OPTIMAL INQUIRY

TAI-WEI HU AND ANDRIY ZAPECHELNYUK

ABSTRACT. A decision maker acquires and processes information about an uncertain state of nature through an inquiry—a contingent sequence of questions culminating in a decision. Inquiries are costly, with costs proportional to their length. We characterize optimal inquiries and uncover two behavioral implications of costly inquiry: attention span reduction (favoring shorter inquiries by deprioritizing some decisions or excluding them from consideration) and confirmation bias (seeking evidence to confirm prior guesses of optimal decisions). Our framework provides a rational foundation to prominent cognitive biases, such as framing and search satisficing in healthcare, and tunnel vision in criminal investigations.

JEL Classification: D81, D83.

Keywords: Bounded rationality, information theory, rational inattention, attention span, confirmation bias, consideration set, framing, search satisficing, tunnel vision

Date: January 28, 2025

Tai-Wei Hu: School of Economics, University of Bristol, 12 Priory Road, Bristol, BS8 1TU, UK. *E-mail:* taiwei.hu@bristol.ac.uk
Andriy Zapechelnyuk: School of Economics, University of Edinburgh, 31 Buccleuch Place, Edinburgh, EH8 9JT, UK. *E-mail:* azapech@gmail.com.

We are grateful to Olivier Compte, Francesc Dilmé, Matthew Elliott, Philippe Jehiel, Benson Leung, Michael Mandler, John Moore, Larry Samuelson, Jakub Steiner, Colin Stewart, Ina Taneva, Tymon Tatur, Chris Tyson, Peter Wagner, and Zaifu Yang, as well as many conference and seminar participants, for helpful comments.

1. Introduction

Inquiry is one of the most frequent and important modes of information processing in our daily life. Examples are abundant. A doctor visit usually consists of a series of questions from reception to actual consultation of the patient's conditions. A crime investigation typically consists of a series of queries and processing the responses. Inquiry about characteristics of products or services is an important aspect of the procurement process in organisations. In all these examples, information to be gathered can be potentially overwhelming, whereas resources available to process it are limited and precious. In this paper, we propose a theory of optimal inquiry that incorporates a dynamic procedure of costly information processing, with novel behavioral implications on attention span and confirmation bias.

We formalize an *inquiry* as a decision maker's strategy of asking questions about a relevant state of nature. It starts with an initial question and a contingent plan that specifies which question to ask depending on the answers to the previous ones. As in the standard Bayesian paradigm, answers to the inquiry determine the posterior information that guides the decision maker's final choice.

Our framework provides an explicit and intuitive procedure for information processing. It has the same backbone motivation as what gave rise to the rational inattention literature (surveyed in Maćkowiak et al., 2023). The main departure of our approach from this literature is that we focus on the dynamic process of inquiry with an endogenous choice of the optimal procedure. This allows us to obtain behavioral implications that are of dynamic nature, such as an endogenous preference for a shorter attention span and a prioritization of certain salient decisions before considering others.

Moreover, our cost of inquiry is directly associated with the acts of asking questions and processing their answers, and hence the cost is independent of the decision maker's beliefs. This cost reflects the burden of the decision maker's cognitive activity or the value of physical resources (such as gathering evidence) needed for the inquiry. For example, the cost of performing a blood glucose test and processing its result (in terms of physical or cognitive resources) is independent of the patient's medical history. In contrast, in the standard rational inattention model, the cost is an entropy-based function of the decision maker's prior beliefs. This dependence on the prior can be unrealistic in certain applications and has a conceptual problem if applied to game situations (Denti et al., 2022).[1]

Our main result is a characterization of optimal inquiry in terms of its dynamic outcomes: the likelihood of different actions to be chosen and the sequence of questions that are asked to arrive at different actions. We utilize two well-known results from the information theory—the Kraft inequality (Kraft, 1949) and the Huffman coding (Huffman, 1952)—to characterize the set of payoff-relevant outcomes implementable by an optimal inquiry. Any such outcome consists of two components—*form* and *content*. The form includes a *consideration set*, which is a subset of feasible actions that are used with a positive probability in that outcome, and a *length profile*, which specifies how many questions are asked to reach each action in the consideration set. The content is an *information partition*, which describes the posterior information about the state upon reaching each action in the consideration set.

Our characterization of optimal inquiry relies on two interconnected principles of optimality. First, we show that the form determines the content: given an optimal consideration set and an associated length profile, the optimal information partition is determined by simple indifference conditions. Second, we show that the content is also informative about the form: given an information partition, the optimal length profile is determined by the Huffman coding. This implies a negative correlation between the ex ante likelihood of choosing an action and the inquiry length that leads to that action. That is, more likely actions are prioritized and considered before other actions.

The above two principles provide a formal insight into the key trade-off that an optimal inquiry balances—the accuracy of information processed against the number of questions needed to achieve it—and lead us to two behavioral implications from this trade-off.

First we consider implications related to the form of optimal inquiry, and define *attention span* as the expected number of questions the decision maker asks before taking an action. We show that the decision maker optimally reduces her attention span as the cost of each question rises. This is achieved either by dropping some

[1]Caplin et al. (2022) also point out that the entropy-based cost function has implications that are empirically counterfactual. Several recent papers, such as Bloedel and Zhong (2024), also consider more general cost structures. We defer the discussion of those papers to the Related Literature.

actions out of the consideration set, or by prioritising some actions over others, or both. At the extreme, when the cost is very low, all feasible options are considered, and it takes as many questions as needed to distinguish them all. At the other extreme, when the cost is very high, no information is processed, and the action is chosen according to the prior belief.

Second, we consider implications to the content of the optimal inquiry. We show that optimal inquiry always exhibits *confirmation bias*: the decision maker optimally seeks information to confirm her prevalent hypothesis of which actions are optimal. This formalizes the informal definition of confirmation bias in psychology such as Nickerson (1998): "It refers usually to unwitting selectivity in the acquisition and use of evidence." We uncover an economic mechanism for the confirmation bias to occur optimally. Because asking questions is costly, the decision maker is willing to make suboptimal choices that are reached after fewer questions. At the same time, ex ante more likely choices are optimally prioritized with fewer questions to confirm them. These two forces together lead to an endogenous confirmation bias.

We apply our model to provide a rational explanation to some documented behavioural biases. Our leading example considers the phenomena of framing and search satisficing that lead to misdiagnosis in primary healthcare (Croskerry et al., 2013). Through the lens of our model, we show that the pressure to end inquiry early can lead to a biased process. The features such as "premature diagnosis" and "search satisficing" can be explained by our confirmation bias, whereas "framing" can be understood as how a doctor's prior beliefs can magnify this bias or determine its direction. Two further case studies are given at the end of the paper. In the case of criminal justice, the literature has argued that cognitive biases such as "tunnel vision" can lead to wrongful convictions (e.g., Gould and Leo, 2010). We show that in our model a higher rate of wrongful conviction can be linked to higher cost associated with a stronger pressure to solve the case fast, and provide a potential explanation of the "tunnel vision". Finally, in a procurement setting, we show that the optimal inquiry process can lead a biased selection, according to which the procurer prioritises some suppliers and shows a favorable bias toward them, even when all suppliers are ex ante symmetric.

Related Literature. This paper makes a conceptual and methodological contribution to three strands of literature.

The first strand includes papers that formulate and study decision making with cognitive limitations. A popular approach in this literature is rational inattention initiated by Sims (2003). It treats limited cognition as costly information acquisition. The cost of acquiring information is postulated as an ex-ante cost function, typically modelled as entropy reduction relative to the prior belief, as in Matějka and McKay (2015) and Jung et al. (2019). More recent papers consider other cost functions. Morris and Strack (2019) introduce an alternative ex-ante cost function motivated by the classic sequential sampling problem of Wald (1945). Hébert and Woodford (2021) propose neighborhood-based cost functions that capture notions of perceptual distance. Pomatto et al. (2023) characterize ex-ante cost functions that satisfy several economically interpretable axioms. Bloedel and Zhong (2024) provide general conditions for ex-ante cost functions to arise from dynamic models of information acquisition. Unlike this literature, we focus on a concrete but intuitive dynamic model where the cost of information is directly associated with asking questions. The dynamic nature of the process and the sequencing of questions matters and has behavioral implications. This approach allows us to capture certain behavioral concepts in a meaningful way with novel insights. In Section 6 we provide a detailed discussion of which of our conclusions are different from those of rational inattention.

Cognitive limitations of a decision maker have also been modeled without reducing them to an ex-ante cost function. Wilson (2014), following the approach of Cover and Thomas (2006), formulates the decision-making process as a finite automaton. The main result in Wilson (2014) is a dynamic-consistency type of result called multi-self consistency. The cognitive constraint is modelled via an exogenously given number of memory states that capture the decision maker's memory capacity. In contrast, our model is dynamically consistent in the conventional sense, and the size of the optimal inquiry is endogenous. Cremer et al. (2007) propose a model of organizational language using codes, with the main trade-off between the use of broader codes, which are easier to process, and the precision of such codes. Dilmé (2024) studies the relationship between precision and complexity of the optimal codes in the context of efficient communication. While our model shares a similar trade-off, it is dynamic in nature with implications on the timing of information processing.

Similar to our paper, Mandler (2024) proposes a model in which the decision maker acquires information by asking questions modelled as a partition of states. The research question addressed in Mandler (2024) is complementary to our paper. The focus of Mandler (2024) is on the implementation of an exogenously given decision rule at minimum cost, with implications on how the inquiries should be structured. In contrast, our model features an endogenous decision rule that is jointly determined with the inquiry tree and information structure, which allows us to show the connection between cognitive costs and behavioral biases.

The second strand of literature includes papers that study behavioral biases with cognitive frictions. These papers range from axiomatic to constrained optimization approaches, the former including Masatlioglu et al. (2012) and Manzini and Mariotti (2014) and the latter including Caplin et al. (2019). While our approach is closer to the latter, we connect the two approaches by showing that our optimal inquiry satisfies certain desirable axioms, such as dynamic consistency and the attention-filter property of Masatlioglu et al. (2012).

The third strand rationalizes confirmation bias. The wisdom from the literature is that frictions in information processing tend to cause a decision maker to favor signals that confirm the prior belief. Wilson's (2014) model generates this form of confirmation bias based on limited memory. However, in her model the decision maker does not seek evidence but passively processes it. In contrast, our decision maker actively seeks evidence to confirm her more likely options. Steiner et al. (2017) obtain a "status quo bias" in a dynamic rational inattention model where the decision maker tends to stick to actions that are optimal ex-ante. Nimark and Sundaresan (2019) also obtain a "confirmation effect," meaning that the decision maker adopts signal structures in favor of the prior belief. All these papers argue that certain implications from the proposed models can be interpreted as confirmation bias and emphasize the importance of the prior belief. Jehiel and Steiner (2020) obtain confirmation bias in a model where the decision maker chooses whether or not to continue to receive more signals, but can only remember the last one received. Confirmation bias here means that the agent is more likely to stop when seeing a signal in favor of the prior. In contrast, our confirmation bias features an endogenous signal structure with the emphasis on the dynamic nature of information processing. Moreover, we provide a broad formal definition of confirmation bias that is not specific to our model and can be applied to other frameworks.

The paper is organized as follows. Section 2 provides a motivational example. Section 3 introduces the model. Section 4 establishes the key principles of optimality, characterizes optimal inquiries, and introduces two main behavioral implications: attention span and confirmation bias. Section 5 presents two more case studies that illustrate potential applications of our model. Section 6 shows the key distinctions between our results and those from the rational inattention literature. The proofs are relegated to the Appendix.

