior distribution. As we see from Corollary 3, if the managers' information acquisition activities lead to a binary posterior, it is almost certain that the candidate optimal ordering quantity that is closer to the state is chosen with lower likelihood, which makes their thinking ineffective. However, if the managers can think one more step, i.e. when the direct instinct tells them that a candidate optimal ordering quantity is less favorable to choose, then it is just the optimal ordering quantity that is closer to the state. In this way, their cognitive biases under the binary posterior after information acquisition can be corrected.

Corollary 4: For a posterior such that $D = \{\alpha_1^*, \alpha_2^*, \dots, \alpha_n^* | \alpha_1^* < \alpha_2^* < \dots < \alpha_n^*\}$, an effective thinking such that the candidate optimal ordering quantity α^* that is most close to θ chosen with highest likelihood is hard to achieve as well.

Therefore, in a general setting, even if they think about more candidates of optimal ordering quantities beyond just two, the managers do not have an effective thinking as well. Can we think that in the general setting, it is the optimal ordering quantity α^* that is furthest from the state chosen with highest likelihood? Not necessarily. Therefore, the best way, or say the most succinct and efficient way of thinking for managers to choose the optimal ordering quantity are try their best to refine their thinking to focus on just two candidate optimal ordering quantities, as long as the managers can, and use the suggestion proposed by us to correct their cognitive biases which naturally arises almost surely as the result of a cognitive activity.

5 How Do Technical Errors Affect Optimal Ordering Decisions by Information Acquisition?

In this section, we consider a setting where the technical errors are in the primitive information, and analyze how the technical errors affect the manager's optimal ordering decision via its

influence on the acquired information on demand. For acquiring information about a demand state θ , it is inevitable that the primitive information contain some technical errors, for example the errors from data collection, computing machines' physical errors, and the estimation errors caused by the imperfect knowledge about the market, etc. Hence, technical errors are prevalent and considering its impact in the information acquisition process is more realistic. Due to the existence of technical errors in the primitive information, it is challenging for the manager to acquire the accurate information about the state, let alone his information acquisition ability limited. For analytical feasibility, we assume the state reflected by the primitive information (θ') is the true demand state (θ) additively affected by an exogenous white noise ε (the technical error), i.e. $\theta' = \theta + \varepsilon$. The probability density distribution of ε is given by $f_\varepsilon(\varepsilon)$. As well, the interpreted information towards an underlying demand θ will be $s' = s + \varepsilon$. In this section, all signals s and the associated signal set S are at the optimum.

According to this specification, the technical error ε in the demand estimate θ' will be completely transferred to the acquired information s' , and the core of s' that reflects the underlying demand information is still s , which is the acquired information of θ when there is no technical error in the primitive information.

s is the complete and the sole substance that reflects the potential demand θ . Due to the existence of the white noise ε in the primitive information, the manager's acquired information is impacted by the white noise and hence the substance of s is hard to keep completely. The specification of signals $s' = s + \varepsilon$ is one way to model the impact of technical errors on acquired information. This specification keeps the influence of ε at its best in the information acquisition process.

Intuitively, when interpreting the primitive information to acquire demand information, although the manager cannot discern ε from θ , he is able to eliminate part of the influence of ε in the interpreted information on demand, and hence the influence of ε in s' is weakened. However, by experimenting several types of technical errors, we find that in some situations, the technical errors' influence in the acquired information is so dominant that the genuine information acquisition activity by the manager is completely overshadowed. Although in our specification the state is independent from the technical error, however we find that the acquired information depend on the specification of technical errors. It means, the type of or the distribution of the technical errors determines the specific form of the information managers acquire from the primitive information on demand.

Define function $m(\cdot) \equiv \arg \max_{\alpha} \int_{-\infty}^{+\infty} [p \min(\theta, \alpha) - c\alpha] f(\theta|\cdot) d\theta$. Thus, $\alpha^* = m(s)$. Define $\alpha^{*'} as the optimal ordering quantity when technical error is incorporated in the acquired information on demand, i.e. the optimal ordering quantity corresponding to s' . Thus, $\alpha^{*' = m(s')$. Because we consider the signal set S at the optimum that leads to the optimal ordering quantities α^* , therefore function $m(\cdot)$ is an expression of the one-to-one correspondence between signal and ordering quantities at the optimum. According to its one-to-one feature, this function is monotonic, and hence it is invertible such that $s = m^{-1}(\alpha^*)$.$

Scaling the value of signals or equivalently scaling of the interpreted results means endowing numerical value to signals. It includes two parts: determining the meaning of signals and the associated value of signals. The scaling is reflected by the

functional form of $m(\cdot)$.

$m(\cdot)$ can be either increasing or decreasing. Whether $m(\cdot)$ is increasing or decreasing reflects the meaning of signals, or literally how the manager interprets the primitive information. The value of signal can be scaled from different perspectives. For example, if s is interpreted as the knowledge “how many people will buy the product,” then α^* and s should exhibit a positive relationship and accordingly $m(\cdot)$ is increasing. If s is interpreted as the knowledge “how many people will not buy the product,” then α^* and s should exhibit a negative relationship and accordingly $m(\cdot)$ is decreasing. No matter whether s reflects “how many people will buy the product” or “how many people will not buy the product,” they are the knowledge around demand θ , but the concept “how many people will buy the product” directly follows the sense of θ , while the concept “how many people will not buy the product” is opposite to θ .

However, supposing the manager’s posterior of his information acquisition activity follows a binary form, by considering the additive technical errors, whether the manager can interpret the primitive information by following the direct or opposite sense of demand as well as what scale he can endows to the signal, i.e. the functional form of $m(\cdot)$, is determined by the following condition:

Proposition 4: Supposing the manager obtains a binary posterior distribution after information acquisition, and it exhibits the form as given in Corollary 3. Therefore, by incorporating an additive white noise ε as the technical error, which follows a density function $f_\varepsilon(\varepsilon)$, the distribution of the manager’s optimal ordering strategy by incorporating the technical error is given by

$$f(\alpha^{*'}) = \sum_{\alpha^* \in D} f_\varepsilon(m^{-1}(\alpha^{*'}) - m^{-1}(\alpha^*))f(\alpha^*) \quad (15)$$

Due to the technical error’s additive existence in the primitive information on demand, the functional form of $m(\cdot)$ has to satisfy $\int_{-\infty}^{+\infty} f(\alpha^{*'})d\alpha^{*'} \equiv 1$.

Therefore, $f(\alpha^{*'})$ depends on $f_\varepsilon(\varepsilon)$. Hence, even if the technical errors can be physically independent from the true state in the primitive information, the information acquisition as a cognitive activity cannot be independent from the technical errors. In the following, we consider three technical error distributions to discuss the insight provided in our Proposition 4 in detail. Besides, we focus on a specific type of interpretation, that is $m(\cdot)$ follows a linear form - $\alpha^* = \frac{1}{k}s$ where $k \neq 0$. If $k > 0$, then the manager interprets the primitive information by following the direct sense of demand, while if $k < 0$, the manager interprets the primitive information by following the opposite sense of demand. Correspondingly, we also call an interpretation where $k > 0$ by a positive interpretation and $k < 0$ by a negative interpretation. The absolute value of k reflects the scale the manager endows to his acquired information. Accordingly, we call a large absolute value of k a wild interpretation and a small absolute value of k a mild interpretation.

In the following, for clarification, we denote $f(\alpha^{*'}) = \alpha_1^*$ and $f(\alpha^{*'}) = \alpha_2^*$ as the likelihood such that by considering the technical errors, the chances that α_1^* and α_2^* , which belong to D as described in Corollary 3, are chosen. The probabilities $f(\alpha^{*'} = \alpha_1^*)$ and $f(\alpha^{*'} = \alpha_2^*)$ are different from $f(\alpha_1^*)$ and

$f(\alpha_2^*)$, which are given in Proposition 2.

1. $f_\varepsilon(\varepsilon)$ follows an exponential distribution: Therefore, $f_\varepsilon(\varepsilon)$ is given by

$$f_\varepsilon(\varepsilon) = \begin{cases} \mu \exp(-\mu\varepsilon) & \varepsilon \geq 0 \\ 0 & \varepsilon < 0 \end{cases}$$

Define $f_e(\alpha^{*'}; \alpha_1^*) = \mu \exp(-\mu k(\alpha^{*'} - \alpha_1^*))f(\alpha_1^*)$ and $f_e(\alpha^{*'}; \alpha_2^*) = \mu \exp(-\mu k(\alpha^{*'} - \alpha_2^*))f(\alpha_2^*)$. Therefore, if $k > 0$,

$$f(\alpha^{*'}) = \begin{cases} 0 & \alpha^{*'} < \alpha_1^* \\ f_e(\alpha^{*'}; \alpha_1^*) & \alpha_1^* \leq \alpha^{*'} < \alpha_2^* \\ f_e(\alpha^{*'}; \alpha_1^*) + f_e(\alpha^{*'}; \alpha_2^*) & \alpha^{*'} \geq \alpha_2^* \end{cases}$$

If $k < 0$,

$$f(\alpha^{*'}) = \begin{cases} 0 & \alpha^{*'} < 0 \\ f_e(\alpha^{*'}; \alpha_1^*) + f_e(\alpha^{*'}; \alpha_2^*) & 0 \leq \alpha^{*'} < \alpha_1^* \\ f_e(\alpha^{*'}; \alpha_2^*) & \alpha_1^* \leq \alpha^{*'} < \alpha_2^* \\ 0 & \alpha^{*'} \geq \alpha_2^* \end{cases}$$

where k takes value such that $\int_{-\infty}^{+\infty} f(\alpha^{*'})d\alpha^{*'} = 1$. We find that

Proposition 5: If the technical errors follow an exponential distribution as expressed above, $f(\alpha^{*'})$ is always decreasing for $\alpha^* \in [\alpha_1^*, +\infty)$ if $k > 0$ and always increasing for $\alpha^* \in [0, \alpha_2^*)$ if $k < 0$. If $k > 0$, it always exists and it is unique. If $k < 0$, it exists if and only if $f(\alpha_1^*)\alpha_1^* + f(\alpha_2^*)\alpha_2^* < \frac{1}{\mu}$ and it is unique. \square

Therefore, given any exponential distribution, no matter what value the parameter μ is, the manager can always find a positive way to interpret the primitive information. Whether or not a negative interpretation is available is up to whether the condition $f(\alpha_1^*)\alpha_1^* + f(\alpha_2^*)\alpha_2^* < \frac{1}{\mu}$ can be satisfied. The monotonicity of $f(\alpha^{*'})$ presented in Proposition 5 implies that, if the technical error follows an exponential distribution, its influence is so powerful that the influence hides the managers’ information acquisition results from a pure primitive information source which is not contaminated by the technical errors.

2. $f_\varepsilon(\varepsilon)$ follows a truncated normal distribution: Suppose the technical error ε follows a truncated normal distribution $TN(0, \sigma^2; 0, +\infty)$, which is modified based on a normal distribution $N(0, \sigma^2)$ by making the truncated density positive for $\varepsilon \in [0, +\infty)$. That is,

$$f_\varepsilon(\varepsilon) = \begin{cases} \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{\varepsilon^2}{2\sigma^2}) & \varepsilon \in [0, +\infty) \\ 0 & \text{otherwise} \end{cases}$$

Define $f_t(\alpha^*; \alpha_1^*) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_1^*)^2)f(\alpha_1^*)$ and $f_t(\alpha^*; \alpha_2^*) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_2^*)^2)f(\alpha_2^*)$. Therefore,

$$f(\alpha^*) = \begin{cases} 0 & \alpha^* < \alpha_1^* \\ f_t(\alpha^*; \alpha_1^*) & \alpha_1^* \leq \alpha^* < \alpha_2^* \\ f_t(\alpha^*; \alpha_1^*) + f_t(\alpha^*; \alpha_2^*) & \alpha^* \geq \alpha_2^* \end{cases}$$

We find that $f(\alpha^*)$ cannot be continuous by any means once ε follows the error distribution $TN(0, \sigma^2; 0, +\infty)$. $f(\alpha^*)$ jumps at $\alpha^* = \alpha_1^*$ and $\alpha^* = \alpha_2^*$, which are the local maxima of $f(\alpha^*)$. Specifically

Proposition 6: Given the error distribution $TN(0, \sigma^2; 0, +\infty)$, k^2 always exists and is unique. If $f(\alpha_1^*) > f(\alpha_2^*)$ and $k^2 = \frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}$, $f(\alpha^* = \alpha_1^*) = f(\alpha^* = \alpha_2^*) > f(\alpha^* \neq \alpha_1^* \text{ or } \alpha_2^*)$.

If $f(\alpha_1^*) > f(\alpha_2^*)$ and $k^2 > \frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}$, $f(\alpha^* = \alpha_1^*) > f(\alpha^* = \alpha_2^*)$. In this situation, for $\alpha^* \in [\alpha_1^*, \alpha_1^* + \sqrt{\frac{2\sigma^2}{k^2} \ln \frac{1}{\exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2) + \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}}]$, $f(\alpha^*)$ is greater than $f(\alpha^* = \alpha_2^*)$.

If $f(\alpha_1^*) \leq f(\alpha_2^*)$, or if $f(\alpha_1^*) > f(\alpha_2^*)$ and $k^2 < \frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}$, $f(\alpha^* = \alpha_1^*) < f(\alpha^* = \alpha_2^*)$.

For $\alpha^* \in [\alpha_2^*, \max \hat{\alpha}^*)$, $f(\alpha^*) > f(\alpha_1^*)$, where $\max \hat{\alpha}^*$ is the largest solution of the equation $\exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_1^*)^2)f(\alpha_1^*) + \exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_2^*)^2)f(\alpha_2^*) = f(\alpha^* = \alpha_1^*)$, which contains four solutions.

Given a technical error distribution, the manager can always find two ways to interpret the primitive information. One follows the direct sense of the demand and the other follows the opposite sense of demand. However, no matter how the manager interprets the acquired information, the posterior distribution is always the same. Unlike Case 1 where technical errors follows an exponential distribution which completely hides the managers' information acquisition attempt from a pure primitive information, in situations where technical errors follows the given truncated normal distribution, the noise cannot completely hide the optimal ordering quantities that are potential to choose without the influence of the technical errors. Specifically, α_2^* is more robust than α_1^* to resist the impact from the technical error. As long as $f(\alpha_2^*) > f(\alpha_1^*)$, even if there exists the influence of the technical error that follows the truncated normal distribution, α_2^* is still the most likely chosen ordering quantity under the influence of the technical error. However, if $f(\alpha_1^*) > f(\alpha_2^*)$, whether α_1^* or α_2^* is the most likely chosen ordering quantity subject to the technical error is up to how the decision maker interprets the primitive information, i.e. the value of k . In this situation, if the interpretation is wild, i.e. $k < -\sqrt{\frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}}$ or $k > \sqrt{\frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}}$,

it is still α_1^* that is most likely chosen under the influence of the technical error. However, if the interpretation is mild, i.e.

$$-\sqrt{\frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}} < k < \sqrt{\frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}}$$

the result of the most likely chosen ordering quantity is shifted to α_2^* . Therefore, if $f(\alpha_1^*) > f(\alpha_2^*)$, only when the interpretation from the primitive information is wild enough, the most likely ordering quantity obtained when information acquisition not affected by technical errors can be kept under the impact of technical errors.

3. $f_\varepsilon(\varepsilon)$ follows a Weibull distribution: Suppose the technical error follows a Weibull distribution, which is given by

$$f_\varepsilon(\varepsilon) = \begin{cases} \frac{a}{b^a} \varepsilon^{a-1} \exp(-(\frac{\varepsilon}{b})^a) & \varepsilon \geq 0 \\ 0 & \varepsilon < 0 \end{cases}$$

where $a > 0$ and $b > 0$. Define $f_w(\alpha^*; \alpha_1^*) = \frac{a}{b^a} k^{a-1} (\alpha^* - \alpha_1^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_1^*}{b})^a) f(\alpha_1^*)$ and $f_w(\alpha^*; \alpha_2^*) = \frac{a}{b^a} k^{a-1} (\alpha^* - \alpha_2^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_2^*}{b})^a) f(\alpha_2^*)$. Therefore,

$$f(\alpha^*) = \begin{cases} 0 & \alpha^* < \alpha_1^* \\ f_w(\alpha^*; \alpha_1^*) & \alpha_1^* \leq \alpha^* < \alpha_2^* \\ f_w(\alpha^*; \alpha_1^*) + f_w(\alpha^*; \alpha_2^*) & \alpha^* \geq \alpha_2^* \end{cases}$$

where k take values such that $\int_{-\infty}^{+\infty} f(\alpha^*) d\alpha^* = 1$ given a and b .

If the technical error distribution follows a Weibull distribution, its influence is also very strong such that can mistaken the information acquisition result from a primitive information which is not contaminated by the technical error. We find that

Proposition 7: If the technical error distribution $f_\varepsilon(\varepsilon)$ follows a Weibull distribution as given above, then if parameter a is an even integer, k can only be positive if it exists. If parameter a is an odd integer, a unique $k < 0$ exists and $k > 0$ does not exist for sure. Given the technical error distribution, if $a \leq 1$, k exists and it is unique and positive.

Therefore, if the technical error distribution follows a Weibull distribution, for a that is an even integer or $a \leq 1$, the interpretation can only follow the direct sense of demand. Proposition 7 also implies that k may not exist if the technical error distribution follows a Weibull distribution. The error distribution determines the existence and valuation of k . The one-to-one correspondence $\alpha^* = m(s)$ always exist. If the error distribution makes k not exist, it indicates that with the given type of error distribution, linear interpretation does not work. If k exists, it indicates that the linear interpretation works subject to the error distribution. The example shows that it is the error distribution that restricts the interpretation into specific forms. In particular, it restricts whether the decision maker will interpret the acquired information by following the direct sense of the underlying state or the opposite sense of the underlying state. Therefore, the example of Weibull distribution as technical error distribution proves that, with the error distribution, the one-to-one correspondence does not exist in some forms. The error distribution restricts the available forms of interpretation at optimum.

Define $\hat{\alpha}_1^* = \alpha_1^* + (\frac{a-1}{a} \frac{b^a}{k^a})^{\frac{1}{a}}$ and $\hat{\alpha}_2^* = \alpha_2^* + (\frac{a-1}{a} \frac{b^a}{k^a})^{\frac{1}{a}}$. $\hat{\alpha}^*$ represent(s) the solutions of the following equation with respect to

$\alpha^{*'}$

$$\frac{a-1}{a} \frac{b^a}{k^a} = \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a \quad (16)$$

where $\phi(\alpha^{*'}) = \frac{u_1}{u_1 + u_2}$, $u_1 = f(\alpha_1^*) \exp(-k^a (\frac{\alpha^{*'} - \alpha_1^*}{b})^a) (\alpha^{*'} - \alpha_1^*)^{a-2}$ and $u_2 = f(\alpha_2^*) \exp(-k^a (\frac{\alpha^{*'} - \alpha_2^*}{b})^a) (\alpha^{*'} - \alpha_2^*)^{a-2}$. The above equation may have $n > 1$ solutions, and we rank all of the solutions according to their value by an ascending order in the following:

$$\hat{\alpha}_{(1)}^* < \hat{\alpha}_{(2)}^* < \dots < \hat{\alpha}_{(m)}^* < \hat{\alpha}_{(m+1)}^* < \dots < \hat{\alpha}_{(n)}^*$$

$\hat{\alpha}^*$ are the local optima of $f(\alpha^{*'})$, which include local minima and local maxima. It is the parameter k that determines how many solutions there are in equation (16). Therefore, like the case where the technical error distribution follows a truncated normal distribution, in this example, we can also find that how to interpret from primitive information determines ordering behavior under the influence of technical errors at optimum.

We are interested in more detailed properties of the technical errors' impact on decision makers' information acquisition activity. Given the noises following the Weibull distribution, we study a case where $\hat{\alpha}_1^* > \alpha_2^*$, i.e. $k^a \leq \frac{a-1}{a} \frac{b^a}{(\alpha_2^* - \alpha_1^*)^a}$. By studying this case, we find that $k^{a-1} > 0$; if $a < 1$, $f(\alpha^{*'})$ equals 0 for $\alpha^* \in [\alpha_1^*, \alpha_2^*)$ with discontinuity at $\alpha^{*'} = \alpha_2^*$; for the case where a is an odd integer and $k < 0$ and the case where $a < 1$, $f(\alpha^{*'})$ is unimodal and the maximum of $f(\alpha^{*'})$ is taken at $\alpha^{*'} = \alpha_2^*$.

For $k > 0$,

1. If $a > 1$, $f(\alpha^{*'})$ increases with respect to $\alpha^{*'} for $\alpha^* \in [\alpha_1^*, \hat{\alpha}_1^*)$, and decreases with respect to $\alpha^{*'} for $\alpha^* \in [\hat{\alpha}_2^*, +\infty)$. If $a < 1$, $f(\alpha^{*'})$ increases with respect to $\alpha^{*'} for $\alpha^* \in [\alpha_2^*, \hat{\alpha}_1^*)$ and decreases with respect to $\alpha^{*'} for $\alpha^* \in [\hat{\alpha}_2^*, +\infty)$.$$$$

2. No matter whether $(-1)^{a-1} \geq 0$, n is an odd integer with generality.

3. No matter whether $a \geq 1$, with generality, for $\alpha^{*'} \in (\hat{\alpha}_m^*, \hat{\alpha}_{m+1}^*)$, if m is an odd integer, $f'(\alpha^{*'}) \leq 0$, and if m is an even integer, $f'(\alpha^{*'}) \geq 0$.

4. No matter whether $a \geq 1$, with generality, there are $\frac{n+1}{2}$ ordering quantities that are chosen with locally highest likelihood, and they are located at $\alpha^{*'} \in [\hat{\alpha}_1^*, \hat{\alpha}_2^*]$. Because it is possible that $\frac{n+1}{2} \geq 2$, the technical error complicates the decision making compared with the situation where there is no technical error and are just two optimal ordering quantities in the posterior.

5. To ensure that local maxima exist for $\alpha^{*'} \in (\alpha_2^*, +\infty)$, it requires that $(a-1)k^a > 0$. Therefore, $a \neq 1$. Once $(a-1)k^a > 0$ is satisfied so that the local maxima exist for $\alpha^{*'} \in (\alpha_2^*, +\infty)$,

we can obtain that $\hat{\alpha}_1^* > \alpha_1^*$ and $\hat{\alpha}_2^* > \alpha_2^*$. Under the condition that $k^a \leq \frac{a-1}{a} \frac{b^a}{(\alpha_2^* - \alpha_1^*)^a}$, $\hat{\alpha}_1^* > \alpha_2^*$. Therefore, under the influence of technical errors, the potential optimal ordering quantities systematically upward bias to those without the influence of technical errors. Suppose a is an integer. The upward bias happens for $k \in (0, (\frac{a-1}{a})^{\frac{1}{a}} \frac{b}{(\alpha_2^* - \alpha_1^*)})$ if a is an odd integer, and for $k \in (-(\frac{a-1}{a})^{\frac{1}{a}} \frac{b}{(\alpha_2^* - \alpha_1^*)}, (\frac{a-1}{a})^{\frac{1}{a}} \frac{b}{(\alpha_2^* - \alpha_1^*)})$ if a is an even integer. Therefore, if a is an odd integer, a mild interpretation following the sense of demand makes the local maximum of $f(\alpha^{*'})$ systematically upward to α_1^* and α_2^* ; if a is an even integer, no matter whether the interpretation follows or is opposite to the sense of demand, an mild interpretation of the primitive information makes the local maximum of $f(\alpha^{*'})$ systematically upward to α_1^* and α_2^* .

Therefore, similar to the information acquisition activity subject to the technical errors following a truncated normal distribution, the impact of technical errors will eclipse the effort by a pure information acquisition activity due to a mild interpretation. The technical error distribution determines the existence and the specific forms of some interpretation. With the error distribution, the ordering decisions at optimum can systematically deviate from the decisions determined with a pure information acquisition activity without the impact of technical errors. The optimal ordering strategy at optimum subject to the technical error distribution can exhibit a multimode form instead of a unimode or a binary form. The multimode distribution form complicates the already biased decision making.

6 Summary and Recommendations for Managers

In this section, we summarize this paper and provide recommendations for managers to make their optimal decisions in inventory orderings in practice. We model information acquisition activities according to rational inattention theory proposed by Sims (1998, 2003, 2006) and the decision problem is presented by **IA-1**. However, information acquisition activity can be understood as a neuronal process according to free energy principle proposed by Friston (2009, 2010) and accordingly, the alternative version of the decision problem is presented by **IA-2**. For **IA-1** and **IA-2**, their expressions are equivalent, but the fundamental meanings are distinct. **IA-1** treats utility maximization as the decision maker's objective and the information acquisition activities assist the decision maker to fulfill this purpose. In **IA-2**, brain is the actual decision maker and it minimizes a free energy for obtaining a steady state. A well-known result in rational inattention theory literature is that there exists a one-to-one correspondence between action and signal at optimum, and this property is also held in **IA-1**. However, in **IA-2**, such a one-to-one correspondence is not necessarily held at steady state. In this paper, we focus on the decision problem formulated according to **IA-1**.

Proposition 1 is a counterpart result to the emergence of the optimal consideration set after information acquisition in rational inattention discrete choice problems in Caplin, Dean and Leahy (2019). Proposition 1 states that scattered points emerge in the action set as a result of information acquisition. The result confirms Jung, King, Matějka and Sims' (2019) conclusion of a countable set of points as potential optimal ordering decisions as a result

of information acquisition activities. Based on Proposition 1, corollary 1 shows that the optimal ordering strategy $f(\alpha^*|\theta)$ is a discrete distribution, so is the unconditional optimal ordering strategy $f(\alpha^*)$.

Proposition 2 presents the closed-form solution of $f(\alpha^*)$, which is given by equation (13), supposing the posterior after information acquisition activities is a binary distribution. The derivation of equation (13) reveals how information acquisition activity is implemented as a neuronal activity, and it implies that an optimal thinking is necessarily a process that *think until nothing to think*. Equation (13) together with estimated $f(\alpha^*)$ can be used as conditions to identify the optimal ordering quantities α^* in the posterior. We call such identification approach by *empirical approach*. However, the empirical approach cannot work well if the primitive information sources are contaminated by technical errors. Proposition 3 provides an example explaining how empirical approach works to identify α^* if $f(\alpha^*)$ is a binary distribution and the technical error distribution possesses the property that $f_\varepsilon(\varepsilon_1 + \varepsilon_2) = f_\varepsilon(\varepsilon_1)f_\varepsilon(\varepsilon_2)$. Weibull distribution, for instance, can possess such property under certain parameter specifications. Corollary 2 provides a condition that enables Proposition 3 to work.