2. The leading example

In this section, we illustrate our main findings by an example in the context of misdiagnosis in primary health care. As Singh et al. (2017) argue:

> "Diagnosis in primary care [...] represents a high-risk area for errors. PCPs [primary care physicians] typically face high patient volumes and make decisions amid uncertainty. [...] PCPs need to carefully balance the risk of missing serious illness with the wise use of often scarce and costly referral and testing resources."

The problem of misdiagnosis is significant. Singh et al. (2014) estimate "a rate of outpatient diagnostic errors of 5.08%, or approximately 12 million US adults every year." Croskerry et al. (2013) attest:

> "...One of the principal factors underlying diagnostic error is bias. Post hoc analyses of diagnostic errors have in fact suggested that flaws in clinical reasoning rather than lack of knowledge underlie cognitive diagnostic errors, and there is some experimental evidence that, at least when problems are complex, errors were associated with intuitive judgments and could be repaired by analytical reasoning. Moreover, a few experimental studies have supported the claim that bias may misdirect diagnostic reasoning, thus leading to errors."

Our example is motivated by a case study in Croskerry et al. (2013), in which a patient complained about constipation but was actually suffering from Cauda Equina

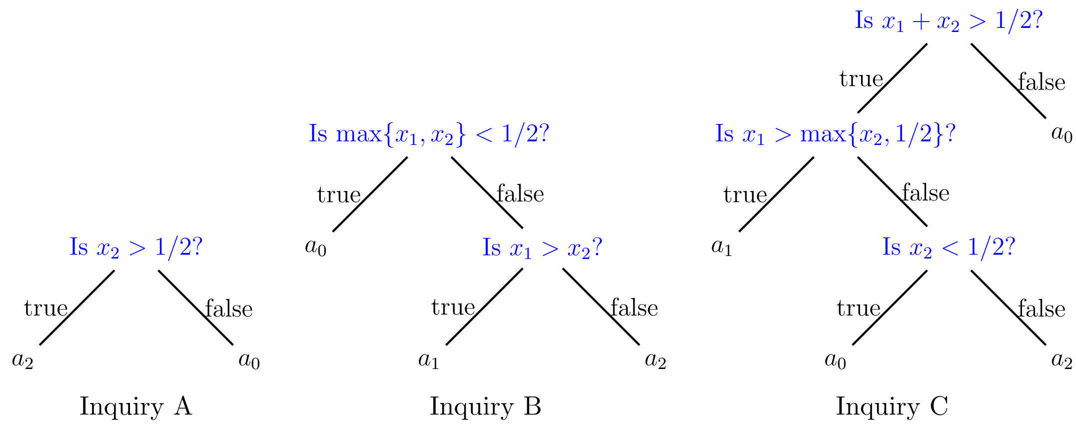


FIGURE 1. Examples of inquiries

Syndrome that requires a specialist or emergency care. Instead, the doctor prescribed a laxative and sent the patient home.

We describe this situation using the following stylized model. The patient's condition is summarized by a state $x = (x_1, x_2) \in [0, 1] \times [0, 1]$, where x_1 represents the diagnostic difficulty of the case and x_2 represents its seriousness. The doctor has three possible actions: to treat the symptom and send the patient home (labelled as action a_0), to refer the patient to a specialist (labeled as action a_1), or to send the patient to emergency (labeled as action a_2). Action a_0 is ideal for conditions that are not too difficult and not too serious, action a_1 is ideal for difficult cases, and action a_2 is ideal for serious cases. Accordingly, the doctor's gross payoffs from these actions are given by the quadratic loss relative to the respective ideal states $(0, 0)$, $(1, 0)$, and $(0, 1)$: $U(a_0, x) = -(x_1)^2 - (x_2)^2$, $U(a_1, x) = -(x_1 - 1)^2 - (x_2)^2$, and $U(a_2, x) = -(x_1)^2 - (x_2 - 1)^2$. For convenience, fix a default action, say, a_0, and consider the utility $u(a, x)$ from each action $a \in \{a_0, a_1, a_2\}$ as compared to a_0, $u(a, x) = U(a, x) - U(a_0, x)$. Thus,

$$u(a_0, x) = 0, \; u(a_1, x) = 2x_1 - 1, \; u(a_2, x) = 2x_2 - 1. \tag{1}$$

The doctor is initially uninformed about x. Note that the doctor does not need to discover x precisely, and she only needs to find out enough to choose a treatment. To learn about x, the doctor follows an *inquiry.* An inquiry is a strategy of how to ask questions, which starts with an initial question, specifies follow-up questions depending on earlier answers, and eventually prescribes an action. Each question asks whether a proposition about x is true or false; these can be propositions about different dimensions of x or propositions about the relationship across different dimensions. Examples of inquiries are shown in Figure 1.

A cost of λ is deducted from the doctor's utility whenever she asks a question. This is interpreted as the opportunity cost of time and cognitive effort spent on a patient that could have been spent to diagnose and treat other patients. In Croskerry et al. (2013) and Singh et al. (2017), this cost is regarded as an important factor that affects the doctor's investigation and the resulting decision. If the doctor reaches an action a after asking ℓ questions, the resulting payoff is $u(a,x) - \lambda\ell$. The doctor chooses an inquiry to maximize her expected utility net of the cost, given her prior knowledge. The prior knowledge is modeled as a prior distribution over x, assumed to have full support and density on $[0,1]^2$.

As a benchmark, suppose that there is no cost of asking questions, $\lambda = 0$. Observe that, by (1), for each $a \in \{a_0, a_1, a_2\}$, it is optimal to choose a if and only if $x \in I_a^0$, where

$$I_{a_0}^0 = \{x : \max\{x_1, x_2\} < 1/2\}, \quad I_{a_1}^0 = \{x : \max\{x_1, x_2\} \geq 1/2 \text{ and } x_1 > x_2\},$$
$$I_{a_2}^0 = \{x : \max\{x_1, x_2\} \geq 1/2 \text{ and } x_1 \leq x_2\},$$

as illustrated by Inquiry B in Figure 1. We refer to $I^0 = \{I_{a_0}^0, I_{a_1}^0, I_{a_2}^0\}$ as an *information partition*. Inquiries B and C (Figure 1) both achieve this outcome. However, when questions are costly, $\lambda > 0$, the two inquiries differ significantly in terms of the cost: when action a_0 is taken, it takes only one question in inquiry B but it may take three questions in inquiry C; when action a_2 is taken, it takes two questions in B but three questions in C. Moreover, the doctor may find it optimal to trade off some accuracy of information about x to reduce the cost of inquiry; in other words, the optimal information partition is endogenously determined by the cost.

Our main results show that the doctor will design the inquiry to induce the optimal information partition to balance the cost against the accuracy. For this example, suppose that action a_0 is the doctor's choice that is most likely to be the correct under the prior distribution (that is, $I_{a_0}^0$ has a higher probability than both $I_{a_1}^0$ and $I_{a_2}^0$). Then, for a range of λ's that are not too high, the optimal inquiry is shown in Figure 2a. It poses one question to reach a_0 and two questions to reach a_1 or a_2. Taking the cost of questions into account, optimal choices are now determined by

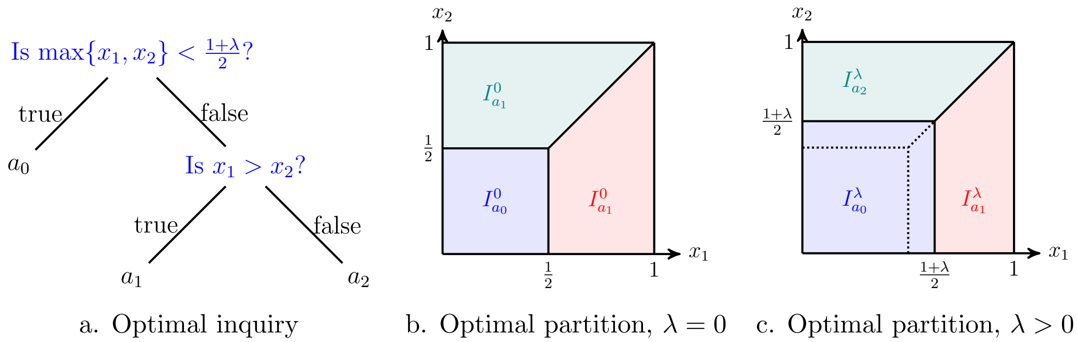


FIGURE 2. Optimal inquiry and optimal information partition

comparing the utilities of the actions net of the costs: $u(a_0, x) - \lambda$, $u(a_1, x) - 2\lambda$, and $u(a_2, x) - 2\lambda$. The optimal information partition $I^\lambda = \{I^\lambda_{a_0}, I^\lambda_{a_1}, I^\lambda_{a_2}\}$ is shown in Figure 2c.

As seen from the comparison of Figures 2b and 2c, as λ increases, the doctor optimally expands the area $I^\lambda_{a_0}$ which leads to a quicker action, namely, a_0, and shrinks the areas where slower actions, a_1 and a_2, are made. Moreover, if λ were to increase further, eventually the doctor would ask only a single question leading to a_0 or one of the other actions, and drop the third action from consideration entirely. This means that, as costs rise, the doctor optimally asks fewer questions (in expectation). We refer to this effect as a reduction of the *attention span.* This also means that the doctor optimally chooses the ex-ante most likely correct action a_0 with an even greater probability, thus actively seeking confirmation of her ex-ante most plausible hypothesis. We refer to this effect as the *confirmation bias.* Misdiagnosis occurs when the state is in $I^\lambda_{a_0} - I^0_{a_0}$ (Figure 2c, the area between the dotted and solid lines), as in this case the doctor sends the patient home when she should have referred them to a specialist or to emergency.

The result that the optimal inquiry under $\lambda > 0$ starts with the question about the ex-ante most likely correct action a_0 and leads to an expansion of $I^\lambda_{a_0}$, hence inducing a_0 with an even higher probability, derives from the well-known Huffman coding (Huffman, 1952). It is a lossless compression algorithm for discrete data which, when adapted to our context, determines the most efficient question tree and implies that more likely actions should be reached after fewer questions. We extend this logic to our setting with a continuum of states and show that it also leads to an endogenous information partition.

Our framework allow us to unify and provide a "rational" explanation to cognitive and behavioral biases discussed in Croskerry et al. (2013): "The principle biases for the physician who saw [the patient] in the clinic were framing, search satisficing and premature diagnostic closure." That is, the doctor identifies the most likely solution to the problem (in this example, a_0) based on her prior knowledge (*framing*) and then *searches* for evidence to confirm that, by prioritizing the confirmation of a_0 over other options and expanding the set of states where a_0 is chosen. When $x \in I^{\lambda}_{a_0} - I^{0}_{a_0}$, the doctor *prematurely closes* the diagnostic inquiry and sends the patient home, whereas ideally she should have continued the inquiry and decide whether an emergency treatment or a specialist referral are needed. Thus, in our framework, misdiagnosis emerges not because of the doctor's intrinsic biases but as an optimal response to costly information processing.

In what follows, we lay out our framework and present the general results of the effect of the cost of processing information on attention span and confirmation bias.

3. Model

3.1. **Primitives.** A decision maker (DM) needs to process information about an uncertain state of nature before taking an action. The DM's utility $u(a, x)$ depends on her action, $a \in A$, and the state, $x \in X$.[2] The set of actions A is finite and contains at least two actions. The set of states X is a convex subset of $\mathbb{R}^L$, $L \in \mathbb{N}$. State x is distributed according to a probability distribution G that is absolutely continuous and has full support on X. We will use notation $\mathbb{P}[\cdot]$ and $\mathbb{E}[\cdot]$ to denote the probability and expectation under G, respectively. Throughout the paper, we assume:

- (A_1) For all $a \in A$, $u(a, x)$ is continuous in x, and $\mathbb{E}[u(a, x)]$ is finite.
- (A_2) For all $a', a'' \in A$ and all $c \in \mathbb{R}$, the set $\{x \in X : u(a', x) - u(a'', x) = c\}$ has empty interior.