Corollary 3 and Corollary 4 show that no matter whether $f(\alpha^*)$ is a binary discrete distribution or a general type of discrete distribution, the information acquisition activity is hard to directly deliver an effective thinking, which means the optimal ordering quantity closest to the demand state is hard to choose with highest likelihood. The failure to directly deliver an effective thinking is caused by the specification of profit function and the information acquisition activity modelled by rational inattention theory. Even though it cannot be obtained directly, an effective thinking by information acquisition can be obtained indirectly by our proposed method to correct the cognitive biases. Our method to correct the cognitive biases is that when thinking, people try their best to refine their thinking result to two possible outcomes, and the one they feel less likely to choose by direct instinct is just the optimal ordering quantity that is closer to the demand state. Therefore, to make thinking effective, it requires us to *think one more step*.

To study how technical errors affect optimal ordering decisions by information acquisition, we consider the case that the technical errors additively affect the state information. Given that the neuronal process of information acquisition results a binary distribution, Proposition 4 exhibits the general form of unconditional optimal ordering strategy if the primitive information sources are contaminated by technical errors. To delve into how technical errors affect information acquisition and resulted decisions, we experiment three technical error distributions by supposing that information acquisition activities generically result a binary distribution if the information sources are not contaminated and the one-to-one correspondence between action and signal is linear such that $\alpha^* = \frac{1}{k}s$. Technical error distributions $f_\varepsilon(\varepsilon)$ smooth out the posterior distribution obtained from the neuronal process of information acquisition from uncontaminated primitive information and we find that the specific form of $f_\varepsilon(\varepsilon)$ and the original result of information acquisition from uncontaminated information source, i.e. $f(\alpha^*)$, determine the shape of the posterior distribution by acquiring information from contaminated information sources, i.e. $f(\alpha^{*'})$.

If $f_\varepsilon(\varepsilon)$ follows an exponential distribution, Proposition 5 shows that a positive interpretation, i.e. $k > 0$, always exists and it is

unique. If $f_\varepsilon(\varepsilon)$ follows a truncated normal distribution, Proposition 6 shows that k^2 , a pair of opposite linear interpretations, always exist and the pair is unique. If $f_\varepsilon(\varepsilon)$ follows a Weibull distribution, our result depends on the parameter a of the presented Weibull distribution: if a is an even integer, only a positive interpretation can exist and the positive interpretation does not exist definitely; if a is an odd integer, only a negative interpretation (i.e. $k < 0$) exists and the interpretation is unique; if $a \leq 1$, only a positive interpretation exists and it is unique.

Based on the theoretical results obtained in this paper, we propose the following method for managers to make optimal inventory ordering decisions in practice subject to the influence of the technical errors in primitive information as well as the limitation of information acquisition as a neuronal activity. Because the research of this paper mainly focuses on the situation where the information acquisition activities result a binary posterior distribution, therefore our proposed method in the following concentrates on the situation where binary posteriors are the result of information acquisition:

1. Make the primitive information as clean as possible: as we find in this paper, technical errors can mistaken the information acquisition results from what the results should have been if the information sources were not contaminated by technical errors. Managers should ensure the cleanness of the primitive information by making the primitive information least contaminated by technical errors before they commence on information acquisition. Only the most clean primitive information can ensure a reliable information acquisition result.

2. Analyze the primitive information thoroughly: step two requires managers analyzing the primitive information obtained from step one thoroughly, i.e. *think until nothing to think*.

Analyzing the primitive information is just an information acquisition activity that we have studied in this paper. The information acquisition activities should be conducted several rounds. In each round, the results of analysis should be refined until managers do not think that extra analysis will improve their conclusions. Ideally, after each round of analysis, managers obtain a binary posterior as a result. When each round of analysis finishes, managers should wait next round of analysis until the impact of latest round of analysis is reduced to a trivial level so that each round of information acquisition is independent. At each round of analysis, managers acquire information from the same primitive information.

Each round of analysis concludes one posterior. The posterior includes a set of optimal ordering quantities and the feeling of the likelihood that each optimal ordering quantity is chosen by direct instinct. The posterior obtained in each round of analysis should be close to each other. Specifically, if managers obtain a binary posterior in each round of analysis, then the value of the less likely chosen optimal ordering quantity and the more likely chosen optimal ordering quantity, plus the feeling of the likelihood by direct instinct that each optimal ordering quantity is chosen, are close across all rounds of analysis. If such status is achieved, information acquisition can stop. The concluded posterior is the average of the results across all rounds of analysis.

If one posterior or few posteriors are divergent from the remaining posteriors, these divergent posteriors should be eliminated and the conclusion in this step is drawn based on the remaining posteriors. If many posteriors are divergent, this step, and even

step one, should be re-implemented until the majority of the posteriors are close to each other. The problem that many posteriors are divergent can be caused by various reasons, for example the primitive information is severely contaminated by technical errors.

3. Suppose managers obtain a binary posterior in each round of analysis and hence conclude a binary posterior after the work in previous two steps. Then, the optimal ordering quantity which managers feel less likely to choose by direct instinct across all rounds of analysis is just the one that is closer to the demand. That is, *think one more step*.

7 Conclusion

This paper contributes to the rational inattention theory by grounding the theory from a neurocognitive perspective to understand and obtain the analytical formula of the optimal action strategies. The analytical formula reveals the essence of the information acquisition process, that is *think until nothing to think*. As a result of the information acquisition process, the acquired information for each contingent state just weighs the likelihood that each possible optimal ordering quantity is chosen given the state. However, the immediate result of information acquisition is hard to be effective. The optimal ordering quantity that is closest to the given state is hard to choose with highest likelihood. However, based on our theoretical findings, we find that an effective thinking can be achieved such that the decision maker can refine his information acquisition result to a binary distribution as long as the decision maker can, and *think one more step*, that is the optimal ordering quantity chosen with lower likelihood given the acquired information on state is just the optimal ordering quantity closer to the state.

Based on our theoretical findings, we propose the *empirical approach* to identify the optimal ordering quantities. That is estimating the unconditional optimal ordering strategies and use the estimates with the theoretical formula of the unconditional optimal ordering strategies to identify the optimal ordering quantities. However, a challenge of applying the empirical approach comprehensively is from the technical errors.

Technical errors are prevalent in our daily works and the technical errors contaminate the primitive information for information acquisition. We study once technical errors contaminate the primitive information, how the contamination affects decision makers' information acquisition. By experimenting three technical error distributions, we find that the technical errors will mistaken the information acquisition results if the primitive information were clean.

Therefore, by looking back the empirical approach we have proposed to identify the optimal ordering quantities obtained from clean primitive information, to separate the impact of technical errors from the estimation is the core to guarantee the validity of the estimation result. It is ideal to expect human decision makers possessing an intelligent mind to separate the technical errors in their cognition. To ensure the estimates as reliable as possible, we suggest the instrumental variable method to use in the estimation of the unconditional likelihood of the ordering decisions. How to generally implement the estimation approach will be our future study.

An alternative formulation of rational inattention decision problems can be understood via Friston's free energy principle, which is a theory to study how an organism, e.g. brain, processes information to get rid of disorder status of states. Therefore, the interpretation of the rational inattention decision models can be founded from a neurocognitive perspective, which provides a solid ground for further research of the rational inattention theory. For example, for information acquisition, it is the dopaminergic system in the brain that conducts the process and the free energy principle applies to interpret the activity by the dopaminergic system. Mapping the interpretation into rational inattention theory, some parameters, for example the shadow price λ , can get explained biologically. λ , with our knowledge about brain science by far, can be understood as a measurement of the density of dopamine, which substance is the infrastructure that has been proven to determine the information acquisition activity modelled by free energy principle or rational inattention theory. If our understanding is proven right, then the empirical studies of rational inattention theory will proceed to a new stage. That is, the questions on human behavior studied in economics can be answered from neuropsychology or neurocognition science, and vice versa.

8 Appendix

A Derivation of the Optimal Ordering Strategy

Proof: The derivation of the optimal ordering strategy is based on the model specification DP-1. The derivation based on model specification DP-2 can follow the same proof demonstrated in the following.

Applying Matějka and McKay (2015)'s manner of deriving the optimal conditional ordering probabilities, we derive the optimal ordering strategy by obtaining the first order condition of the manager's profit maximization problem. From (5), (7) and (8), the Lagrangian function of the optimization problem is

$$\begin{aligned} L(f(\alpha^*|\theta)) = & \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi(f(\theta|\alpha^*))f(\alpha^*|\theta)g(\theta)d\theta d\alpha^* \\ & - \lambda \left(- \int_{-\infty}^{+\infty} f(\alpha^*) \log f(\alpha^*) d\alpha^* \right. \\ & + \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(\alpha^*|\theta) \log f(\alpha^*|\theta) d\alpha^* \right) g(\theta) d\theta \\ & + \int_{-\infty}^{+\infty} \lambda_{\alpha^*}(\theta) f(\alpha^*|\theta) g(\theta) d\theta \\ & \left. - \int_{-\infty}^{+\infty} \lambda(\theta) \left(\int_{-\infty}^{+\infty} f(\alpha^*|\theta) d\alpha^* - 1 \right) g(\theta) d\theta \right) \end{aligned}$$

where λ , $\lambda_{\alpha^*}(\theta)$ and $\lambda(\theta)$ are Lagrangian multipliers. Because $f(\alpha^*|\theta)$ is non-negative, hence $\lambda_{\alpha^*}(\theta) = 0$. Because the probability density function satisfies $\int_{-\infty}^{+\infty} f(\alpha^*|\theta) d\alpha^* = 1$, hence $\lambda(\theta) > 0$. The first order condition of this optimization problem is given by

$$p \min(\theta, \alpha^*) - c\alpha^* - \lambda(\theta) + \lambda(\log f(\alpha^*) - \log f(\alpha^*|\theta)) = 0$$

Arranging this first order condition, we obtain

$$f(\alpha^*|\theta) = f(\alpha^*) \exp\left(\frac{p \min(\theta, \alpha^*) - c\alpha^* - \lambda(\theta)}{\lambda}\right) \quad (17)$$

Put equation (17) back into (8), and we get

$$\exp\left(\frac{\lambda(\theta)}{\lambda}\right) = \int_{-\infty}^{+\infty} f(\alpha^*) \exp\left(\frac{p \min(\theta, \alpha^*) - c\alpha^*}{\lambda}\right) d\alpha^* \quad (18)$$

Combining equations (17) and (18), we get

$$f(\alpha^*|\theta) = \frac{f(\alpha^*) \exp\left(\frac{p \min(\theta, \alpha^*) - c\alpha^*}{\lambda}\right)}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp\left(\frac{p \min(\theta, \alpha^*) - c\alpha^*}{\lambda}\right) d\alpha^*}$$

Note that the above equation is still established for $f(\alpha^*) = 0$. Therefore the relationship $f(\alpha^*) = \int_{-\infty}^{+\infty} f(\alpha^*|\theta)g(\theta)d\theta$ is maintained.

B Proof of Proposition 1

Proof: As stated in the main context of the paper, the proofs of all results in the following are based on the model specification DP-1.

Caplin, Dean and Leahy (2019) studies rational inattention discrete choice models. They find that after information acquisition, an optimal consideration set emerges. Our proof of Proposition 1 follows the proof of their Proposition 1, even if we study a continuous-state continuous-action model. Jung, King, Matějka and Sims (2019) studies rational inattention decision models that follows Boreil measures and the decision variables are continuous. They find that the resulted posterior consists of a countable set of points with no limit points. Our Proposition 1 is consistent with the result.

The first order condition of the rational inattention decision problem can be written by $\pi(\theta, \alpha^*) + \lambda_{\alpha^*}(\theta) - \lambda(\theta) + \lambda(\log f(\alpha^*) - \log f(\alpha^*|\theta)) = 0$. Therefore,

$$\begin{aligned} f(\alpha^*|\theta) &= f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*) + \lambda_{\alpha^*}(\theta) - \lambda(\theta)}{\lambda}\right) \\ &= f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*) - \lambda(\theta)}{\lambda}\right) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right) \end{aligned}$$

Define set D by $D \equiv \{\alpha^*|f(\alpha^*) \geq 0\}$. Therefore, for $\alpha^* \notin D$, $f(\alpha^*) = 0$, and hence $f(\alpha^*|\theta) = 0$ due to its non-negativity. For $\alpha^* \in D$, $\lambda_{\alpha^*}(\theta) = 0$ according to complementarity slackness. Because $\int_{-\infty}^{+\infty} f(\alpha^*|\theta)d\alpha^* = 1$, therefore $\int_{-\infty}^{+\infty} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right) d\alpha^* = \exp\left(\frac{\lambda(\theta)}{\lambda}\right)$. Therefore,

$$f(\alpha^*|\theta) = \frac{f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right)}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right) d\alpha^*}$$

However,

$$\begin{aligned} &\int_{-\infty}^{+\infty} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right) d\alpha^* \\ &= \int_{\alpha^* \in D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right) d\alpha^* \\ &+ \int_{\alpha^* \notin D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right) d\alpha^* \\ &= \int_{\alpha^* \in D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^* \end{aligned}$$

Therefore, for $\alpha^* \in D$,

$$f(\alpha^*|\theta) = \frac{f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right)}{\int_{\alpha^* \in D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^*}$$

and

$$\int_{-\infty}^{+\infty} \frac{\exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right)g(\theta)}{\int_{\alpha^* \in D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^*} d\theta = 1$$

For $\alpha^* \notin D$,

$$f(\alpha^*|\theta) = \frac{f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right)}{\int_{\alpha^* \in D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^*}$$

Therefore,

$$\int_{-\infty}^{+\infty} f(\alpha^*|\theta)g(\theta)d\theta = \int_{-\infty}^{+\infty} \frac{f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right)g(\theta) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right)}{\int_{\alpha^* \in D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^*} d\theta$$

and hence for $\alpha^* \notin D$,

$$\int_{-\infty}^{+\infty} \frac{\exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right)}{\int_{\alpha^* \in D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^*} g(\theta) \exp\left(\frac{\lambda_{\alpha^*}(\theta)}{\lambda}\right) d\theta = 1$$

Therefore, due to the non-negativity of $\lambda(\theta)$,

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} \frac{\exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right)}{\int_{\alpha^* \in D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^*} \exp\left(\frac{\lambda(\theta)}{\lambda}\right)g(\theta)d\theta \\ &\geq \int_{-\infty}^{+\infty} \frac{\exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right)g(\theta)}{\int_{\alpha^* \in D} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^*} d\theta \\ &= \int_{-\infty}^{+\infty} \frac{\exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right)g(\theta)}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^*} d\theta \\ &= F(\alpha^*) \end{aligned}$$

Therefore, for $\alpha^* \notin D$, these α^* are not chosen for any value of θ and $F(\alpha^*) \leq 1$. Otherwise, for $\alpha^* \in D$, which are chosen for any value of θ , $F(\alpha^*) = 1$. Therefore, Proposition 1 is proven.

C Proof of Corollary 1

Proof: $F(\alpha^*)$ is analytic. $F'(\alpha^*) = 0$ just happens for scattered points on the real line. Therefore, if $F(\alpha^*) = 1$ has solutions, then a discrete $f(\alpha^*)$ will be obtained, and $f(\alpha^*|\theta)$ is discrete as well. Therefore, even though the prior is Lesbegue, and the decision variable is continuous, the information acquisition activity of our model finally leads to a discrete posterior. Therefore, Corollary 1 is proven.

D Proof of Proposition 2

Proof: Recall equation (10):

$$\int_{-\infty}^{+\infty} \frac{\exp(\frac{p \min(\theta, \alpha^*) - c \alpha^*}{\lambda})}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{p \min(\theta, \alpha^*) - c \alpha^*}{\lambda}) d\alpha^*} g(\theta) d\theta = 1.$$
It should be held for all value of α^* . Define $F(\alpha^*) = \int_{-\infty}^{+\infty} \frac{\exp(\frac{p \min(\theta, \alpha^*) - c \alpha^*}{\lambda})}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{p \min(\theta, \alpha^*) - c \alpha^*}{\lambda}) d\alpha^*} g(\theta) d\theta$, and $A(\theta) = \int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{\pi(\theta, \alpha^*)}{\lambda}) d\alpha^* = \int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{p \min(\theta, \alpha^*) - c \alpha^*}{\lambda}) d\alpha^*$. Therefore, $\frac{\partial F(\alpha^*)}{\partial \alpha^*} = 0$, $\frac{\partial^2 F(\alpha^*)}{\partial \alpha^{*2}} = 0$, ..., $\frac{\partial^n F(\alpha^*)}{\partial \alpha^{*n}} = 0$ and so on for higher order derivatives.

Further, $F(\alpha^*) = \int_{\alpha^*}^{+\infty} \frac{e^{\frac{(p-c)\alpha^*}{\lambda}}}{A(\theta)} g(\theta) d\theta + \int_{-\infty}^{\alpha^*} \frac{e^{\frac{p\theta - c\alpha^*}{\lambda}}}{A(\theta)} g(\theta) d\theta = 1$. Therefore,

$$\begin{aligned} \frac{\partial F(\alpha^*)}{\partial \alpha^*} &= \frac{p-c}{\lambda} e^{\frac{(p-c)\alpha^*}{\lambda}} \int_{\alpha^*}^{+\infty} \frac{g(\theta)}{A(\theta)} d\theta \\ &\quad - \frac{c}{\lambda} e^{-\frac{c\alpha^*}{\lambda}} \int_{-\infty}^{\alpha^*} e^{\frac{p\theta}{\lambda}} \frac{g(\theta)}{A(\theta)} d\theta \\ &= \psi(\alpha^*) \end{aligned} \quad (19)$$

and

$$\begin{aligned} \frac{\partial^2 F(\alpha^*)}{\partial \alpha^{*2}} &= \left(\frac{p-c}{\lambda}\right)^2 e^{\frac{(p-c)\alpha^*}{\lambda}} \int_{\alpha^*}^{+\infty} \frac{g(\theta)}{A(\theta)} d\theta \\ &\quad + \frac{c^2}{\lambda^2} e^{-\frac{c\alpha^*}{\lambda}} \int_{-\infty}^{\alpha^*} e^{\frac{p\theta}{\lambda}} \frac{g(\theta)}{A(\theta)} d\theta \\ &\quad - \frac{p}{\lambda} e^{\frac{(p-c)\alpha^*}{\lambda}} \frac{g(\alpha^*)}{A(\alpha^*)} \\ &= \eta(\alpha^*) \end{aligned} \quad (20)$$

Timing $\frac{c}{\lambda}$ on both sides of equation (19), equation (19) becomes

$$\frac{c^2}{\lambda^2} e^{-\frac{c\alpha^*}{\lambda}} \int_{-\infty}^{\alpha^*} e^{\frac{p\theta}{\lambda}} \frac{g(\theta)}{A(\theta)} d\theta = \frac{c(p-c)}{\lambda^2} e^{\frac{(p-c)\alpha^*}{\lambda}} \int_{\alpha^*}^{+\infty} \frac{g(\theta)}{A(\theta)} d\theta + \varepsilon(\alpha^*) \quad (21)$$

where $\varepsilon(\alpha^*) = \psi(\alpha^*) \frac{c}{\lambda}$. Combining equation (20) and (21), therefore we obtain that

$$\frac{p(p-c)}{\lambda^2} \int_{\alpha^*}^{+\infty} \frac{g(\theta)}{A(\theta)} d\theta = \frac{p}{\lambda} \frac{g(\alpha^*)}{A(\alpha^*)} + (\eta(\alpha^*) - \varepsilon(\alpha^*)) \exp\left(-\frac{(p-c)\alpha^*}{\lambda}\right) \quad (22)$$

Therefore, according to equation (22), the function $\frac{p(p-c)}{\lambda^2} \int_{\alpha^*}^{+\infty} \frac{g(\theta)}{A(\theta)} d\theta - \frac{p}{\lambda} \frac{g(\alpha^*)}{A(\alpha^*)} - (\eta(\alpha^*) - \varepsilon(\alpha^*)) \exp\left(-\frac{(p-c)\alpha^*}{\lambda}\right)$ always equation to 0 for any value of α^* . Therefore, the derivative of $\frac{p(p-c)}{\lambda^2} \int_{\alpha^*}^{+\infty} \frac{g(\theta)}{A(\theta)} d\theta - \frac{p}{\lambda} \frac{g(\alpha^*)}{A(\alpha^*)} - (\eta(\alpha^*) - \varepsilon(\alpha^*)) \exp\left(-\frac{(p-c)\alpha^*}{\lambda}\right)$ with respect to α^* is always equal to 0. Therefore, by taking derivative with respect to α^* on both sides of equation (22) and substitute the variable α^* by θ , it is obtained that

$$\begin{aligned} \left(\frac{g(\theta)}{A(\theta)}\right)' + \frac{p-c}{\lambda} \frac{g(\theta)}{A(\theta)} &= ((\eta(\theta) - \varepsilon(\theta)) \frac{p-c}{p} - (\eta'(\theta) - \varepsilon'(\theta)) \frac{\lambda}{p}) \exp\left(-\frac{p-c}{\lambda} \theta\right) \end{aligned} \quad (23)$$

Solving ODE (23), we obtain that

$$\begin{aligned} \frac{g(\theta)}{A(\theta)} &= \xi \exp\left(-\frac{p-c}{\lambda} \theta\right) \\ &\quad + ((\eta(\theta) - \varepsilon(\theta)) \frac{p-c}{p} - (\eta'(\theta) - \varepsilon'(\theta)) \frac{\lambda}{p}) \theta \exp\left(-\frac{p-c}{\lambda} \theta\right) \end{aligned}$$

where ξ is a constant of which value is to determine requiring extra conditions, e.g. its value is known at a particular θ_0 such that $\xi = \frac{g(\theta_0)}{A(\theta_0)} \exp\left(\frac{p-c}{\lambda} \theta_0\right) - ((\eta - \varepsilon) \frac{p-c}{p} - (\eta' - \varepsilon') \frac{\lambda}{p}) \theta_0$. Therefore, according to the definition of $A(\theta)$

$$\begin{aligned} &\int_{-\infty}^{+\infty} f(\alpha^*) \exp\left(\frac{\pi(\theta, \alpha^*)}{\lambda}\right) d\alpha^* \\ &= \frac{g(\theta)}{\xi \exp\left(-\frac{p-c}{\lambda} \theta\right) + ((\eta(\theta) - \varepsilon(\theta)) \frac{p-c}{p} - (\eta'(\theta) - \varepsilon'(\theta)) \frac{\lambda}{p}) \theta \exp\left(-\frac{p-c}{\lambda} \theta\right)} \end{aligned} \quad (24)$$

Suppose the posterior distribution is a binary distribution such that $D \equiv \{\alpha_1^*, \alpha_2^*\}$ where $\alpha_1^* < \alpha_2^*$. Therefore, in the binary posterior distribution scenario, equation (24) can be written by

$$f(\alpha_1^*) \exp\left(\frac{\pi(\theta, \alpha_1^*)}{\lambda}\right) + (1 - f(\alpha_1^*)) \exp\left(\frac{\pi(\theta, \alpha_2^*)}{\lambda}\right) = T(\theta)$$

where

$$T(\theta) \equiv \frac{g(\theta)}{\xi \exp\left(-\frac{p-c}{\lambda} \theta\right) + ((\eta(\theta) - \varepsilon(\theta)) \frac{p-c}{p} - (\eta'(\theta) - \varepsilon'(\theta)) \frac{\lambda}{p}) \theta \exp\left(-\frac{p-c}{\lambda} \theta\right)}$$

Therefore,

$$f(\alpha_1^*) = \frac{T(\theta) - \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda}) - \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})} \quad (25)$$

Because $\frac{\partial f(\alpha_1^*)}{\partial \theta} = 0$, therefore taking derivative with respect to θ of both sides of equation (25), it is obtained that

$$f(\alpha_1^*) = \frac{T'(\theta) - \frac{\partial \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta}}{\frac{\partial \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda})}{\partial \theta} - \frac{\partial \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta}}$$

Suppose for $k - 1 > 1$, $f(\alpha_1^*) =$

$$\frac{T^{(k-1)}(\theta) - \frac{\partial^{k-1} \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^{k-1}}}{\frac{\partial^{k-1} \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda})}{\partial \theta^{k-1}} - \frac{\partial^{k-1} \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^{k-1}}}, \text{ where } T^{k-1}(\theta) \text{ is the}$$

$k - 1$ -th order derivative of $T(\theta)$. Because $\frac{\partial f(\alpha_1^*)}{\partial \theta} = 0$, therefore by taking derivative with respect to θ for the $f(\alpha_1^*)$ where $T^{(k-1)}(\theta)$ is a component in it, it is obtained that

$$f(\alpha_1^*) = \frac{T^{(k)}(\theta) - \frac{\partial^k \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^k}}{\frac{\partial^k \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda})}{\partial \theta^k} - \frac{\partial^k \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^k}}. \text{ Therefore, by deduc-}$$

tion, it is obtained that

$$f(\alpha_1^*) = \frac{T^{(n)}(\theta) - \frac{\partial^n \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^n}}{\frac{\partial^n \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda})}{\partial \theta^n} - \frac{\partial^n \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^n}}$$

where $T^{(n)}(\theta)$ is the n -th order derivative of $T(\theta)$ with respect to θ and $n = 1, 2, 3, \dots$. Because $\frac{\partial f(\alpha_1^*)}{\partial \theta} = 0$ is always held, therefore in our model, the value of n equals the number which is the minimum rounds of differentiation to make all components in $f(\alpha_1^*)$, which are $\frac{\partial^n \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda})}{\partial \theta^n}$, $\frac{\partial^n \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^n}$ and $T^{(n)}(\theta) = 0$, independent of θ , i.e. $\frac{\partial^{n+1} \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda})}{\partial \theta^{n+1}}$, $\frac{\partial^{n+1} \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})}{\partial \theta^{n+1}} = 0$ and $T^{(n+1)}(\theta) = 0$. Therefore, Proposition 2 is proven.