Assumption (A_1) is needed for the DM's optimization problem to be well defined. Assumption (A_2) means that the utility curves of any two actions are almost never

[2]Variable x can be interpreted as a profile of observables or signals with quantitative information about the true underlying state of nature (which may be ultimately unobservable) that the DM can ask questions about.

parallel to each other. Most utility functions that naturally emerge in applications satisfy (A_1) and (A_2).[3]

3.2. **Inquiries.** When confronted with a state x, the DM does not observe x directly. Instead, to obtain information about x, she relies on an inquiry: a series of true/false questions formulated as propositions. A proposition is a statement about x in the form "$x \in Y$" that can be either true or false. We denote the collections of Borel subsets of X by $\mathcal{B}(X)$, and identify a proposition with a set $Y \in \mathcal{B}(X)$. We say that proposition Y is *true* at x if $x \in Y$ and it is *false* if $x \notin Y$.

An *inquiry* $Q = \langle N, T, \sigma, \chi, d \rangle$ is a finite binary tree. Non-terminal nodes of the tree are associated with propositions, and terminal nodes are associated with actions. Specifically:

- a finite set N of nodes contains a root n^o and a nonempty set T of terminal nodes (note that the tree may consist of a single terminal node, i.e., $N = T = \{n^o\}$);
- each non-terminal node $n \in N - T$ is followed by exactly two edges labelled *true* and *false*;
- successor function σ assigns to each non-terminal node $n \in N - T$ and each edge $e = \{true, false\}$ a child $\sigma(n, e) \in N$ of node n following edge e;
- proposition mapping χ assigns to each non-terminal node $n \in N - T$ a proposition $\chi(n) \in \mathcal{B}(X)$;
- decision rule d assigns to each terminal node $t \in T$ an action $d_t \in A$.

We denote by $\mathcal{Q}_X$ the set of all possible inquiries given a set of states X.

Given a state of nature $x \in X$, an inquiry $Q = \langle N, T, \sigma, \chi, d \rangle$ begins with the proposition $\chi(n^o)$ at the root. At each non-terminal node $n \in N - T$, the inquiry asks whether it is true that $x \in \chi(n)$. If true, then the inquiry proceeds to the node $\sigma(n, true)$; otherwise, it proceeds to the node $\sigma(n, false)$. When a terminal node $t \in T$ is reached, the DM takes action d_t.

The inquiry transforms a quantitative assessment, say, "$x \geq r$", into a qualitative one, say, "yes" or "no", eventually leading to a qualitative recommendation of which action to choose. In our example in Section 2, the doctor's main concern is whether

[3] In fact, we impose (A_2) for notational convenience. It guarantees that the optimal inquiry is essentially unique as indifferences can only occur with probability zero. This assumption can be substantially relaxed at the expense of more cumbersome notation.

a given action is appropriate for the state of the patient; this ultimately is a qualitative assessment. In a world with limited cognitive and other resources for the exact assessment of the patient's condition in quantitative terms, our model captures the endogenous coarsening of the information partition through asking questions.

Formally, the inquiry *categorizes* states of nature into subsets through a series of questions. When arriving at any (terminal or non-terminal) node $n \in N$, the DM's information about the state is summarized by a subset of states, denoted by $I_n(Q)$. That is, given the answers to the questions in the previous nodes, the DM can infer that the true state belongs to $I_n(Q)$, recursively defined as follows. At the root, all states are possible, and hence $I_{n^o}(Q) = X$. Given a non-terminal node $n \in N - T$, define

$$I_{n^{true}}(Q) = I_n(Q) \cap \chi(n) \quad \text{and} \quad I_{n^{false}}(Q) = I_n(Q) \cap \big(X - \chi(n)\big), \tag{2}$$

where n^{true} and n^{false} are the successors of n after "true" and "false" answers to the proposition $\chi(n)$, respectively. For a terminal node t, $I_t(Q)$ consists of all states under which t is reached, and we call it a *category* of states induced by Q. Note that the collection of categories $\{I_t(Q) : t \in T\}$ forms a partition of X. It is the information partition at the end of the inquiry.

As zero probability events do not matter for payoffs, we adopt and use throughout the paper a measure-based notion of partition that disregards sets of measure zero under G. Specifically:

Definition 3.1. A collection of disjoined sets $\{X_1, X_2, ..., X_K\}$ is a *partition of* X if $\mathbb{P}(X_k) > 0$ for each k, and $\sum_k \mathbb{P}(X_k) = \mathbb{P}(X) = 1$.

3.3. **Payoffs.** We assume that asking questions is costly. Let the DM's cost of any single question be $\lambda > 0$. Given an inquiry Q, let $\ell_t(Q)$ be the length of the path from n^o to t in the tree, that is, $\ell_t(Q)$ is the number of questions asked to reach terminal node t. Then, the cost of inquiry at terminal node t is equal to $\lambda \ell_t(Q)$.

We can now formulate the DM's optimization problem. Given an inquiry Q and a state x, if the inquiry reaches the terminal node t for the given x, the DM's ex-post payoff net of the cost is

$$u(d_t, x) - \lambda \ell_t(Q).$$

Because each terminal node $t \in T$ is reached whenever the state x is in $I_t(Q)$, the DM's ex ante expected utility is given by

$$W(Q;\lambda) = \sum_{t\in T} \int_{x\in I_t(Q)} \Big(u(d_t,x) - \lambda \ell_t(Q)\Big) G(\mathrm{d}x). \tag{3}$$

An inquiry Q is *optimal* if it solves

$$\max_{Q\in\mathcal{Q}_X} W(Q;\lambda). \tag{4}$$

3.4. **Outcomes.** Here we show that it suffices to describe an optimal inquiry by its payoff-relevant outcome based on two observations. First, if an inquiry is optimal, then every node must be reached with positive probability. Indeed, if a node is never reached, then the question leading to this node is redundant. Second, an optimal inquiry cannot induce the same action in two terminal nodes. For example, in Inquiry C (Figure 1), action a_0 is chosen after a single question if $x_1 + x_2 \leq 1/2$ and after three questions if $x_1 + x_2 > 1/2$ and $\max\{x_1, x_2\} < 1/2$. For such an inquiry, we can always construct another inquiry that leads to the same final information partition with fewer questions. In this example, Inquiry B (Figure 1), in fact, does the job.

As a result, each terminal node corresponds to a unique action in A, and from now on we identify terminal nodes with actions they induce. Let $D(Q)$ be the set actions induced under Q. We refer to $D(Q)$ as the *consideration set.*[4] For each action $d \in D(Q)$, let $\ell_d(Q)$ denote the length of inquiry leading to the terminal node where d is chosen, and let $I_d(Q)$ denote the information set or the *category* induced by Q in that terminal node. Let $\ell(Q) = (\ell_d(Q))_{d\in D(Q)}$ and $I(Q) = (I_d(Q))_{d\in D(Q)}$. We will refer to the triple $(D(Q), \ell(Q), I(Q))$ as the *outcome* induced by Q. The pair $(D(Q), \ell(Q))$ captures the shape of the inquiry tree, and hence referred as the *form* of Q. The partition $I(Q)$ captures the informational content in terminal nodes of the inquiry tree, and hence referred as the *content* of Q.

An outcome (D, ℓ, I) captures all we need to know to evaluate the DM's expected payoff of an inquiry that leads to that outcome. Indeed, for any two different inquiries Q and Q' that implement the same outcome (D, ℓ, I), the expected utility in (4)

[4]We adopt this terminology from Masatlioglu et al. (2012). Our optimal inquiry satisfies a minimal rationality property for consideration set required by Masatlioglu et al. (2012): independence of irrelevant alternatives, or the "attention filter". It is defined as follows. Suppose that D is a strict subset of A. The attention filter property holds if, for any smaller action set $A' \subset A$ that contains D, the optimal consideration set is still D.

implies that they deliver the same ex ante payoff to the DM. Conversely, we can ask whether a given profile (D, ℓ, I) with $D \subset A$, $\ell = (\ell_d)_{d \in D} \in \mathbb{N}^{|D|}$ and I being a partition of X to $|D|$ categories is attainable as outcome of an inquiry. If the answer is yes, then we say that the outcome (D, ℓ, I) is *implementable*. The following lemma characterizes implementable outcomes.

Lemma 3.1. *An outcome (D, ℓ, I) is implementable if and only if*

$$\sum_{d \in D} 2^{-\ell_d} = 1. \tag{5}$$

Equality (5) follows from the Kraft inequality (Kraft, 1949) in the information theory that characterizes the path lengths of binary trees.[5] For any given form (D, ℓ) that satisfies (5) and any given content I with $|D|$ categories, the proof of Lemma 3.1 constructs an inquiry with the corresponding outcome. This lemma also fully characterizes possible length configurations for a given consideration set. For example, with $|D| = 3$, only three length profiles satisfy equality (5), namely, $\ell = (1, 2, 2)$, $\ell = (2, 1, 2)$, and $\ell = (2, 2, 1)$. With $|D| = 4$, there are 13 length profiles that satisfy (5), namely, the uniform profile, $(2, 2, 2, 2)$, and 12 distinct permutations of the profile $(1, 2, 3, 3)$.

4. Optimal Inquiries and behavioural implications

4.1. **Optimal Inquiries.** We have shown that an inquiry can be summarized by its outcome (D, ℓ, I). Moreover, Lemma 3.1 shows that the information partition I does not affect whether or not an outcome profile if implementable. This characterization allows us to solve the optimal inquiry problem in two stages. We first fix an arbitrary form (D, ℓ) that satisfies (5), and solve for the optimal content $I = I^*(D, \ell)$. Then, we maximize over all possible forms (D, ℓ).

In the first stage, taking (D, ℓ) as given, we find an information partition $I^*(D, \ell)$ that maximizes the DM's expected utility. Specifically, let $I^*(D, \ell) = \{I^*_d(D, \ell)\}_{d \in D}$, where

$$I^*_d(D, \ell) = \Big\{x \in X : u(d, x) - \lambda \ell_d > \max_{a \in D - \{d\}} u(a, x) - \lambda \ell_a \Big\}. \tag{6}$$

[5] Here we have equality instead of inequality because in our inquiry trees each non-terminal node has precisely two outgoing branches.

That is, for each action $d \in D$, $I^*_d(D,\ell)$ is the set of states where d is the unique best-response action among all actions in D when the DM takes into account the cost of inquiry associated with each action.[6] Thus, the DM weakly prefers $I^*(D,\ell)$ to any content I. Formally, using the notation $W(D,\ell,I;\lambda)$ for the DM's expected utlity from an outcome (D,ℓ,I), we have

$$\begin{aligned} W(D,\ell,I^*(D,\ell);\lambda) &= \int_{x\in X}\left(\max_{d\in D}(u(d,x)-\lambda\ell_d)\right)G(\mathrm{d}x) \\ &\geq \sum_{d\in D}\int_{x\in I_d}\Big(u(d,x)-\lambda\ell_d\Big)G(\mathrm{d}x) = W(D,\ell,I;\lambda). \end{aligned} \tag{7}$$

This observation is a key to solving the optimal inquiry problem. As mentioned earlier, the outcomes of an inquiry include both a continuous element I and discrete element (D,ℓ). The optimal content I is determined by the form (D,ℓ) through $I^*(D,\ell)$, and problem (4) can now be reduced to the choice of the form, (D,ℓ).[7]

Let $\mathcal{F}^*$ be the set of all forms (D,ℓ) with nonempty $D \subseteq A$ and ℓ satisfying (5). We have the following theorem.