E Proof of Proposition 3

Proof: Recall equation (10), which is

$$\int_{-\infty}^{+\infty} \frac{\exp(\frac{p \min(\theta, \alpha^*) - c \alpha^*}{\lambda}) g(\theta)}{\int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{p \min(\theta, \alpha^*) - c \alpha^*}{\lambda}) d\alpha^*} d\theta = 1. \text{ If the posterior distribution is a binary distribution such that } D \text{ contains two elements } \alpha_1^* \text{ and } \alpha_2^* \text{ which are chosen with } f(\alpha_1^*) \text{ and } f(\alpha_2^*) \text{ respectively, where } f(\alpha_1^*) + f(\alpha_2^*) = 1, \text{ then the denominator of the integrand becomes } \int_{-\infty}^{+\infty} f(\alpha^*) \exp(\frac{p \min(\theta, \alpha^*) - c \alpha^*}{\lambda}) d\alpha^* = f(\alpha_1^*) \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda}) + f(\alpha_2^*) \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda}). \text{ Hence, in this situation, equation (10) becomes } \int_{-\infty}^{+\infty} \frac{\exp(\frac{\pi(\theta, \alpha^*)}{\lambda}) g(\theta)}{f(\alpha_1^*) \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda}) + f(\alpha_2^*) \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})} d\theta = 1.$$

According to Proposition 2, $f(\alpha_1^*)$ is a function with respect to α_1^* and α_2^* , so is $f(\alpha_2^*)$. Therefore, the equation

$$\int_{-\infty}^{+\infty} \frac{\exp(\frac{\pi(\theta, \alpha^*)}{\lambda}) g(\theta)}{f(\alpha_1^*) \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda}) + f(\alpha_2^*) \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})} d\theta = 1 \text{ has two unknown variables } \alpha_1^* \text{ and } \alpha_2^*. \text{ Neither } \alpha_1^* \text{ nor } \alpha_2^* \text{ can be obtained by}$$

just solving a single equation. One more condition is required.

Recall equation (15). For a binary posterior where $\alpha_1^* \alpha_2^* \in D$, equation (15) is written by $f(\alpha^{*'}) = f_\varepsilon(m^{-1}(\alpha^{*'}) - m^{-1}(\alpha_1^*))f(\alpha_1^*) + f_\varepsilon(m^{-1}(\alpha^{*'}) - m^{-1}(\alpha_2^*))f(\alpha_2^*) = f(\alpha_1^*)(f_\varepsilon(m^{-1}(\alpha^{*'}) - m^{-1}(\alpha_1^*)) - f_\varepsilon(m^{-1}(\alpha^{*'}) - m^{-1}(\alpha_2^*))) + f_\varepsilon(m^{-1}(\alpha^{*'}) - m^{-1}(\alpha_2^*))$. If the technical error distribution $f_\varepsilon(\varepsilon)$ satisfies that $f_\varepsilon(\varepsilon_1 + \varepsilon_2) = f_\varepsilon(\varepsilon_1)f_\varepsilon(\varepsilon_2)$, then the equation of $f(\alpha^{*'})$ can be further arranged to $f(\alpha^{*'}) = (f(\alpha_1^*)f_\varepsilon(-m^{-1}(\alpha_1^*)) - f(\alpha_1^*)f_\varepsilon(-m^{-1}(\alpha_2^*)) + f_\varepsilon(-m^{-1}(\alpha_2^*)))f_\varepsilon(m^{-1}(\alpha^{*'}))$. Define $\gamma \equiv f(\alpha_1^*)f_\varepsilon(-m^{-1}(\alpha_1^*)) - f(\alpha_1^*)f_\varepsilon(-m^{-1}(\alpha_2^*)) + f_\varepsilon(-m^{-1}(\alpha_2^*))$. Therefore, the previous equation can be written by $f(\alpha^{*'}) = \gamma f_\varepsilon(m^{-1}(\alpha^{*'}))$. The data of $f(\alpha^{*'})$ and $\alpha^{*'}$ are feasible to collect. The one-to-one correspondence relationship $m(\cdot)$ can be obtained theoretically according to the procedures we present in the three examples in section 5. $m(\cdot)$ is invertible due to its one-to-one correspondence feature. Hence, the data of $\alpha^{*'}$ can be transformed to the data of $f_\varepsilon(m^{-1}(\alpha^{*'}))$. Therefore, by estimating the linear relationship between $f(\alpha^{*'})$ and $f_\varepsilon(m^{-1}(\alpha^{*'}))$, i.e. estimating $f(\alpha^{*'}) = \gamma f_\varepsilon(m^{-1}(\alpha^{*'})) + v$, where v is the estimation error, γ can be obtained statistically.

By obtaining the value of γ , we rearrange the definition equation of γ and obtain that $f(\alpha_1^*) = \frac{\gamma - f_\varepsilon(-m^{-1}(\alpha_2^*))}{f_\varepsilon(-m^{-1}(\alpha_1^*)) - f_\varepsilon(-m^{-1}(\alpha_2^*))}$. The equation together with equation (13) construct second equation of α_1^* and α_2^* in equation (14). Therefore, two equations are obtained for α_1^* and α_2^* : $\int_{-\infty}^{+\infty} \frac{\exp(\frac{\pi(\theta, \alpha^*)}{\lambda}) g(\theta)}{f(\alpha_1^*) \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda}) + f(\alpha_2^*) \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})} d\theta = 1$ and $f(\alpha_1^*) = \frac{\gamma - f_\varepsilon(-m^{-1}(\alpha_2^*))}{f_\varepsilon(-m^{-1}(\alpha_1^*)) - f_\varepsilon(-m^{-1}(\alpha_2^*))}$. If the equation group composed by the two equations have solutions, then it can be said that $\alpha_1^*, \alpha_2^* \in D$ exist and α_1^* and α_2^* can be obtained by solving the equation group. Otherwise, α_1^* and α_2^* do not exist and therefore a binary posterior distribution does not exist. Therefore, Proposition 3 is proven.

F Proof of Corollary 2

Proof: Define $F_1(\alpha_1^*, \alpha_2^*) = \alpha_1^* \int_{-\infty}^{+\infty} \frac{\exp(\frac{\pi(\theta, \alpha^*)}{\lambda}) g(\theta)}{f(\alpha_1^*) \exp(\frac{\pi(\theta, \alpha_1^*)}{\lambda}) + f(\alpha_2^*) \exp(\frac{\pi(\theta, \alpha_2^*)}{\lambda})} d\theta$ and $F_2(\alpha_1^*, \alpha_2^*) = \alpha_2^* - f(\alpha_1^*) + \frac{\gamma - f_\varepsilon(-m^{-1}(\alpha_2^*))}{f_\varepsilon(-m^{-1}(\alpha_1^*)) - f_\varepsilon(-m^{-1}(\alpha_2^*))}$. If the prior distribution $g(\theta) \geq 0 \in [a, b] \subset (0, +\infty)$ and $g(\theta) = 0$ otherwise, then by solving equation group (14), the obtained $\alpha_1^*, \alpha_2^* \in [a, b]$ as well. $F_1(\alpha_1^*, \alpha_2^*)$ and $F_2(\alpha_1^*, \alpha_2^*)$ are continuous functions with respect to $(\alpha_1^*, \alpha_2^*) \in [a, b] \times [a, b]$. Therefore, according to Brouwer's fixed point theorem, for the following equation group, which is equivalent to equation group (14),

$$\begin{cases} F_1(\alpha_1^*, \alpha_2^*) = \alpha_1^* \\ F_2(\alpha_1^*, \alpha_2^*) = \alpha_2^* \end{cases}$$

a fixed point of (α_1^*, α_2^*) exists. Therefore, if the prior distribution $g(\theta)$ satisfies that $g(\theta) \geq 0 \in [a, b] \subset (0, +\infty)$ and $g(\theta) = 0$ otherwise, then equation group (14) has at least one solution of $(\alpha_1^*, \alpha_2^*) \in [a, b] \times [a, b]$. Therefore, Corollary 2 is proven.

G Proof of Corollary 3

Proof: Consider the following four situations.

I. Suppose $0 < \alpha_2^* - \theta < \theta - \alpha_1^*$. Hence, $\frac{\alpha_1^* + \alpha_2^*}{2} < \theta$. If $f(\alpha_2^*|\theta) > f(\alpha_1^*|\theta)$, therefore by equation (9), it is obtained that $\theta > \frac{\lambda}{p} \ln \frac{f(\alpha_1^*)}{f(\alpha_2^*)} \exp(\frac{p\alpha_1^* - c(\alpha_1^* - \alpha_2^*)}{\lambda})$. Therefore, with the prerequisite condition that $\frac{\alpha_1^* + \alpha_2^*}{2} < \theta$, the result $\theta > \frac{\lambda}{p} \ln \frac{f(\alpha_1^*)}{f(\alpha_2^*)} \exp(\frac{p\alpha_1^* - c(\alpha_1^* - \alpha_2^*)}{\lambda})$ is held indicates that $\frac{\alpha_1^* + \alpha_2^*}{2} \leq \frac{\lambda}{p} \ln \frac{f(\alpha_1^*)}{f(\alpha_2^*)} \exp(\frac{p\alpha_1^* - c(\alpha_1^* - \alpha_2^*)}{\lambda})$.

II. Suppose $\alpha_1^* < \alpha_2^* < \theta$. This situation also implies $\frac{\alpha_1^* + \alpha_2^*}{2} < \theta$. If $f(\alpha_2^*|\theta) > f(\alpha_1^*|\theta)$, therefore by equation (9), it is obtained that $\frac{f(\alpha_2^*)}{f(\alpha_1^*)} > \frac{\exp(\frac{(p-c)\alpha_1^*}{\lambda})}{\exp(\frac{(p-c)\alpha_2^*}{\lambda})}$.

III. Suppose $\alpha_2^* - \theta > \theta - \alpha_1^* > 0$. Hence, $\frac{\alpha_1^* + \alpha_2^*}{2} > \theta$. If $f(\alpha_2^*|\theta) < f(\alpha_1^*|\theta)$, therefore by equation (9), it is obtained that $\theta < \frac{\lambda}{p} \ln \frac{f(\alpha_1^*)}{f(\alpha_2^*)} \exp(\frac{p\alpha_1^* - c(\alpha_1^* - \alpha_2^*)}{\lambda})$. Therefore, with the prerequisite condition that $\frac{\alpha_1^* + \alpha_2^*}{2} > \theta$, the result $\theta < \frac{\lambda}{p} \ln \frac{f(\alpha_1^*)}{f(\alpha_2^*)} \exp(\frac{p\alpha_1^* - c(\alpha_1^* - \alpha_2^*)}{\lambda})$ is held indicates that $\frac{\alpha_1^* + \alpha_2^*}{2} \geq \frac{\lambda}{p} \ln \frac{f(\alpha_1^*)}{f(\alpha_2^*)} \exp(\frac{p\alpha_1^* - c(\alpha_1^* - \alpha_2^*)}{\lambda})$.

IV. Suppose $\theta < \alpha_1^* < \alpha_2^*$. This situation also implies $\frac{\alpha_1^* + \alpha_2^*}{2} > \theta$. If $f(\alpha_2^*|\theta) < f(\alpha_1^*|\theta)$, therefore by equation (9), it is obtained that $\frac{f(\alpha_2^*)}{f(\alpha_1^*)} < \frac{\exp(\frac{c\alpha_2^*}{\lambda})}{\exp(\frac{c\alpha_1^*}{\lambda})}$.

Therefore, if the result that $f(\alpha_2^*|\theta) > f(\alpha_1^*|\theta)$ if $|\alpha_2^* - \theta| < |\alpha_1^* - \theta|$ and $f(\alpha_2^*|\theta) < f(\alpha_1^*|\theta)$ if $|\alpha_2^* - \theta| > |\alpha_1^* - \theta|$ always happens, the parameters should always simultaneously satisfy the conditions listed in above four scenarios, all of which together is the necessary and sufficient condition of the result. These conditions can be summarized by

$$\begin{cases} \frac{\alpha_1^* + \alpha_2^*}{2} = \frac{\lambda}{p} \ln \frac{f(\alpha_1^*)}{f(\alpha_2^*)} \exp(\frac{p\alpha_1^* - c(\alpha_1^* - \alpha_2^*)}{\lambda}) \\ \frac{\exp(\frac{c\alpha_2^*}{\lambda})}{\exp(\frac{c\alpha_1^*}{\lambda})} > \frac{f(\alpha_2^*)}{f(\alpha_1^*)} > \frac{\exp(\frac{(p-c)\alpha_1^*}{\lambda})}{\exp(\frac{(p-c)\alpha_2^*}{\lambda})} \end{cases}$$

However, the equation that $\frac{\alpha_1^* + \alpha_2^*}{2} = \frac{\lambda}{p} \ln \frac{f(\alpha_1^*)}{f(\alpha_2^*)} \exp(\frac{p\alpha_1^* - c(\alpha_1^* - \alpha_2^*)}{\lambda})$ restricts the available parameter specifications to satisfy the above necessary and sufficient conditions. Almost surely, it is hard to find parameter specifications to satisfy the necessary and sufficient condition. Therefore, Corollary 3 is proven.

H Proof of Corollary 4

Proof: Consider a general posterior such that $D = \{a_1^*, a_2^*, \dots, a_n^* | a_1^* < a_2^* < \dots < a_n^*\}$ where $n \geq 3$. Our proof begins from focusing on the α^* that is most close to θ denoted by $\alpha_{(1)}^*$ and the second most close to θ denoted by $\alpha_{(2)}^*$. According to Corollary 3, it is known that almost surely it cannot be held that $f(\alpha_{(2)}^*|\theta) > f(\alpha_{(1)}^*|\theta)$ if $|\alpha_{(2)}^* - \theta| < |\alpha_{(1)}^* - \theta|$ and $f(\alpha_{(2)}^*|\theta) < f(\alpha_{(1)}^*|\theta)$ if $|\alpha_{(2)}^* - \theta| > |\alpha_{(1)}^* - \theta|$. Therefore, the α^* that is most close to θ is hard to choose with higher likelihood than the α^* that is second most close to θ . Therefore, it is hard for

the α^* most close to θ to choose with highest likelihood.

I Proof of Proposition 4

Proof: According to Bayes theorem, $f(\alpha^{s'}) = \sum_{\alpha^* \in D} f(\alpha^{s'}|\alpha^*)f(\alpha^*) = \sum_{\alpha^* \in D} f(\alpha^* + \varepsilon|\alpha^*)f(\alpha^*)$. Given α^* , it is ε that determines the value of $\alpha^{s'}$ according to $s' = s + \varepsilon$. Because $\alpha^* = m(s)$ and the function $m(\cdot)$ is invertible, therefore an ε corresponding to $\alpha^{s'}$ can be expressed by $\varepsilon = m^{-1}(\alpha^{s'}) - m^{-1}(\alpha^*)$ given α^* . Therefore, $f(\alpha^{s'}|\alpha^*) = f_\varepsilon(m^{-1}(\alpha^{s'}) - m^{-1}(\alpha^*))$. Therefore, $f(\alpha^{s'}) = \sum_{\alpha^* \in D} f_\varepsilon(m^{-1}(\alpha^{s'}) - m^{-1}(\alpha^*))f(\alpha^*)$.

Unlike the situation where the information acquisition activity is not affected by the technical error and hence any form of $m(\cdot)$ makes $\int_{-\infty}^{+\infty} f(\alpha^*)d\alpha^* = 1$, the one-to-one correspondence $m(\cdot)$ does not make $\int_{-\infty}^{+\infty} f(\alpha^{s'})d\alpha^{s'} = 1$ by default. Therefore, due to the technical error, the form of $m(\cdot)$ is restricted to satisfy $\int_{-\infty}^{+\infty} f(\alpha^{s'})d\alpha^{s'} = 1$. Proposition 4 is proven.

J Proof of Proposition 5

Proof: Suppose $f_\varepsilon(\varepsilon)$ follows an exponential distribution. Therefore, $f_\varepsilon(\varepsilon)$ is given by

$$f_\varepsilon(\varepsilon) = \begin{cases} \mu \exp(-\mu\varepsilon) & \varepsilon \geq 0 \\ 0 & \varepsilon < 0 \end{cases}$$

Define $f_e^m(\alpha^{s'}; \alpha_1^*) = \mu \exp(-\mu(m^{-1}(\alpha^{s'}) - m^{-1}(\alpha_1^*)))f(\alpha_1^*)$ and $f_e^m(\alpha^{s'}; \alpha_2^*) = \mu \exp(-\mu(m^{-1}(\alpha^{s'}) - m^{-1}(\alpha_2^*)))f(\alpha_2^*)$. Given $\alpha^* \in \{\alpha_1^*, \alpha_2^*\}$, $m^{-1}(\alpha^{s'}) - m^{-1}(\alpha^*) \geq 0$ divides the range of $\alpha^{s'}$ to differentiate between the situations where $f_\varepsilon(m^{-1}(\alpha^{s'}) - m^{-1}(\alpha^*)) \geq 0$ and where $f_\varepsilon(m^{-1}(\alpha^{s'}) - m^{-1}(\alpha^*)) = 0$. Therefore, according to $f(\alpha^{s'}) = \sum_{\alpha^* \in D} f_\varepsilon(m^{-1}(\alpha^{s'}) - m^{-1}(\alpha^*))f(\alpha^*)$, it is obtained that if $m'(\cdot) > 0$,

$$f(\alpha^{s'}) = \begin{cases} 0 & \alpha^{s'} < \alpha_1^* \\ f_e^m(\alpha^{s'}; \alpha_1^*) & \alpha_1^* \leq \alpha^{s'} < \alpha_2^* \\ f_e^m(\alpha^{s'}; \alpha_1^*) + f_e^m(\alpha^{s'}; \alpha_2^*) & \alpha^{s'} \geq \alpha_2^* \end{cases}$$

If $m'(\cdot) < 0$,

$$f(\alpha^{s'}) = \begin{cases} f_e^m(\alpha^{s'}; \alpha_1^*) + f_e^m(\alpha^{s'}; \alpha_2^*) & \alpha^{s'} < \alpha_1^* \\ f_e^m(\alpha^{s'}; \alpha_2^*) & \alpha_1^* \leq \alpha^{s'} < \alpha_2^* \\ 0 & \alpha^{s'} \geq \alpha_2^* \end{cases}$$

where μ takes value such that $\int_{-\infty}^{+\infty} f(\alpha^{s'})d\alpha^{s'} = 1$. Define $f_e(\alpha^{s'}; \alpha_1^*) = \mu \exp(-\mu k(\alpha^{s'} - \alpha_1^*))f(\alpha_1^*)$ and $f_e(\alpha^{s'}; \alpha_2^*) = \mu \exp(-\mu k(\alpha^{s'} - \alpha_2^*))f(\alpha_2^*)$. Because we specify

that $\alpha^* = \frac{1}{k}s$, therefore if $k > 0$,

$$f(\alpha^{*'}) = \begin{cases} 0 & \alpha^{*'} < \alpha_1^* \\ f_e(\alpha^{*'}; \alpha_1^*) & \alpha_1^* \leq \alpha^{*'} < \alpha_2^* \\ f_e(\alpha^{*'}; \alpha_1^*) + f_e(\alpha^{*'}; \alpha_2^*) & \alpha^{*'} \geq \alpha_2^* \end{cases}$$

If $k < 0$,

$$f(\alpha^{*'}) = \begin{cases} 0 & \alpha^{*'} < 0 \\ f_e(\alpha^{*'}; \alpha_1^*) + f_e(\alpha^{*'}; \alpha_2^*) & 0 \leq \alpha^{*'} < \alpha_1^* \\ f_e(\alpha^{*'}; \alpha_2^*) & \alpha_1^* \leq \alpha^{*'} < \alpha_2^* \\ 0 & \alpha^{*'} \geq \alpha_2^* \end{cases}$$

where k takes value such that $\int_{-\infty}^{+\infty} f(\alpha^{*'})d\alpha^{*'} = 1$.

Define $F_1(k; \mu) = \int_{\alpha_1^*}^{+\infty} \mu \exp(-\mu k(\alpha^{*'} - \alpha_1^*))f(\alpha_1^*)d\alpha^{*'} + \int_{\alpha_2^*}^{+\infty} \mu \exp(-\mu k(\alpha^{*'} - \alpha_2^*))f(\alpha_2^*)d\alpha^{*'}$ and $F_2(k; \mu) = \int_0^{\alpha_1^*} \mu \exp(-\mu k(\alpha^{*'} - \alpha_1^*))f(\alpha_1^*)d\alpha^{*'} + \int_0^{\alpha_2^*} \mu \exp(-\mu k(\alpha^{*'} - \alpha_2^*))f(\alpha_2^*)d\alpha^{*'}$. For $k > 0$, given $\mu > 0$, $F_1(0; \mu) = +\infty$ and $F_1(+\infty; \mu) = 0$. Because $\frac{\partial F_1(k; \mu)}{\partial k} < 0$, therefore there exists a unique $k = \hat{k} > 0$ such that $F(\hat{k}; \mu) = \int_{-\infty}^{+\infty} f(\alpha^{*'})d\alpha^{*'} = 1$. For $k < 0$, given $\mu > 0$, $F_2(-\infty; \mu) = +\infty$ and $F_2(0; \mu) = \mu(f(\alpha_1^*)\alpha_1^* + f(\alpha_2^*)\alpha_2^*)$. Because $\frac{\partial F_2(k; \mu)}{\partial k} > 0$, therefore there exists a unique $k = \hat{k} < 0$ such that $F(\hat{k}; \mu) = \int_{-\infty}^{+\infty} f(\alpha^{*'})d\alpha^{*'} = 1$ if and only if $f(\alpha_1^*)\alpha_1^* + f(\alpha_2^*)\alpha_2^* < \frac{1}{\mu}$. Therefore, Proposition 5 is proven.

K Proof of Proposition 6

Proof: Suppose $f_\varepsilon(\varepsilon)$ follows a truncated normal distribution $TN(0, \sigma^2; 0, +\infty)$. Therefore, $f_\varepsilon(\varepsilon)$ is given by

$$f_\varepsilon(\varepsilon) = \begin{cases} \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{\varepsilon^2}{2\sigma^2}) & \varepsilon \in [0, +\infty) \\ 0 & \text{otherwise} \end{cases}$$

Define $f_t(\alpha^{*'}; \alpha_1^*) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^{*'} - \alpha_1^*)^2)f(\alpha_1^*)$ and $f_t(\alpha^{*'}; \alpha_2^*) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^{*'} - \alpha_2^*)^2)f(\alpha_2^*)$. Given $\alpha^* \in \{\alpha_1^*, \alpha_2^*\}$, $m^{-1}(\alpha^{*'}) - m^{-1}(\alpha^*) \geq 0$ divides the range of $\alpha^{*'}$ to differentiate between the situations where $f_\varepsilon(m^{-1}(\alpha^{*'}) - m^{-1}(\alpha^*)) \geq 0$ and where $f(m^{-1}(\alpha^{*'}) - m^{-1}(\alpha^*)) = 0$. Therefore, according to $f(\alpha^{*'}) = \sum_{\alpha^* \in D} f_\varepsilon(m^{-1}(\alpha^{*'}) - m^{-1}(\alpha^*))f(\alpha^*)$, it is obtained that

$$f(\alpha^{*'}) = \begin{cases} 0 & \alpha^{*'} < \alpha_1^* \\ f_t(\alpha^{*'}; \alpha_1^*) & \alpha_1^* \leq \alpha^{*'} < \alpha_2^* \\ f_t(\alpha^{*'}; \alpha_1^*) + f_t(\alpha^{*'}; \alpha_2^*) & \alpha^{*'} \geq \alpha_2^* \end{cases}$$

Define $F(k^2) = \int_{-\infty}^{+\infty} f(\alpha^{*'})d\alpha^{*'} = \int_{\alpha_1^*}^{+\infty} \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^{*'} - \alpha_1^*)^2)f(\alpha_1^*)d\alpha^{*'} + \int_{\alpha_2^*}^{+\infty} \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^{*'} - \alpha_2^*)^2)f(\alpha_2^*)d\alpha^{*'}$. $\frac{\partial F(k^2)}{\partial k^2} < 0$ for all k^2 , $F(0) = +\infty$ and $F(+\infty) = 0$. Therefore, there exists a unique k^2 such that $F(k^2) = \int_{-\infty}^{+\infty} f(\alpha^{*'})d\alpha^{*'} = 1$. The k^2 can be obtained by solving the equation $F(k^2) = 1$.

In the posterior distribution affected by technical errors, $f(\alpha^{*'} = \alpha_1^*) = f(\alpha^{*'} = \alpha_2^*)$ equivalently indicates that $f(\alpha_2^*) = f(\alpha_1^*)(1 - \exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2))$. Only if $f(\alpha_2^*) < f(\alpha_1^*)$, $f(\alpha_2^*) = f(\alpha_1^*)(1 - \exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2))$ happens. Solving $f(\alpha_2^*) = f(\alpha_1^*)(1 - \exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2))$, we obtain that $k^2 = \frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}$. Because $f'(\alpha^{*'}) < 0$

for either $\alpha^{*'} \in [\alpha_1^*, \alpha_2^*)$ or $\alpha^{*'} \in (\alpha_2^*, +\infty)$, therefore $f(\alpha^{*'} = \alpha_1^*) = f(\alpha^{*'} = \alpha_2^*) > f(\alpha^{*'} \neq \alpha_1^* \text{ or } \alpha_2^*)$.

Consider the case where $f(\alpha^{*'} = \alpha_1^*) > f(\alpha^{*'} = \alpha_2^*)$. $f(\alpha^{*'} = \alpha_1^*) > f(\alpha^{*'} = \alpha_2^*)$ indicates that $f(\alpha_2^*) < f(\alpha_1^*)(1 - \exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2))$. Only if $f(\alpha_1^*) > f(\alpha_2^*)$, $f(\alpha_2^*) < f(\alpha_1^*)(1 - \exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2))$ happens. By rearranging $f(\alpha_2^*) < f(\alpha_1^*)(1 - \exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2))$, it is obtained that $k^2 > \frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}$. Because $f'(\alpha^{*'}) < 0$ for either $\alpha^{*'} \in [\alpha_1^*, \alpha_2^*)$ or $\alpha^{*'} \in (\alpha_2^*, +\infty)$, therefore for $\alpha^{*'} \in [\alpha_1^*, \alpha_1^* + \sqrt{\frac{2\sigma^2}{k^2} \ln \frac{1}{\exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2) + \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}}]$, $f(\alpha^{*'}) > f(\alpha^{*'} = \alpha_2^*)$. $\alpha_1^* + \sqrt{\frac{2\sigma^2}{k^2} \ln \frac{1}{\exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2) + \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}}$

is obtained by solving the equation $f(\alpha^{*'} = \alpha_2^*) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^{*'} - \alpha_1^*)^2)f(\alpha_1^*)$. Figure 1 illustrates the result.