Theorem 4.1. *An inquiry Q is optimal if and only if $(D(Q),\ell(Q))$ solves*

$$\max_{(D,\ell)\in\mathcal{F}^*}\int_{x\in X}\left(\max_{d\in D}(u(d,x)-\lambda\ell_d)\right)G(\mathrm{d}x), \tag{8}$$

and $I(Q)$ is identical to $I^*(D(Q),\ell(Q))$ up to a measure zero set.

Since $\mathcal{F}^*$ is a finite set, Theorem 4.1 implies that an optimal inquiry exists. We also remark here that optimal inquiry satisfies dynamic consistency in that the DM could not benefit from altering any sub-inquiry in the middle of the process if she was given the opportunity. As a result, it makes no difference whether the DM commits to an optimal inquiry ex ante or she is free to update her strategy at any interim stage.

[6] Note that $I^*(D,\ell)$ is a partition of X according to Definition 3.1, because, by assumption (A_2), the set $\left(X - \bigcup_{d\in D} I^*_d(D,\ell)\right)$ has measure zero.

[7] Following the methodology of Bloedel and Zhong (2024), it is possible to solve (4) by first calculating the cost of an induced information partition (by determining the optimal tree leading to this partition using Huffman coding and then calculating the expected cost for this inquiry tree), and then optimizing over all information partitions. However, this approach would be impractical since, unlike Bloedel and Zhong (2024), we have a continuum of states and thus a continuum of information partitions with discontinuous cost structures resulted from Huffman coding. In contrast, our approach allows us to reduce (4) to a finite optimization problem.

The set $\mathcal{F}^*$ of possible forms can be large for large consideration sets. Here we give an auxiliary optimality condition that can substantially reduce the number of candidate solutions. For a given content $\{I_d\}_{d\in D}$, this optimality condition states that ℓ must minimize the average length with respect to the probability distribution $(\mathbb{P}(I_d))_{d\in D}$ subject to the constraint (5). This is a well-known problem in information theory, and the solution is described by the algorithm called *Huffman coding*.

To illustrate this, consider the example in Section 2. Given the consideration set $D = \{a_0, a_1, a_2\}$, let a corresponding partition $\{I_{a_0}, I_{a_1}, I_{a_2}\}$ be given. If $\mathbb{P}(I_{a_0}) > \max\{\mathbb{P}(I_{a_1}), \mathbb{P}(I_{a_2})\}$, then the algorithm requires $\ell_{a_0} = 1$ and $\ell_{a_1} = \ell_{a_2} = 2$, that is, the shortest length is associated with the highest probability to minimize the overall expected length. The generalization to arbitrary D follows the same principle. We refer to Cover and Thomas (2006, Section 5.6) for formal details of the Huffman coding algorithm. We have the following theorem.

Theorem 4.2. *If (D, ℓ, I) is an optimal outcome, then:*

(a) ℓ is obtained from the Huffman coding w.r.t. the distribution $(\mathbb{P}(I_d))_{d\in D}$;
(b) for all $d, d' \in D$, if $\mathbb{P}(I_d) > \mathbb{P}(I_{d'})$, then $\ell_d \leq \ell_{d'}$.

We claimed in Section 2 that, as long as $\mathbb{P}(I^0_{a_0}) > \max\{\mathbb{P}(I^0_{a_1}), \mathbb{P}(I^0_{a_2})\}$, the optimal length profile is to have $\ell_{a_0} = 1$, that is, to inquire about a_0 first for a range of relatively small λ's. This result immediately follows from Theorem 4.2 and the fact that the boundaries of the optimal categories determined by (6) are continuous in λ. As a result, $\mathbb{P}(I^\lambda_{a_0}) > \max\{\mathbb{P}(I^\lambda_{a_1}), \mathbb{P}(I^\lambda_{a_2})\}$ holds for small λ's (see Figure 2c for illustration), and Huffman coding demands a_0 to be inquired about first.

4.2. **Attention Span.** Our model of inquiry can be interpreted as an attention strategy, whereby the DM focuses on various decisions during her inquiry process. With this interpretation, a natural question is then how the cost λ affects the DM's attention span, defined as how long she would concentrate on the task of gathering information before taking an action. Formally, we measure *attention span* in our framework as the expected inquiry length given by

$$\bar{\ell}(D, \ell, I) = \sum_{d\in D} \ell_d \mathbb{P}(I_d). \tag{9}$$

Importantly for our purpose, it captures whether there is a lot of probability weight on a few actions with short inquiry length, or whether this weight is more spread out among many actions. A smaller $\bar{\ell}(D, \ell, I)$ means a shorter attention span. In the extreme, the DM has no attention span at all when she chooses a single action without asking any questions, in which case we have $\bar{\ell}(D, \ell, I) = 0$. The opposite extreme occurs when the lengths are mostly equal and the probabilities are spread out. A special case of this is when each action in D is reached after exactly the same number of questions, referred to as the *uniform* length profile. Note that this can only happen if $|D| = 2^k$ for some $k \in \mathbb{N}$. Formally:

Definition 4.1. Given an outcome (D, ℓ, I), we say that length profile ℓ is *uniform* if it assigns the same length to all actions in D, so $\ell_d = \ell_{d'}$ for all $d, d' \in D$.

The next theorem shows that higher cost always shortens the attention span, and strictly so as long as the optimal inquiry length is not uniform.

Theorem 4.3. *Let $\lambda_1 < \lambda_2$ and let $(D^{\lambda_j}, \ell^{\lambda_j}, I^{\lambda_j})$ be an optimal outcome under λ_j, $j = 1, 2$. Then, $\bar{\ell}(D^{\lambda_1}, \ell^{\lambda_1}, I^{\lambda_1}) \geq \bar{\ell}(D^{\lambda_2}, \ell^{\lambda_2}, I^{\lambda_2})$. Moreover, this inequality is strict, unless $|D^{\lambda_1}| = |D^{\lambda_2}|$ and both ℓ^{λ_1} and ℓ^{λ_2} are uniform.*

The intuition for Theorem 4.3 is based on the following trade-off that the optimal inquiry resolves. On the one hand, to achieve a high (expected) utility from actions, it needs to minimize the mismatch between the sets of states where this action is chosen and where it is ex post optimal. On the other hand, it needs to minimize the expected length of inquiry. As the cost increases, the latter motive becomes more important, and optimal inquiry shifts probabilities toward actions with shorter inquiries at the expense of more mismatches.

This preference for shorter inquiries generates a "bias" if we compare the information partition thus generated to the one that would be optimally used under zero cost. This bias is generated by the motive to decrease the expected inquiry length, and this can be achieved by adjusting the inquiry either through the form or through the content. The form affects the extensive margin, and the DM can simply drop certain actions from the consideration set and in this way the overall inquiry length may be reduced. We have the following result regarding the form.

Proposition 4.1. *Let*

$$A^* = \{a \in A: \; u(a,x) > \max_{a' \in A - \{a\}} u(a',x) \text{ for some } x \in X\}.$$

There exist two thresholds $\lambda_2 > \lambda_1 > 0$ such that for all $\lambda < \lambda_1$, the optimal consideration set is $D = A^*$; and for all $\lambda > \lambda_2$, the optimal consideration set is a singleton, $|D| = 1$. Moreover, if actions a and a' are such that $\sup_{x \in X} |u(a',x) - u(a'',x)| < \lambda$, then at most one of them will be in the optimal consideration set.

4.3. **Confirmation Bias.** Taking the form of the inquiry as given, we now consider how the content is affected by the cost of inquiry. We will show that the DM optimally expands the categories associated with the more likely actions, relative to the zero-cost benchmark. This can be interpreted as the DM searching for evidence to confirm the desirability of the actions in D that are most likely to be optimal. We call this effect *confirmation bias*.

To define confirmation bias, let us consider the zero-cost case as a benchmark, and compare the probability that the most likely actions are taken under the optimal inquiry with and without the cost, conditional on each state. To do so, we first rank the actions according to their likelihood under the optimal inquiry. Let Q be an inquiry, let (D, ℓ, I) be the outcome induced by Q, and let $K = |D|$. We order the actions in D according to how likely they are chosen under inquiry Q, so $D = \{d_k\}_{k=1}^K$, such that

$$\begin{gathered} \mathbb{P}(I_{d_1}) \geq \mathbb{P}(I_{d_2}) \geq ... \geq \mathbb{P}(I_{d_K}), \\ \text{with a tie-breaking rule } \; \mathbb{P}(I_{d_k}) = \mathbb{P}(I_{d_{k+1}}) \implies \ell_{d_k} \leq \ell_{d_{k+1}}. \end{gathered} \tag{10}$$

For each $k = 1, ..., K-1$ and each $x \in X$, let $p_Q(d_1, ..., d_k|x)$ be the probability that an action in $\{d_1, ..., d_k\}$ (i.e., one of k most likely actions) is selected under an inquiry Q conditional on state x.[8] Also, let $p_0(d_1, ..., d_k|x)$ be the probability that an action in $\{d_1, ..., d_k\}$ is optimally selected under the zero-cost benchmark conditional on x.

Definition 4.2. An inquiry Q with outcome (D, ℓ, I) has *confirmation bias* if for each order $(d_k)_{k=1}^K$ that satisfies (10), for all $k = 1, 2, ..., K-1$ and almost all $x \in X$,

$$p_Q(d_1, ..., d_k|x) \geq p_0(d_1, ..., d_k|x). \tag{11}$$

[8] For each state, the inquiry prescribes a path through the tree to a terminal node associated with an action, so $p_Q(d_1, ..., d_k|x)$ is well defined.

It has *strict confirmation bias* if, in addition, there exist $k \in \{1, ..., K-1\}$ and a nonempty open set $X' \subset X$ such that (11) holds strictly for all $x \in X'$.

In words, the DM has confirmation bias if, for each $k = 1, ..., K-1$, she confirms to k most likely actions: no matter what states realize, she chooses one of these actions with a greater probability, as compared to what she would have done without the cost. The difference $p_Q(d_1, ..., d_k|x) - p_0(d_1, ..., d_k|x)$ is the probability of making an erroneous decision in state x by choosing one of k most likely actions when the zero-cost benchmark prescribes otherwise. In any subset of states X' where this difference is positive, the DM is biased towards the initially most likely actions and may choose an action that is suboptimal from the perspective of the DM who has no cost of information processing.

This definition formalizes the notion of confirmation bias usually adopted in psychology. According to Nickerson (1998), "it refers usually to unwitting selectivity in the acquisition and use of evidence," which he believes is also the definition used by general psychologists. Our definition has the advantage of formally defining both confirming sets and biases: the actions the DM confirms to are the most likely ones in her optimal strategy, and the errors are defined against the zero-cost benchmark.

Theorem 4.4. *For every $\lambda > 0$, every optimal inquiry has confirmation bias. Moreover, an optimal inquiry has strict confirmation bias if and only if its length profile is not uniform.*

Confirmation bias of the optimal inquiry emerges for the following reason. From Theorem 4.2(b) we know that the most likely actions are associated with shorter inquiry lengths. At the same time, given the optimal length profile, the optimal information partition given by (6) has the feature that actions associated with shorter inquiry lengths will be chosen on a larger set of states relative to the zero-cost benchmark. These two factors reinforce each other and give rise to confirmation bias.

Note that, because every inquiry Q induces an information partition I, it is easy to see that $p_Q(d_1, ..., d_k|x)$ is always either 0 or 1. Moreover, the unconditional probability of choosing an action in $\{d_1, ..., d_k\}$ is the same as the probability that the state belongs to $I_{d_1} \cup ... \cup I_{d_k}$, so

$$\int_X p_Q(d_1, ..., d_k|x) G(\mathrm{d}x) = \sum_{d \in \{d_1, ..., d_k\}} \mathbb{P}(I_d).$$

Denoting by $I^0 = (I_d^0)_{d\in A}$ the optimal information partition under the zero-cost benchmark (i.e., I_d^0 is given by (6) with $\lambda = 0$ for each $d \in A$), Theorem 4.4 implies that the DM optimally expands the categories associated with the more likely actions, relative to the zero-cost benchmark. This can be interpreted as the DM searching for evidence to confirm the desirability of the actions in D that are most likely to be optimal. A related implication of Theorem 4.4 is that the probability distribution over $(d_k)_{k=1}^K$ induced by the optimal inquiry first order stochastically dominates that induced by the zero-cost benchmark:

$$\sum_{d\in\{d_1,...,d_k\}} \mathbb{P}(I_d) \geq \sum_{d\in\{d_1,...,d_k\}} \mathbb{P}(I_d^0), \quad \text{for each } k = 1, ..., K-1, \tag{12}$$

and the inequality is strict for some k if the confirmation bias is strict.

We have defined confirmation bias as a comparison against the benchmark case with zero cost. Now we show that this bias is increasing in the cost, in the sense that the probability of choosing one of k ex ante most likely actions increases state by state as λ increases. As before, in doing so we keep an optimal form (D, ℓ) constant.

Definition 4.3. Suppose that the form (D, ℓ) associated with an optimal inquiry Q^λ remains constant for some interval of costs, $\lambda \in [\lambda_1, \lambda_2]$. We say that confirmation bias is *increasing over* $[\lambda_1, \lambda_2]$ if for each order $(d_k)_{k=1}^K$ that satisfies (10), for all $k = 1, ..., K-1$ and almost all $x \in X$,

$$p_{Q^\lambda}(d_1, ..., d_k|x) \geq p_{Q^{\lambda'}}(d_1, ..., d_k|x)) \text{ for all } \lambda, \lambda' \text{ with } \lambda_1 \leq \lambda < \lambda' \leq \lambda_2. \tag{13}$$

Moreover, confirmation bias is *strictly increasing over* $[\lambda_1, \lambda_2]$ if there exist $k \in \{1, ..., K-1\}$ and a nonempty open set $X' \subset X$ such that (13) holds strictly for all $x \in X'$.

Denoting by I^λ the content induced by Q^λ, we obtain that increasing confirmation bias implies that, for each $k = 1, ..., K-1$, the set $I_1^\lambda \cup ... \cup I_k^\lambda$ expands as λ increases. Hence, the likelihood of choosing one of k most likely actions increases with λ, and this increase becomes strict under strict increase of confirmation bias. We have the following result.

Proposition 4.2. *Let $\lambda > 0$.*

(a) There exists an optimal inquiry Q^λ whose form (D, ℓ) is constant over an interval $[\lambda_1, \lambda_2]$ that contains λ.

(b) For every optimal inquiry Q^λ whose form (D, ℓ) is constant over some interval $[\lambda_1, \lambda_2]$ that contains λ, the confirmation bias of Q^λ is increasing over $[\lambda_1, \lambda_2]$. Moreover, it is strictly increasing if and only if ℓ is not uniform.

Proposition 4.2 shows that as the cost rises, the DM would optimally make more "errors" and is biased more toward the most likely actions. This result is related to Theorem 4.3, which states that as the cost rises, the DM optimally shortens the average inquiry length. Proposition 4.2 shows that the way to optimally achieve that is by decreasing the accuracy of her choices in favor of most likely actions, which are also actions associated with the shortest inquiry lengths.

Recall our example in Section 2. As mentioned, under the assumption that $\mathbb{P}(I^0_{a_0}) > \max\{\mathbb{P}(I^0_{a_1}), \mathbb{P}(I^0_{a_2})\}$, the optimal inquiry prioritises action a_0 for a range of λ's that are not too large, and the doctor first inquire about the optimality of prescription to treat the symptom before investigating the optimality of referrals or emergency treatment. Theorem 4.4 and Proposition 4.2 then imply that the doctor would optimally expand the category of states under which a_0 is optimal, as shown in Figure 2c. Moreover, since we show that confirmation bias holds in general in our model, the details of the example, such as the underlying distribution of states and the detailed specification of the utility functions, do not matter for the main conclusion.

This confirmation bias toward the most likely option provides a potential economic mechanism for the behavioral biases highlighted by Croskerry et al. (2013), as the doctor designs the optimal inquiry that prioritises the most likely action (*framing*) with an expands the category for that action to be appropriate (*search satisficing*). Our model also has another novel prediction: the action that the DM has confirmation bias toward is also investigated first. Thus, we can then define "premature closure" of the diagnosis, which is listed as one of the biases in Croskerry et al. (2013). Indeed, when states in $I^\lambda_{a_0} - I^0_{a_0}$ occur, it would be optimal to inquire further instead of closing the inquiry, and this may be regarded as a bias from the outside observer's perspective. However, our theory also shows that there is nothing irrational per se but what is important is to take the cost of inquiry into account.

5. Case studies

To illustrate potential applications of our model, we offer two further case studies. In the first case we consider the phenomenon of "tunnel vision" in criminal investigations, where the authorities adopt inquiries with a focus on specific suspects that lead to wrongful convictions. In the second case we consider endogenous salience in a procurement setting in which a procurer may prioritise some suppliers over others in a group of ex ante symmetric suppliers.

5.1. **Wrongful Conviction.**

Gould and Leo (2010) review the literature on the extent and factors leading into wrongful convictions and believe that it is the process and factors affecting the process that are important. In their words, "...it is better to understand the sources of wrongful convictions not so much as dichotomous causes—a witness correctly or incorrectly identified the defendant and the identification directly led the jury to convict—but as contributing factors in a path analysis that might have been broken at some point before conviction." Among the leading factors the article identifies, we are interested in "tunnel vision", which is described in Gould and Leo (2010) as "the more law enforcement practitioners become convinced of a conclusion—in this case, a suspect's guilt—the less likely they are to consider alternative scenarios that conflict with this conclusion."

We illustrate this tunnel vision with the following example. Suppose that there are two suspects, A and B, and one of them is surely guilty. Given all the possible observables that the police can investigate, suppose that state $x \in [0, 1]$ represents the posterior belief that A is guilty. There are three actions: charge A, B, or neither, denoted by a_A, a_B, and $a_\varnothing$, respectively. Suppose that the police obtains utility θ_A if they charge A when A is guilty, θ_B if charge B when B is guilty, -1 if a wrong suspect (either A or B) is charged, and 0 if neither. Thus,

$$u(a_A, x) = \theta_A x + (-1)(1 - x), \quad u(a_B, x) = \theta_B(1 - x) + (-1)x, \quad u(a_\varnothing, x) = 0.$$

Assume that x is uniformly distributed on $[0, 1]$ and that $1 > \theta_A > \theta_B \geq 0.52$, so that ex ante the most likely suspect is A, and no action is dominated. We depict the optimal categories in Figure 3, where $\bar{x}_1$ and $\bar{x}_2$ are thresholds between the categories under optimal inquiry, and $\bar{x}_1^0$ and $\bar{x}_2^0$ are thresholds under zero-cost benchmark. As indicated in the figure, a positive λ leads to an expansion of the category for charging

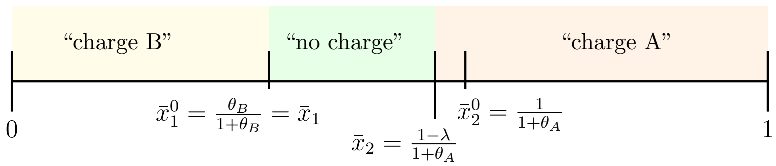


FIGURE 3. Optimal categories for the police inquiry

A, who is the prime suspect. In other words, the confirmation bias leads the police to lower the threshold of evidence needed to charge suspect A relative to the benchmark case without cost. Specifically, on the interval of states $(\bar{x}_2, \bar{x}_2^0)$, the police makes an error by charging A when they should have let them go. Moreover, this interval expands with λ.

Furthermore, if the cost were to rise even higher, the police would have found it optimal to drop the no-charge option out of the consideration set altogether. Thus, if we define "type-I" error as the situation where the police does not charge anyone, and "type-II" error as the situation where police charges the wrong suspect, then the optimal inquiry is always biased toward the type-II error. That is, it always has higher type-II error than the no-cost benchmark.

The above effects may provide an explanation of the tunnel vision. The police under pressure to end the investigation optimally focuses on the prime suspect and is willing to charge the prime suspect even with relatively weak evidence. Moreover, if the pressure is high enough, the police optimally focuses on fewer options that what they would have considered under less pressure.

5.2. **Biased Procurement.** Here we consider an application of our model to a procurement setting. Consider an organisation (procurer) that decides on a procurement order from one of several potential suppliers, say, a government agency requires a large order of supply for a specific drug.[9] The values of the products/services provided by the suppliers are uncertain and are independently distributed across the suppliers. The procurer does not directly observe the values and has to rely on costly expertise to evaluate and compare them. In the example of drug procurement, while the results of clinical trials of drug quality or efficacy are readily available, they are

[9]E.g., the U.S. Department of Veterans Affairs purchased over \$9 billion in medical supplies in 2023 (see the 2023 U.S. Federal Market Procurement Report, p. 72, available at https://thecgp.org/images/2024/06/23-Market-Report-2.pdf.

conducted on different sample sizes using different methodologies, and require costly expertise for lengthy clinical trials reports to be read and interpreted.

Formally, suppose that there are L suppliers. Let $x_l \geq 0$ denote the value of supplier l's product to the procurer. Each x_l is independently distributed according to a distribution function G_l. Using the notations of our model, let $X = \mathbb{R}_+^L$. The procurer chooses one of L suppliers to contract with, so $A = \{a_1, ..., a_L\}$. Let

$$u(a_l, x) = x_l \text{ for all } l = 1, ..., L \text{ and all } x \in X.$$

We use this setting to show that the cost of expertise, modelled as cost of inquiry, may lead to"biased" selection in procurement, giving priority to some suppliers and deprioritizing or completely excluding others. This bias is most apparent when all the suppliers are ex ante symmetric (i.e., their values are identically distributed), and there is no a priori reason to treat them asymmetrically. Nevertheless, it follows from Theorem 4.3 that, as the cost increases, the attention span of the inquiry must shorten. This can be achieved in two ways. First, when it is too costly to compare all suppliers, the procurer may have a reduced consideration set where only a subset of suppliers are considered and their values are potentially compared. Second, within the consideration set, some suppliers can be endogenously prioritized so that they are selected after fewer questions in the inquiry process as compared to others. It then follows from Theorem 4.4 that this leads to confirmation bias: a prioritized supplier is selected in a larger set of states than any of those with a lower priority, despite being ex ante symmetric. This is not due to any conflict of interest but solely coming from the motive to save inquiry cost.

To illustrate the last point, suppose that there are three ex ante symmetric suppliers, say, Alice, Bakir, and Chen. The optimal inquiry needs to minimize the number of questions and this leads to an endogenous asymmetric treatment of the suppliers. Specifically, the affirmative answer to the initial question must lead to the selection of one of the suppliers, say, Alice. The distinction between Bakir and Chen is made only if the answer to the initial question is negative. As selecting Alice requires one question, while selecting Bakir or Chen requires two questions, the optimal inquiry leads the procurer to choose the supplier with the highest value net of the cost (as stated in Theorem 4.1): $x_1 - \lambda$, $x_2 - 2\lambda$, and $x_3 - 2\lambda$ for Alice, Bakir, and Chen, respectively. Thus, the procurer has an optimal "bias" towards Alice by setting a lower bar with the margin of λ for her to be selected. Of course, since the three suppliers are all symmetric, the choice to prioritise Alice is completely arbitrary and it is equally optimal to prioritise any other supplier. The point, however, is that any optimal inquiry will prioritise *some* supplier, and we refer to this effect as *endogenous salience*, which is demanded by the optimal inquiry.