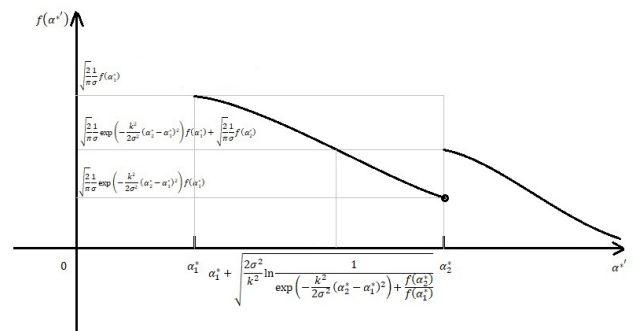


Figure 1. The case of $f(\alpha^{*'})$ where $f(\alpha^{*'} = \alpha_1^*) > f(\alpha^{*'} = \alpha_2^*)$. The black curve depicts $f(\alpha^{*'})$. The grey lines are auxiliary lines for framing the curve. Source: Figure 1 is drawn by the author of this paper.

Consider the case where $f(\alpha^{*'} = \alpha_1^*) < f(\alpha^{*'} = \alpha_2^*)$. $f(\alpha^{*'} = \alpha_1^*) < f(\alpha^{*'} = \alpha_2^*)$ indicates that $f(\alpha_2^*) > f(\alpha_1^*)(1 - \exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2))$. If $f(\alpha_2^*) \geq f(\alpha_1^*)$, for any $k \in \mathbb{R}$ $f(\alpha_2^*) > f(\alpha_1^*)(1 - \exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2))$ is

held. Suppose $f(\alpha_2^*) < f(\alpha_1^*)$. In this situation, by rearranging $f(\alpha_2^*) > f(\alpha_1^*)(1 - \exp(-\frac{k^2}{2\sigma^2}(\alpha_2^* - \alpha_1^*)^2))$, it is obtained that $k^2 < \frac{2\sigma^2}{(\alpha_2^* - \alpha_1^*)^2} \ln \frac{1}{1 - \frac{f(\alpha_2^*)}{f(\alpha_1^*)}}$. Because $f'(\alpha^*) < 0$ for either $\alpha^* \in [\alpha_1^*, \alpha_2^*)$ or $\alpha^* \in (\alpha_2^*, +\infty)$, therefore for $\alpha^* \in [\alpha_1^*, \max \hat{\alpha}^*)$, $f(\alpha^*) > f(\alpha_1^*)$. $\max \hat{\alpha}^*$ is the largest solution of the equation $f(\alpha^* = \alpha_1^*) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_1^*)^2) f(\alpha_1^*) + \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_2^*)^2) f(\alpha_2^*)$. $\sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_1^*)^2) f(\alpha_1^*) + \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_2^*)^2) f(\alpha_2^*)$ as a function with respect to $\alpha^* \in \mathbb{R}$ has two local maximum. One is at $\alpha^* = \alpha_1^*$ and the other at $\alpha^* = \alpha_2^*$. It also has a local minimum $\min \alpha^* \in (\alpha_1^*, \alpha_2^*)$. The function increases with respect to $\alpha^* \in (-\infty, \alpha_1^*)$ or $(\min \alpha^*, \alpha_2^*)$, while decreases with respect to $\alpha^* \in (\alpha_1^*, \min \alpha^*)$ or $(\alpha_2^*, +\infty)$. Therefore, the equation $f(\alpha^* = \alpha_1^*) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_1^*)^2) f(\alpha_1^*) + \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp(-\frac{k^2}{2\sigma^2}(\alpha^* - \alpha_2^*)^2) f(\alpha_2^*)$ has four solutions. Figure 2 illustrates the result.

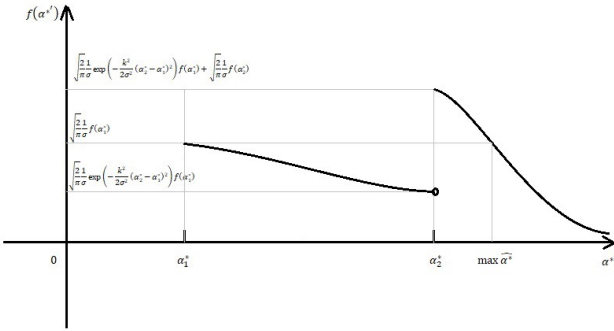


Figure 2. The case of $f(\alpha^*)$ where $f(\alpha^* = \alpha_1^*) < f(\alpha^* = \alpha_2^*)$. The black curve depicts $f(\alpha^*)$. The grey lines are auxiliary lines for framing the curve. Source: Figure 2 is drawn by the author of this paper.

Therefore, Proposition 6 is proven.

L Proof of Proposition 7

Proof: Suppose $f_\varepsilon(\varepsilon)$ follows Weibull distribution:

$$f_\varepsilon(\varepsilon) = \begin{cases} \frac{a}{b^a} \varepsilon^{a-1} \exp(-(\frac{\varepsilon}{b})^a) & \varepsilon \geq 0 \\ 0 & \varepsilon < 0 \end{cases}$$

where $a > 0$ and $b > 0$. Define $f_w(\alpha^*; \alpha_1^*) = \frac{a}{b^a} k^{a-1} (\alpha^* - \alpha_1^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_1^*}{b})^a) f(\alpha_1^*)$ and $f_w(\alpha^*; \alpha_2^*) = \frac{a}{b^a} k^{a-1} (\alpha^* - \alpha_2^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_2^*}{b})^a) f(\alpha_2^*)$. Given $\alpha^* \in \{\alpha_1^*, \alpha_2^*\}$, $m^{-1}(\alpha^*) - m^{-1}(\alpha^*) \geq 0$ divides the range of α^* to differentiate between the situations where $f_\varepsilon(m^{-1}(\alpha^*) - m^{-1}(\alpha^*)) \geq 0$ and where $f(m^{-1}(\alpha^*) - m^{-1}(\alpha^*)) = 0$. Therefore, according to $f(\alpha^*) = \sum_{\alpha^* \in D} f_\varepsilon(m^{-1}(\alpha^*) - m^{-1}(\alpha^*)) f(\alpha^*)$, it is obtained that

$$f(\alpha^*) = \begin{cases} 0 & \alpha^* < \alpha_1^* \\ f_w(\alpha^*; \alpha_1^*) & \alpha_1^* \leq \alpha^* < \alpha_2^* \\ f_w(\alpha^*; \alpha_1^*) + f_w(\alpha^*; \alpha_2^*) & \alpha^* \geq \alpha_2^* \end{cases}$$

Define $T(k) \equiv \int_{-\infty}^{+\infty} f(\alpha^*) d\alpha^* = \frac{a}{b} k^{a-1} g(k)$, where $g(k) = \int_{\alpha_1^*}^{+\infty} (\alpha^* - \alpha_1^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_1^*}{b})^a) f(\alpha_1^*) d\alpha^* + \int_{\alpha_2^*}^{+\infty} (\alpha^* - \alpha_2^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_2^*}{b})^a) f(\alpha_2^*) d\alpha^*$. Hence, $g'(k) = -\alpha k^{a-1} (\int_{\alpha_1^*}^{+\infty} (\alpha^* - \alpha_1^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_1^*}{b})^a) f(\alpha_1^*) (\frac{\alpha^* - \alpha_1^*}{b})^a d\alpha^* + \int_{\alpha_2^*}^{+\infty} (\alpha^* - \alpha_2^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_2^*}{b})^a) f(\alpha_2^*) (\frac{\alpha^* - \alpha_2^*}{b})^a d\alpha^*)$. $T'(k) = \frac{a}{b} ((a-1)k^{a-2} g(k) + k^{a-1} g'(k))$. Therefore, if $a-1$ is an even integer and $k < 0$, $T'(k) < 0$; otherwise, $T(k)$ is not monotonic with respect to k .

Therefore, if $a-1$ is an odd integer, k can only be positive if it exists to ensure that $T(k) > 0$. However, if $a-1$ is an odd integer, $T(k)$ is not monotonic and therefore the existence of k is not guaranteed. If $a-1$ is an even integer and $k < 0$, $T(-\infty) = +\infty$, $T(0) = 0$ and $T'(k) < 0$. Therefore, there exists a unique $k \in (-\infty, 0)$ such that $T(k) \equiv \int_{-\infty}^{+\infty} f(\alpha^*) d\alpha^* = 1$. If $a-1$ is an even integer and $k > 0$, again $T(k)$ is not monotonic and therefore the existence of k is not guaranteed.

If $a \leq 1$, $k < 0$ cannot be held because it will make $T(k)$ an imaginary number. Therefore, in this situation, $k > 0$. Given that $a \leq 1$, $T(0) = +\infty$, $T(+\infty) = 0$ and $T'(k) < 0$. Therefore, there exists a unique $k \in (0, +\infty)$ such that $T(k) \equiv \int_{-\infty}^{+\infty} f(\alpha^*) d\alpha^* = 1$. Proposition 7 is proven.

M Properties of $f(\alpha^*)$ where the technical error follows a Weibull distribution

Proof: For $\alpha^* \in [\alpha_1^*, \alpha_2^*)$, $f(\alpha^*) = \frac{a}{b^a} k^{a-1} (\alpha^* - \alpha_1^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_1^*}{b})^a) f(\alpha_1^*)$. Hence, $f'(\alpha^*) = \frac{a}{b^a} k^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_1^*}{b})^a) f(\alpha_1^*) ((a-1)(\alpha^* - \alpha_1^*)^{a-2} - k^a \frac{a}{b} (\alpha^* - \alpha_1^*)^{a-1} (\frac{\alpha^* - \alpha_1^*}{b})^{a-1})$. Therefore, $f'(\alpha^*) \geq 0$ equivalently indicates that $\frac{b^a}{a} \frac{a-1}{k^a} \geq (\alpha^* - \alpha_1^*)^a$. Because $f(\alpha^*) \geq 0$ for $\alpha^* \in [\alpha_1^*, \alpha_2^*)$, therefore $k^{a-1} > 0$. $f'(\alpha^*) \geq 0$ can be equivalently rearranged to $\frac{b^a}{a} \frac{a-1}{k^a} \geq (\alpha^* - \alpha_1^*)^a$. Therefore, if $a < 1$, $f'(\alpha^*) < 0$ for $\alpha^* \in [\alpha_1^*, \alpha_2^*)$. In this situation, $\lim_{\alpha^* \rightarrow \alpha_1^*} f(\alpha^*) = 0$. Therefore, to ensure $f(\alpha^*) \geq 0$, in this situation $f(\alpha^*)$ has to be normalized such that for $\alpha^* \in [\alpha_1^*, \alpha_2^*)$, $f(\alpha^*) = 0$. Therefore, in this situation, there exists a discontinuity point $\alpha^* = \alpha_2^*$ for $f(\alpha^*)$ because $f(\alpha^* = \alpha_2^*) > 0$; if $a > 1$, $f(\alpha^*) = \frac{a}{b^a} k^{a-1} (\alpha^* - \alpha_1^*)^{a-1} \exp(-k^a (\frac{\alpha^* - \alpha_1^*}{b})^a) f(\alpha_1^*)$ increases with respect to α^* for $\alpha^* \in [\alpha_1^*, \alpha_1^*)$, and decreases with respect to α^* for $[\hat{\alpha}_1^*, +\infty)$. Because we focus on the situation where $\hat{\alpha}_1^* > \alpha_2^*$, therefore $f(\alpha^*) > 0$ for $\alpha^* \in [\alpha_1^*, \alpha_2^*)$

if $a > 1$.

For $\alpha^{*'} \in [\alpha_2^*, +\infty)$, $f(\alpha^{*'}) = \frac{a}{b^a} k^{a-1} \{(\alpha^{*'} - \alpha_1^*)^{a-1} \exp(-k^a (\frac{\alpha^{*'} - \alpha_1^*}{b})^a) f(\alpha_1^*) + (\alpha^{*'} - \alpha_2^*)^{a-1} \exp(-k^a (\frac{\alpha^{*'} - \alpha_2^*}{b})^a) f(\alpha_2^*)\}$. $f'(\alpha^{*'}) \geq 0$ can be equivalently rearranged to $\frac{a-1}{a} \frac{b^a}{k^a} \geq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a$, where $\phi(\alpha^{*'}) = \frac{f(\alpha_1^*) \exp(-k^a (\frac{\alpha^{*'} - \alpha_1^*}{b})^a) (\alpha^{*'} - \alpha_1^*)^{a-2}}{f(\alpha_1^*) \exp(-k^a (\frac{\alpha^{*'} - \alpha_1^*}{b})^a) (\alpha^{*'} - \alpha_1^*)^{a-2} + f(\alpha_2^*) \exp(-k^a (\frac{\alpha^{*'} - \alpha_2^*}{b})^a) (\alpha^{*'} - \alpha_2^*)^{a-2}}$. Therefore, for $\alpha^{*'} \in [\alpha_2^*, +\infty)$, the local optimum of $f(\alpha^{*'})$ are the solutions of equation

$$\frac{a-1}{a} \frac{b^a}{k^a} = \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a$$

which is equation (16). The equation can have $n \geq 1$ solutions. Equation (16) is equivalent to $f'(\alpha^{*'}) = 0$ for $\alpha^{*'} \geq \alpha_2^*$. Therefore, the n solutions of equation (16), $\hat{\alpha}_{(1)}^* < \hat{\alpha}_{(2)}^* < \dots < \hat{\alpha}_{(m)}^* < \hat{\alpha}_{(m+1)}^* < \dots < \hat{\alpha}_{(n)}^*$ are greater than α_2^* .