Endogenous salience can occur even if the number of suppliers is 2^k for some $k \in \mathbb{N}$; thus, although it is possible to treat them symmetrically using the uniform length profile of inquiry, it may not be optimal to do so. For illustration, suppose that we now have a fourth supplier, Diana, in the pool, and that λ is not too high so that all four suppliers are considered. In this case, Lemma 3.1 allows for two types of length profiles: the uniform profile $\ell = (2, 2, 2, 2)$, and the asymmetric profile $\ell = (1, 2, 3, 3)$ and all its permutations.

Under the drug procurement example, the prior distribution of each x_l would reflect recovery/survival times.[10] Typical distributions to model survival times include exponential and log-logistics (e.g., Jackson et al., 2010; Ishak et al., 2013). For the exponential distribution we normalize the mean to unity, and for log-logistic distribution we normalize the mode to unity and consider the mean of $\pi/2$ and $4\pi/3\sqrt{3}$. In all these cases, for small λ's, the optimal inquiry treats all four suppliers symmetrically: it requires exactly two questions to select the supplier with the highest value, yielding the net payoff of $\max\{x_1, x_2, x_3, x_4\} - 2\lambda$. However, for a range of intermediate λ's,[11] the optimal inquiry prioritises one of the four suppliers, say, Alice again, and the affirmative answer to the first question would lead to her selection, but it requires two questions for, say, Bakir, and three for Chen and Diana. Accordingly, the bar for Alice's selection is lowered, by the margin of λ against Bakir and 2λ against Chen and Diana, yielding the net payoff of $\max\{x_1 - \lambda, x_2 - 2\lambda, x_3 - 3\lambda, x_4 - 3\lambda\}$.

The last example highlights that two procurers in completely identical procurement situations who only differ in their cost of processing information may adopt very different optimal inquiries. The procurer with a slightly higher cost would prioritise

[10]According to Ishak et al. (2013), "For chronic conditions and progressive diseases, health-economic models must provide insights over the lifetimes of individuals, which often exceed the span of available data from clinical trials."

[11]Specifically, for the exponential distribution, the range is $[0.52, 0.54]$; for the log-logistic with mean $\pi/2$ and $4\pi/3\sqrt{3}$, the range is $[0.56, 1.01]$ and $[0.83, 2.48]$, respectively.

one of the suppliers at the expense of all others, while the procurer with a slightly lower cost would treat them symmetrically.

Some observers claim that the decision-making processes in government procurements are potentially subject to biases related to cognitive factors, such as mental shortcuts and confirmation biases (see, for example, the BIS report on procurements of transportation projects in the UK[12]) due to the pervasiveness of their presence in human decision-making. However, different from these observers, our approach emphasizes that apparent biases can be an outcome of the optimal responses to the underlying friction in the decision-making process, and hence, to address the biases, it is important the recognize the endogenous mechanisms behind the behaviour.

6. Comparison with Rational Inattention Models

Our model shares a similar motivation to that for the rational inattention (RI) model (the standard formulation can be found in, e.g., Matějka and McKay, 2015, and Caplin et al., 2019), and aims at studying the implications of costly information processing. When the two models share the same primitives—a state space X, a finite set of actions A, and a utility function $u(a, x)$—their results can be compared according to the probabilities of induced choices. A standard RI model is a static model of information processing where the cost depends on the entropy reduction w.r.t. the DM's prior. In contrast, we consider a simple dynamic model of information processing where the cost is directly associated with the acts of asking questions and processing their answers, and thus it is independent of the DM's prior. This difference in the modeling approach leads to differences in the implications.

The first key difference is that we can use our model to formally define and explicitly study the attention span. Theorem 4.3 shows that a rise in the cost of inquiry leads to a reduction in the attention span. Since the attention span is an intrinsically dynamic concept, there is no direct counterpart of it in RI models. However, this result does have implications to choice probabilities. For illustration, recall our procurement example with several ex ante symmetric suppliers (Section 5.2). The RI model leads

[12]This report is conducted by the Behavioural Insight Team for the UK government, titled "An Exploration of Potential Behavioural Biases in Project Delivery in the Department for Transport," and can be found in https://assets.publishing.service.gov.uk/media/5a8203f640f0b62305b92065/exploration-of-behavioural-biases.pdf

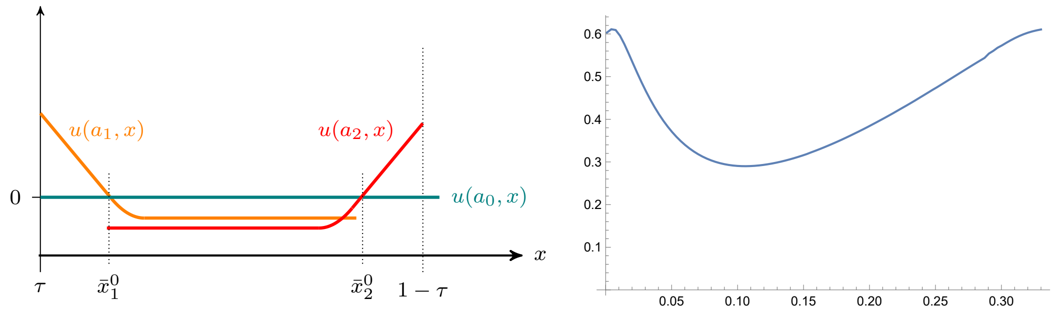
to the symmetric choice probabilities, that is, each supplier is contracted equally likely. In contrast, our model can generate endogenous asymmetric choice probabilities, as the procurer endogenously prioritizes some suppliers over others to save the cost of distinguishing them. Theorem 4.3 implies that this asymmetric treatment can be optimal even when the uniform treatment is available, as in our example with four suppliers. Indeed, at $\lambda = 0$, the value from either the asymmetric profile $\ell = (1,2,3,3)$ or the uniform one $\ell = (2,2,2,2)$ is the same. By the Envelope Theorem, the derivative of the value with respect to λ is the optimal average length, which strictly increases under the asymmetric profile by Theorem 4.3 but it stays constant for the uniform one, so for some positive λ the asymmetric profile becomes better than the uniform one.

The second key difference is that outcomes according to the RI model do not necessarily exhibit confirmation bias.[13] To show this, consider the following example in the setting of options trade. Suppose that the future price ρ of an asset depends on the today's economic fundamentals, summarized by x, and a noise, summarized by ε, where $\rho = x + \varepsilon$. The DM initially observes neither x nor ε. She can inquire about x, but no information about ε available, besides its prior distribution.

Two financial instruments are available to the DM: (i) a put option on the asset with a strike price ρ_1, which can be purchased at a premium α_1, and (ii) a call option on the asset with a strike price ρ_2, which can be purchased at a premium α_2. The DM chooses one of three actions: not invest (action a_0), buy the put option (action a_1), or buy the call option (action a_2). The DM is risk neutral and evaluates her choices according to their expected net financial gains, with no discounting (for simplicity).

For this illustration, let τ and θ be parameters satisfying $0 < 2\tau < \theta < 1/2$, and assume that $\rho_1 = \theta$ and $\rho_2 = 1 - \theta$, and that x is uniformly distributed on $[\tau, 1-\tau]$ and hence $X = [\tau, 1-\tau]$ and $G(x) = (x-\tau)/(1-2\tau)$, and ε is uniformly distributed on $[-\tau, \tau]$. Then, the expected net financial gain from each action, conditional on each $x \in X$, is given by

[13] Note that our Definition 4.2 is applicable to the RI model as well.



a. The utility functions

b. Optimal choice probability for a_0

FIGURE 4. The RI solution in the setting of option choices

$$u(a_0, x) = 0 \text{ for all } x \in X,$$

$$u(a_1, x) = \begin{cases} \theta - x - \alpha_2, & \text{if } x < \theta - \tau \\ \frac{(\theta+\tau-x)^2}{4\tau} - \alpha_2, & \text{if } x \in [\theta - \tau, \theta + \tau] \\ -\alpha_2, & \text{if } x > \theta + \tau. \end{cases},$$

$$u(a_2, x) = \begin{cases} -\alpha_1, & \text{if } x < 1 - \theta - \tau \\ \frac{(x-(1-\theta-\tau))^2}{4\tau} - \alpha_1, & \text{if } x \in [1 - \theta - \tau, 1 - \theta + \tau] \\ x - (1-\theta) - \alpha_1, & \text{if } x > 1 - \theta + \tau. \end{cases}.$$

See Figure 4a for a graphic depiction of the utilities.[14] Consider the following values of the parameters: $\tau = 0.04$, $\theta = 0.25$, $\alpha_1 = 0.025$ and $\alpha_2 = 0.03$. Both options generate a negative expected payoff under the prior. Hence, the DM would invest only if she is sufficiently convinced that x is far from the middle. If there is no cost of information, in both the RI model and ours, the DM would buy the put option for $x < \bar{x}_1^0 = 0.227$ and buy the call option for $x > \bar{x}_2^0 = 0.78$. Accordingly, under the zero-cost benchmark, the ex ante probability of not investing, denoted by p_0^0, is given by $p_0^0 = 0.6$.

Following Matějka and McKay (2015), in the RI model with cost parameter $\lambda > 0$, the DM chooses an information structure, modelled a joint distribution, $\Phi(s, x)$, where

[14]Note that $u(a, x)$ does not satisfy our assumption (A_2). However, our argument is still valid, as $u(a, x)$ can be slightly perturbed, to make sure that (A_2) holds, without any qualitative consequences to the results.

$s \in [0,1]$ is the signal to be received by the DM, and an action plan, $a(s)$, to maximize

$$\int_{x\in X}\int_s u(a(s),x)\Phi(\mathrm{d}s|x)G(\mathrm{d}x) - \lambda\Big(H(G) - \mathbb{E}_s[H(\Phi(\cdot|s))]\Big)$$
$$\text{subject to } \int_s \Phi(\mathrm{d}s,x) = G(x) \text{ for all } x \in X,$$

where H denotes the entropy of the distribution, and $\Phi(\cdot|x)$ and $\Phi(\cdot|s)$ denote the conditional distributions on x and s, respectively. The constraint means that the marginal distribution over x is equal to the prior distribution $G(x)$. In terms of choice outcomes, Lemma 2 in Matějka and McKay (2015) implies that the ex ante choice probabilities, with p_k^λ denoting the probability of choosing action a_k from the optimal solution, solve

$$\max_{(p_0^\lambda,p_1^\lambda,p_2^\lambda)\in\Delta_3}\int_{x=0}^1 \log\left[\sum_{k=0}^2 p_k^\lambda e^{\frac{u(a_k,x)}{\lambda}}\right]G(\mathrm{d}x).$$

We depict the optimal probability of not investing, p_0^λ, in Figure 4b, as λ changes (for the range under which all three actions are considered in the RI model). Observe that there is an interval of costs where p_0^λ decreases as λ goes up.[15] Thus, for a range of relatively small costs, the RI model predicts that the DM (who would not invest absent any information) may be *more likely* to invest as the cost of information increases. It also means that the optimal solution in the RI does not have the property of confirmation bias, according to our definition. This is because confirmation bias implies that the probability of choosing the most likely action must increase with λ. This does not hold in the RI model, as for $\lambda \in (0.01, 0.05)$ action a_0 is the most likely choice but its probability p_0^λ is decreasing, as shown in Figure 4b.

In contrast, under our model, as λ increases, the DM expands the category for the action of not investing, a_0, and shrinks the categories for the two investment actions, as depicted in Figure 5. For higher λ's, it is optimal for the DM to consider $\{a_0\}$ only and hence stay away from the option market altogether. In other words, according to our model, agents with higher cost of processing financial information will expand the region of states where they would deem either type of investment not worthwhile.

[15]The range for this to happen is $\lambda \in (0.01, 0.1)$; moreover, for $\lambda < 0.05$, a_0 has the highest choice probability.

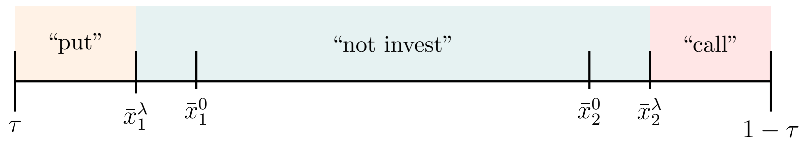


FIGURE 5. Optimal categories for option choices

Only those with the expertise, modelled as very small λ's, would invest when it is in fact profitable to do so.[16]

APPENDIX A. PROOFS

A.1. **Proof of Lemma 3.1.** The necessity follows from Theorem 5.2.1 in Cover and Thomas (2006). We have an equality instead of inequality, because in our inquiry tree every non-terminal node has exactly two branches.

To prove the sufficiency, let $K \in \{2, ..., |A|\}$, let $D = \{d_1, ..., d_K\} \subset A$, let $I = \{I_k\}_{k=1}^K$ be a partition of X, and let $\ell = (\ell_1, ..., \ell_K) \in \mathbb{N}^K$ be a length profile such that (5) holds. We construct an inquiry $Q = \langle N, T, \sigma, \chi, d \rangle$ with $T = D$ that satisfies

$$I_{d_k}(Q) = I_k \quad \text{and} \quad \ell_{d_k}(Q) = \ell_k \quad \text{for all } k = 1, ..., K. \tag{14}$$

By Theorem 5.2.1 in Cover and Thomas (2006) and equation (5), there exists a finite binary tree with a set of nodes N and a successor relation over N, with K terminal nodes labeled $t_1, ..., t_K$, such that, for each $k = 1, ..., K$, the length of the path from the root to each terminal node t_k is exactly ℓ_k. For each nonterminal node $n \in N - T$, let us associate two edges leading out of n with $true$ and $false$, and define the map σ so that $\sigma(n, true) = n^{true}$ if $n \rightsquigarrow n^{true}$ along the edge labelled $true$ and $\sigma(n, false) = n^{false}$ if $n \rightsquigarrow n^{false}$ along the edge labelled $false$. Let decision rule d be the identity mapping that associates each terminal node t_k with action d_k for each $k = 1, ..., K$.

It remains to construct a proposition mapping χ that yields the partition I in the terminal nodes. First, we associate each node in N with a set, $I_n(Q)$, as follows. For each $k = 1, ..., K$, let $I_{t_k}(Q) = I_k$. Then, by backward induction, for each nonterminal node $n \in N - T$, let $I_n(Q) = I_{\sigma(n,true)}(Q) \cup I_{\sigma(n,false)}(Q)$. This implies that $I_{n^o}(Q)$ at the root n^o is equal to X up to a measure-zero set, since $\{I_k\}_{k=1}^K$ is a partition, and

[16]Note that the cost parameter λ has different meanings in the RI model and in ours, and hence we emphasize the qualitative difference.

we can place the measure-zero set anywhere in the propositions used along the tree anywhere without affecting the payoffs.

Finally, define a proposition map χ as follows. For each nonterminal node $n \in N - T$, let $\chi(n) = I_{\sigma(n,true)}(Q)$. By induction from the root of the tree, it is straightforward to verify that χ satisfies (2), so, for each $n \in N$, $I_n(Q)$ is indeed the information set induced by Q at node n. □

A.2. **Proof of Theorem 4.1.** The result that an inquiry is optimal if and only if its form solves (8) is immediate by equations (3) and (6), inequality (7), and the definition of $\mathcal{F}^*$ together with Lemma 3.1.

It remains to show that if (D, ℓ, I) is an optimal outcome, then I is identical to $I^*(D, \ell)$ up to a measure zero set. Let (D, ℓ) be given, and let $I^* = I^*(D, \ell)$. For any partition $I = \{I_d : d \in D\}$, let

$$W(D, \ell, I) = \sum_{d \in D} \int_{I_d} [u(d, x) - \lambda \ell_d] G(\mathrm{d}x).$$

By (6), for any I and any $d \in D$, if $x \in I_d^* \cap I_{d'}$ with $d \neq d'$ then

$$u(d, x) - \lambda \ell_d > u(d', x) - \lambda \ell_{d'}.$$

Thus, since $\mathbb{P}(X - \cup_{d \in D} I_d^*) = 0$ by (A_2) and the fact that G has full support,

$$\begin{aligned} & W(D, \ell, I^*) - W(D, \ell, I) \\ = \; & \sum_{d, d' \in D} \int_{I_d^* \cap I_{d'}} \left[(u(d, x) - \lambda \ell_d) - (u(d', x) - \lambda \ell_{d'}) \right] G(\mathrm{d}x) \\ - \; & \sum_{d \in D} \int_{I_d \cap (X - \cup_{d \in D} I_d^*)} [u(d, x) - \lambda \ell_d] G(\mathrm{d}x) \\ = \; & \sum_{d \neq d' \in D} \int_{I_d^* \cap I_{d'}} \left[(u(d, x) - \lambda \ell_d) - (u(d', x) - \lambda \ell_{d'}) \right] G(\mathrm{d}x) \geq 0, \end{aligned}$$

and the inequality is strict if $\mathbb{P}(I_d^* \cap I_{d'}) > 0$ for some $d \neq d'$. □

A.3. **Proof of Theorem 4.2.** Let $Z = (D, \ell, I)$ optimal.

(a) Given I, the length profile ℓ must deliver the lowest average length among those satisfying (5), for otherwise by Lemma 3.1 we can find another inquiry that implements the same expected utility from actions but with lower expected cost.

The length profile must be given by the Huffman coding (Cover and Thomas, 2006, Theorem 5.8.1).

(b) Suppose, by contradiction, that $\mathbb{P}(I_d) > \mathbb{P}(I_{d'})$ and $\ell_d > \ell_{d'}$. Let $Z' = (D, \ell', I)$ be another outcome which is the same as Z, except $\ell'_d = \ell_{d'}$ and $\ell'_{d'} = \ell_d$. Note that Z' satisfies (5) and hence, by Lemma 3.1, can be induced by an inquiry. As Z and Z' share the same (D, I), they implement the same actions in the same states, and thus have the same gross utility. However, Z' is strictly less costly, as it has a shorter expected length than Z:

$$\sum_{d \in D} \ell'_d \mathbb{P}(I_d) - \sum_{d \in D} \ell_d \mathbb{P}(I_d) = [\mathbb{P}(I_d)\ell'_d + \mathbb{P}(I_{d'})\ell'_{d'}] - [\mathbb{P}(I_d)\ell_d + \mathbb{P}(I_{d'})\ell_{d'}]$$
$$= [\mathbb{P}(I_d)\ell_{d'} + \mathbb{P}(I_{d'})\ell_d] - [\mathbb{P}(I_d)\ell_d + \mathbb{P}(I_{d'})\ell_{d'}] = -(\mathbb{P}(I_d) - \mathbb{P}(I_{d'}))(\ell_d - \ell_{d'}) < 0.$$

This is a contradiction to the optimality of Z. □

A.4. **Auxiliary Lemma.** We introduce notation and provide an auxiliary lemma that will be used in the proofs of Theorems 4.3 and 4.4 and Proposition 4.2 below.

Let $(D, \ell) \in \mathcal{F}^*$, let $K = |D|$, and let $I = I^*(D, \ell)$ be given by (6). We order the actions in D according to how likely they are chosen, so $D = \{d_k\}_{k=1}^K$, such that $(d_1, ..., d_k)$ satisfies (10). Let E_k^λ be the event that an action in $\{d_1, ..., d_k\}$ (i.e., one of k most likely actions) is preferred to all other actions when the cost λ is taken into account:

$$E_k^\lambda = \left\{ x \in X : \max_{k'=1,...,k} u(d_{k'}, x) - \lambda \ell_{d_{k'}} > \max_{m=k+1,...,K} u(d_m, x) - \lambda \ell_{d_m} \right\}, \tag{15}$$

Similarly, let E_k^0 be the event of optimally choosing an action in $\{d_1, ..., d_k\}$ in the zero-cost benchmark, $\lambda = 0$.

It can be easily seen that when (D, ℓ) is optimal, E_k^λ coincides with $\bigcup_{d \in D_k} I_d^*(D, \ell)$ except, possibly, on a measure zero set. In words, conditional on event E_k^λ, the inquiry almost surely leads to an action in $\{d_1, ..., d_k\}$.

Lemma A.1. *Let (D, ℓ, I) be an optimal outcome, let $K = |D|$, and let actions in D be ordered according to (18). For each $\lambda_1, \lambda_2 \in \mathbb{R}_+$ with $\lambda_1 < \lambda_2$,*

$$E_k^{\lambda_1} \subseteq E_k^{\lambda_2} \quad \textit{for all } k = 1, 2, ..., K-1. \tag{16}$$

Moreover, if ℓ is not uniform, then there exists $k^* \in \{1, ..., K-1\}$ such that

$$\ell_{k^*} < \ell_{k^*+1} \text{ and the set } E_{k^*}^{\lambda_2} - E_{k^*}^{\lambda_1} \text{ has a non-empty interior.} \tag{17}$$

Proof. First, we prove (16). By (10) and Theorem 4.2(b), we have

$$\ell_{d_1} \leq \ell_{d_2} \leq ... \leq \ell_{d_K}. \tag{18}$$

Let $k \in \{1, ..., K-1\}$. Suppose by contradiction that there exists $x \in E_k^{\lambda_1}$ such that $x \notin E_k^{\lambda_2}$. By (15), $x \in E_k^{\lambda_1}$ and $x \notin E_k^{\lambda_2}$ imply that there exist k^* and m^*, with $k^* \leq k < m^*$, such that

$$u(d_{k^*}, x) - \lambda_1 \ell_{d_{k^*}} = \max_{k'=1,...,k} u(d_{k'}, x) - \lambda_1 \ell_{d_{k'}} > u(d_{m^*}, x) - \lambda_1 \ell_{d_{m^*}}, \tag{19}$$

$$u(d_{k^*}, x) - \lambda_2 \ell_{d_{k^*}} \leq \max_{k'=k+1,...,K} u(d_{k'}, x) - \lambda_2 \ell_{d_{k'}} = u(d_{m^*}, x) - \lambda_2 \ell_{d_{m^*}}. \tag{20}$$

Combining (19) and (20), we obtain

$$\lambda_2(\ell_{d_{m^*}} - \ell_{d_{k^*}}) \leq u(d_{m^*}, x) - u(d_{k^*}, x) < \lambda_1(\ell_{d_{m^*}} - \ell_{d_{k^*}}).$$

This is a contradiction to $\lambda_2 > \lambda_1$ and $\ell_{d_{m^*}} \geq \ell_{d_{k^*}}$, where the last inequality is by (18) and $m^* > k^*$. This proves (16).