It is known that $\min\{(\alpha^{*'} - \alpha_1^*)^a, (\alpha^{*'} - \alpha_2^*)^a\} \leq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a \leq \max\{(\alpha^{*'} - \alpha_1^*)^a, (\alpha^{*'} - \alpha_2^*)^a\}$. Because $\alpha^{*'} \in [\alpha_2^*, +\infty)$, therefore $(\alpha^{*'} - \alpha_2^*)^a \leq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a \leq (\alpha^{*'} - \alpha_1^*)^a$. Define $T(\alpha^{*'}; \alpha^*) = (\alpha^{*'} - \alpha^*)^a$, which is a function with respect to $\alpha^{*'}$ given α^* . If $(-1)^{a-1} < 0$, then for $\alpha^{*'} < \alpha^*$, $T'(\alpha^{*'}; \alpha^*) < 0$; for $\alpha^{*'} > \alpha^*$, $T'(\alpha^{*'}; \alpha^*) > 0$. If $(-1)^{a-1} > 0$, $T'(\alpha^{*'}; \alpha^*) \geq 0$ for all $\alpha^{*'}$. Denote the solution of $\frac{a-1}{a} \frac{b^a}{k^a} = (\alpha^{*'} - \alpha_1^*)^a$ by $\hat{\alpha}_1^*$, and the solution of $\frac{a-1}{a} \frac{b^a}{k^a} = (\alpha^{*'} - \alpha_2^*)^a$ by $\hat{\alpha}_2^*$. Solving the two equations, we obtain that $\hat{\alpha}_1^* = \alpha_1^* + (\frac{a-1}{a} \frac{b^a}{k^a})^{\frac{1}{a}}$ and $\hat{\alpha}_2^* = \alpha_2^* + (\frac{a-1}{a} \frac{b^a}{k^a})^{\frac{1}{a}}$. $f(\alpha^{*'})$ having local optima for $\alpha^{*'} \in [\alpha_2^*, +\infty)$ requires that $(a-1)k^a > 0$ so that the equation $\frac{a-1}{a} \frac{b^a}{k^a} = \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a$ can be held. Therefore, $\hat{\alpha}_1^* > \alpha_1^*$ and $\hat{\alpha}_2^* > \alpha_2^*$.

$\hat{\alpha}_1^*$ is not necessarily greater than α_2^* . It is the local optima of $f(\alpha^{*'})$ for $\alpha^{*'} \in [\alpha_2^*, +\infty)$ that are our concern, and that's why, with generality, we focus on a case where $\hat{\alpha}_1^* > \alpha_2^*$. For $(-1)^{a-1} > 0$, $(\alpha^{*'} - \alpha_2^*)^a \leq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a \leq (\alpha^{*'} - \alpha_1^*)^a$ is always held with respect to all $\alpha^{*'}$. For $(-1)^{a-1} < 0$, $T'(\alpha^{*'}; \alpha_1^*) > 0$ for $\alpha^{*'} > \alpha_1^*$ and $T'(\alpha^{*'}; \alpha_2^*) > 0$ for $\alpha^{*'} > \alpha_2^*$. Because of $\alpha_1^* < \alpha_2^*$, hence for $\alpha^{*'} \geq \alpha_2^*$, both $T'(\alpha^{*'}; \alpha_1^*) > 0$ and $T'(\alpha^{*'}; \alpha_2^*) > 0$ are held. Therefore, for $(-1)^{a-1} < 0$, $(\alpha^{*'} - \alpha_2^*)^a \leq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a \leq (\alpha^{*'} - \alpha_1^*)^a$ is always held with respect to $\alpha^{*'}$ in $[\alpha_2^*, +\infty)$.

Therefore, given that $\alpha_2^* < \hat{\alpha}_1^*$, no matter whether $(-1)^{a-1} \geq 0$, for $\alpha^{*'} \in [\alpha_2^*, +\infty)$, both $T'(\alpha^{*'}; \alpha_1^*) > 0$ and $T'(\alpha^{*'}; \alpha_2^*) > 0$ are held, and $(\alpha^{*'} - \alpha_2^*)^a \leq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a \leq (\alpha^{*'} - \alpha_1^*)^a$ is always held with respect to $\alpha^{*'}$ in $[\alpha_2^*, +\infty)$. Therefore, solving the equations $\frac{a-1}{a} \frac{b^a}{k^a} = (\alpha^{*'} - \alpha_2^*)^a \equiv T(\alpha^{*'}; \alpha_2^*)$,

$\frac{a-1}{a} \frac{b^a}{k^a} = \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a$ and $\frac{a-1}{a} \frac{b^a}{k^a} = (\alpha^{*'} - \alpha_1^*)^a \equiv T(\alpha^{*'}; \alpha_1^*)$, due to the relationship $(\alpha^{*'} - \alpha_2^*)^a \leq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a \leq (\alpha^{*'} - \alpha_1^*)^a$, therefore $\hat{\alpha}_1^* \leq \hat{\alpha}_2^* \leq \hat{\alpha}_1^*$, where $\hat{\alpha}_2^*$ is any solution of the equation $\frac{a-1}{a} \frac{b^a}{k^a} = \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a$.

Therefore, for $\alpha_2^* < \alpha^{*'} < \hat{\alpha}_1^*$, $(\alpha^{*'} - \alpha_2^*)^a \leq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a \leq (\alpha^{*'} - \alpha_1^*)^a < \frac{a-1}{a} \frac{b^a}{k^a}$, which implies that $f'(\alpha^{*'}) > 0$; for $\hat{\alpha}_2^* < \alpha^{*'} < +\infty$, $\frac{a-1}{a} \frac{b^a}{k^a} < (\alpha^{*'} - \alpha_2^*)^a \leq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a \leq (\alpha^{*'} - \alpha_1^*)^a$, which implies that $f'(\alpha^{*'}) < 0$. Therefore, for $a > 1$, because $f(\alpha^{*'}) \geq 0$ for $\alpha^{*'} \in [\alpha_1^*, \alpha_2^*)$, and $f(\alpha_2^*) = \lim_{\alpha^{*'} \rightarrow \alpha_2^*}$ and hence continuity at $\alpha^{*'} = \alpha_2^*$ is ensured, therefore for $\alpha^{*'} \in [\alpha_1^*, \hat{\alpha}_1^*)$, $f'(\alpha^{*'}) \geq 0$ and $f(\alpha^{*'})$ is continuous. However, for $a < 1$, as it is known, $f(\alpha^{*'}) = 0$ for $\alpha^{*'} \in [\alpha_1^*, \alpha_2^*)$. Therefore, in this situation, $f(\alpha^{*'})$ increases with respect to $\alpha^{*'}$ in $[\alpha_2^*, \hat{\alpha}_1^*)$.

$\hat{\alpha}_1^* > \alpha_2^*$ can be equivalently rearranged to $k^a \leq \frac{a-1}{a} \frac{b^a}{(\alpha_2^* - \alpha_1^*)^a}$. Therefore, under the condition that $k^a \leq \frac{a-1}{a} \frac{b^a}{(\alpha_2^* - \alpha_1^*)^a}$, the local optima of $f(\alpha^{*'})$ for $\alpha^{*'} \in [\alpha_2^*, +\infty)$, which belong to $\alpha^{*'} \in [\hat{\alpha}_1^*, \hat{\alpha}_2^*]$, systematically upward bias to the value of the optimal ordering quantity α_1^* and α_2^* obtained without the influence of the technical error. Recall the existence condition $(a-1)k^a > 0$. Hence, ensuring the existence of local optimum of $f(\alpha^{*'})$ for $[\alpha_2^*, +\infty)$, $a \neq 1$. Therefore, if a is an odd integer, the existence condition implies that $k > 0$, and therefore $k^a \leq \frac{a-1}{a} \frac{b^a}{(\alpha_2^* - \alpha_1^*)^a}$ in this situation can be equivalently rearranged to $k \in (0, (\frac{a-1}{a})^{\frac{1}{a}} \frac{b}{(\alpha_2^* - \alpha_1^*)})$. If a is an even integer, $k^a \leq \frac{a-1}{a} \frac{b^a}{(\alpha_2^* - \alpha_1^*)^a}$ in this situation can be equivalently rearranged to $k \in (-(\frac{a-1}{a})^{\frac{1}{a}} \frac{b}{(\alpha_2^* - \alpha_1^*)}, (\frac{a-1}{a})^{\frac{1}{a}} \frac{b}{(\alpha_2^* - \alpha_1^*)})$.

The n solutions of $\frac{a-1}{a} \frac{b^a}{k^a} = \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a$ are the local optima of $f(\alpha^{*'})$ for $\alpha^{*'} \in [\alpha_2^*, +\infty)$, which include local minima and local maxima. According to the continuity of $f(\alpha^{*'})$ for $\alpha^{*'} \in [\alpha_2^*, +\infty)$, with generality, the local minima and local maxima is located along the order that *...maximum-minimum-maximum-minimum...* as $\alpha^{*'}$ increases. No matter whether $(-1)^{a-1} \geq 0$, $f(\alpha^{*'})$ increases with respect to $\alpha^{*'}$ for $\alpha^{*'} \in [\alpha_2^*, \hat{\alpha}_1^*)$ and decreases with respect to $\alpha^{*'}$ for $\alpha^{*'} \in [\hat{\alpha}_2^*, +\infty)$. Therefore, according to the continuity of $f(\alpha^{*'})$, the first and last optimum of $f(\alpha^{*'})$ for $\alpha^{*'} \in [\hat{\alpha}_1^*, \hat{\alpha}_2^*] \subset [\alpha_2^*, +\infty)$ are local maxima. Due to the arrangement order that *...maximum - minimum - maximum - minimum...*, therefore n is an odd integer with generality. There are $\frac{n-1}{2}$ local minima and $\frac{n+1}{2}$ local maxima. Besides, with generality, according to the arrangement order that *...maximum - minimum - maximum - minimum...*, for $\alpha^{*'} \in [\hat{\alpha}_m^*, \hat{\alpha}_{m+1}^*)$, if m is an odd integer, $f'(\alpha^{*'}) \leq 0$; if m is an even integer $f'(\alpha^{*'}) \geq 0$.

From above discussion, it has been known that for $\alpha^{*'} \in [\alpha_2^*, +\infty)$, if and only if $\frac{a-1}{a} \frac{b^a}{k^a} \leq \phi(\alpha^{*'}) (\alpha^{*'} - \alpha_1^*)^a + (1 - \phi(\alpha^{*'})) (\alpha^{*'} - \alpha_2^*)^a$, $f'(\alpha^{*'}) \leq 0$. Therefore, if a

is an odd integer and $k < 0$, or if $a < 1$ and hence it determines that $k > 0$ by Proposition 7, $f'(\alpha^{*'}) \leq 0$ for all $\alpha^{*'} \in [\alpha_2^*, +\infty)$. Therefore, no matter whether $f(\alpha^{*'})$ is continuous or discontinuous at $\alpha^{*'} = \alpha_2^*$, $f(\alpha^{*'})$ is unimodal for $\alpha^{*'} \in \mathbb{R}$ and the mode is taken at $\alpha^{*'} = \alpha_2^*$.

REFERENCES

- [1] Boyacı T, Akçay Y (2017) Pricing When Customers Have Limited Attention. *Management Science*. ePub ahead of print May 25. <https://doi.org/10.1287/mnsc.2017.2755>
- [2] Caplin A, Dean M, Leahy J (2019) Rational Inattention, Optimal Consideration Sets, and Stochastic Choice. *Review of Economic Studies*. 86(3): 1061-1094. <https://doi.org/10.1093/restud/rdy037>
- [3] Cover T, Thomas J (2006) *Elements of Information Theory*, 2nd ed. (John Wiley & Sons, Inc).
- [4] Friston K (2009) The Free Energy Principle: a Rough Guide to the Brain? *Trends in Cognitive Sciences*. 13(7): 293-301. doi:10.1016/j.tics.2009.04.005
- [5] Friston K (2010) Free Energy Principle: a Unified Brain Theory? *Nature Reviews Neuroscience*. Advance online publication, 13 Jan 2010. doi:10.1038/nrn2787
- [6] Jung J, Kim J H, Matějka F, Sims C A (2019) Discrete Actions in Information-Constrained Decision Problems. rdz011. <https://doi.org/10.1093/restud/rdz011>
- [7] Kahneman D (1973) *Attention and Effort*. (Prentice-Hall, Inc., Englewood Cliffs, New Jersey).
- [8] Lachman R, Lachman J L, Butterfield B C (1979) *Cognitive Psychology and Information Processing: An Introduction*. (Psychology Press, Taylor & Francis Group, London and New York).
- [9] 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.
- [10] Matějka F (2016) Rationally Inattentive Seller: Sales and Discrete Pricing. *Review of Economic Studies*. 83(3): 1125-1155. <https://doi.org/10.1093/restud/rdv049>
- [11] Schweitzer M E, Cachon G P (2000) Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence. *Management Science*. 46(3): 404-420.
- [12] Shannon C E (1948) A Mathematical Theory of Communication. *The Bell System Technical Journal*. 27: 379-423, 623-656.
- [13] Simon H (1955) A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*. 69(1): 99-118.
- [14] Sims C A (1998) Stickiness. *Carnegie-Rochester Conference Series on Public Policy*. 49: 317-356.
- [15] Sims C A (2003) Implications of Rational Inattention. *Journal of Monetary Economics*. 50(3): 665-690.
- [16] Sims C A (2006) *Rational Inattention: A Research Agenda*. Working Paper, Princeton University.
- [17] Tutino A (2011) 'Rational Inattention' guides overloaded brains, help economists understand market behavior. *Economic Letter*, Federal Reserve Bank of Dallas, vol. 6 (Mar).
- [18] Yang M (2015) Coordination with Flexible Information Acquisition. *Journal of Economic Theory*. 158 (Part B): 721-738.