Next, suppose ℓ is not uniform. Then, by definition of the uniform profile and (18), there exists $k^* \in \{1, ..., K-1\}$ such that

$$\ell_{d_1} = ... = \ell_{d_{k^*}} < \ell_{d_{k^*+1}} \leq ... \leq \ell_{d_K}. \tag{21}$$

We now prove that the set $E_{k^*}^{\lambda_2} - E_{k^*}^{\lambda_1}$ has a non-empty interior. Define

$$\bar{u}(x) = \max_{k=1,...,k^*} u(d_k, x) \text{ and } w_\lambda(x) = \left(\max_{m=k^*+1,...,K} u(d_m, x) - \lambda \ell_{d_m}\right) - (\bar{u}(x) - \lambda \ell_{d_{k^*}}).$$

For each $x \in X$ there exists $m^*(x) \in \{k^*+1, ..., K\}$ such that

$$\begin{aligned} w_{\lambda_2}(x) &= u(d_{m^*(x)}, x) - \lambda_2 \ell_{d_{m^*(x)}} + \lambda_2 \ell_{d_{k^*}} - \bar{u}(x) \\ &< u(d_{m^*(x)}, x) - \lambda_1 \ell_{d_{m^*(x)}} + \lambda_1 \ell_{d_{k^*}} - \bar{u}(x) \leq w_{\lambda_1}(x), \end{aligned}$$

where the strict inequality follows from $\lambda_1 < \lambda_2$ and $\ell_{d_{k^*}} < \ell_{d_{m^*(x)}}$. We thus obtain

$$w_{\lambda_2}(x) < w_{\lambda_1}(x) \text{ for all } x \in X \text{ and all } \lambda_1 < \lambda_2. \tag{22}$$

Next, by (15) and (21), we have

$$x \in E_{k^*}^{\lambda} \iff w_\lambda(x) < 0. \tag{23}$$

Fix $\lambda_1 < \lambda_2$. To show (17), we construct an open set in which $w_{\lambda_1}(x) \geq 0 > w_{\lambda_2}(x)$. By assumption ($A_2$), the sets $E_{k^*}^{\lambda_1}$ and $X - E_{k^*}^{\lambda_2}$ have nonempty interiors. Let

$$y \in Int(E_{k^*}^{\lambda_1}) \text{ and } z \in Int(X - E_{k^*}^{\lambda_2}).$$

Since X is convex and points y and z are in $Int(X)$, we have $\alpha y + (1-\alpha)z \in Int(X)$ for all $\alpha \in (0,1)$. Thus, since $w_{\lambda_1}(x)$ and $w_{\lambda_2}(x)$ are continuous in x by (A_1), there exist $\alpha_1, \alpha_2 \in (0,1)$ such that

$$w_{\lambda_1}(\alpha_1 y + (1-\alpha_1)z) = 0 \text{ and } w_{\lambda_2}(\alpha_2 y + (1-\alpha_2)z) = 0.$$

Moreover, by (22), and (23), we have $\alpha_2 < \alpha_1$. Consider $\alpha^* \in (\alpha_2, \alpha_1)$, and let $x^* = \alpha^* y + (1-\alpha^*)z$, so $w_{\lambda_1}(x^*) > 0 > w_{\lambda_2}(x^*)$. Since $w_{\lambda_1}(x)$ and $w_{\lambda_2}(x)$ are continuous in x by (A_1), there exists an open neighborhood O_{x^*} of x^* such that $w_{\lambda_1}(x) > 0 > w_{\lambda_2}(x)$ for all $x \in O_{x^*}$. By (23), this proves (17). □

A.5. **Proof of Theorem 4.3.** Let $\lambda_1 < \lambda_2$. For $j = 1, 2$, let Q_{λ_j} be an optimal inquiry for λ_j, and let $Z^{\lambda_j} = (D^{\lambda_j}, \ell^{\lambda_j}, I^{\lambda_j})$ be the associated outcome. Denote

$$\bar{u}(Z^{\lambda_j}) \equiv \sum_{d \in D^{\lambda_j}} \int_{x \in I_d^{\lambda_j}} u(d,x) G(\mathrm{d}x), \quad j = 1, 2.$$

By (3) and (9), we have $W(Q_{\lambda_j}; \lambda_j) = \bar{u}(Z^{\lambda_j}) - \lambda_j \bar{\ell}(Z^{\lambda_j})$. By the optimality of Z^{λ_j} given λ_j, for each $j = 1, 2$, we have

$$\bar{u}(Z^{\lambda_1}) - \lambda_1 \bar{\ell}(Z^{\lambda_1}) \geq \bar{u}(Z^{\lambda_2}) - \lambda_1 \bar{\ell}(Z^{\lambda_2}) \text{ and } \bar{u}(Z^{\lambda_2}) - \lambda_2 \bar{\ell}(Z^{\lambda_2}) \geq \bar{u}(Z^{\lambda_1}) - \lambda_2 \bar{\ell}(Z^{\lambda_1}).$$

Combining these inequalities yields

$$\lambda_1 \left(\bar{\ell}(Z^{\lambda_1}) - \bar{\ell}(Z^{\lambda_2}) \right) \leq \bar{u}(Z^{\lambda_1}) - \bar{u}(Z^{\lambda_2}) \leq \lambda_2 \left(\bar{\ell}(Z^{\lambda_1}) - \bar{\ell}(Z^{\lambda_2}) \right).$$

Thus, $\bar{\ell}(Z^{\lambda_1}) \geq \bar{\ell}(Z^{\lambda_2})$ whenever $\lambda_1 < \lambda_2$.

Next, we show that $\bar{\ell}(Z^{\lambda_1}) \neq \bar{\ell}(Z^{\lambda_2})$ unless $|D^{\lambda_1}| = |D^{\lambda_2}|$ and both ℓ^{λ_1} and ℓ^{λ_2} are uniform. Suppose that both ℓ^{λ_1} and ℓ^{λ_2} are uniform, but $|D^{\lambda_1}| \neq |D^{\lambda_2}|$. Then, by definition of uniform profiles, we have $\bar{\ell}(Z^{\lambda_1}) = \log_2 |D^{\lambda_1}|$ and $\bar{\ell}(Z^{\lambda_2}) = \log_2 |D^{\lambda_2}|$. Consequently, $|D^{\lambda_1}| \neq |D^{\lambda_2}|$ implies $\bar{\ell}(Z^{\lambda_1}) \neq \bar{\ell}(Z^{\lambda_2})$.

It remains to consider the case where ℓ^{λ_1} or ℓ^{λ_2} are non-uniform. Suppose that ℓ^{λ_1} is non-uniform; the other case is symmetric. For any given form (D, ℓ), denote by $I^{\lambda}(D, \ell)$ the optimal content given by (6) under λ. We have two cases.

First, suppose that there exists $\tilde{\lambda} \in (\lambda_1, \lambda_2)$ such that $(D^{\lambda_1}, \ell^{\lambda_1})$ is optimal over $[\lambda_1, \tilde{\lambda}]$. To simplify notation, let $(D, \ell) = (D^{\lambda_1}, \ell^{\lambda_1})$. Let $K = |D|$, and let the actions $d_1, ..., d_K$ in D be ordered according to (10) under λ_1. Then, by (10), the optimality of (D, ℓ), and Theorem 4.2(b), we have

$$\ell_{d_1} \leq \ell_{d_2} \leq ... \leq \ell_{d_K}. \tag{24}$$

Let $Z^{\tilde{\lambda}} = (D, \ell, I^{\tilde{\lambda}})$, and let $E_k^{\tilde{\lambda}}$ be given by (15) under cost $\tilde{\lambda}$. By the optimality of (D, ℓ) and Theorem 4.1, for each $\lambda = \{\lambda_1, \tilde{\lambda}\}$ we have

$$\mathbb{P}(E_k^{\lambda}) = \mathbb{P}\left(\bigcup_{k'=1}^{k} I_{d_{k'}}^{\lambda}(D, \ell)\right) = \sum_{k'=1}^{k} \mathbb{P}\big(I_{d_{k'}}^{\lambda}(D, \ell)\big).$$

By the optimality of (D, ℓ) under both λ_1 and $\tilde{\lambda}$, $\lambda_1 < \tilde{\lambda}$, and Lemma A.1, we have

$$\mathbb{P}\big(E_k^{\tilde{\lambda}}\big) \geq \mathbb{P}\big(E_k^{\lambda_1}\big), \quad \text{for each } k = 1, ..., K-1, \tag{25}$$

and there exists $k^* \in \{1, ..., K-1\}$ such that

$$\ell_{d_{k^*}} < \ell_{d_{k^*+1}} \quad \text{and} \quad \mathbb{P}\big(E_{k^*}^{\tilde{\lambda}}\big) > \mathbb{P}\big(E_{k^*}^{\lambda_1}\big). \tag{26}$$

Thus, we obtain

$$\begin{aligned}
\bar{\ell}(Z^{\tilde{\lambda}}) - \bar{\ell}(Z^{\lambda_1}) &= \sum_{k=1}^{K} \Big[\ell_{d_k} \mathbb{P}\big(I_{d_k}^{\tilde{\lambda}}(D, \ell)\big) - \ell_{d_k} \mathbb{P}\big(I_{d_k}^{\lambda_1}(D, \ell)\big)\Big] \\
&= \sum_{k=1}^{K-1} (\ell_{d_k} - \ell_{d_{k+1}}) \left[\sum_{k'=1}^{k} \mathbb{P}\big(I_{d_{k'}}^{\tilde{\lambda}}(D, \ell)\big) - \sum_{k'=1}^{k} \mathbb{P}\big(I_{d_{k'}}^{\lambda_1}(D, \ell)\big)\right] \\
&= \sum_{k=1}^{K-1} (\ell_{d_k} - \ell_{d_{k+1}}) \Big[\mathbb{P}\big(E_k^{\tilde{\lambda}}\big) - \mathbb{P}\big(E_k^{\lambda_1}\big)\Big] < 0,
\end{aligned} \tag{27}$$

where the first equality is by (9), the second equality is by rearrangement, and the inequality is by (24), (25), and (26). We already established that $\lambda_2 \geq \tilde{\lambda}$ implies $\bar{\ell}(Z^{\lambda_2}) \leq \bar{\ell}(Z^{\tilde{\lambda}})$. Thus, by (27), we obtain $\bar{\ell}(Z^{\lambda_2}) \leq \bar{\ell}(Z^{\tilde{\lambda}}) < \bar{\ell}(Z^{\lambda_1})$.

Second, suppose that $(D^{\lambda_1}, \ell^{\lambda_1})$ is not optimal for λ in $[\lambda_1, \tilde{\lambda}]$ for all $\tilde{\lambda} > \lambda_1$. Because $\mathcal{F}^*$ is finite, there exists a different form $(D, \ell) \in \mathcal{F}^*$, $(D, \ell) \neq (D^{\lambda_1}, \ell^{\lambda_1})$, and an interval $(\lambda_1, \tilde{\lambda}]$, with $\tilde{\lambda} \in (\lambda_1, \lambda_2)$, such that (D, ℓ) is an optimal form for all $\lambda \in (\lambda_1, \tilde{\lambda}]$. Thus, for all $\lambda \in (\lambda_1, \tilde{\lambda}]$ we have $V(D, \ell; \lambda) > V(D^{\lambda_1}, \ell^{\lambda_1}; \lambda)$, where

$$V(D, \ell; \lambda) = \int_X \max_{d \in D} (u(d, x) - \lambda \ell_d) \, G(\mathrm{d}x) \tag{28}$$

denotes the optimal value under (D, ℓ). Clearly, V is continuous and convex in λ. Hence, the right derivatives of V w.r.t. λ always exist, denoted by V'_+. It then follows from $V(D, \ell; \lambda) > V(D^{\lambda_1}, \ell^{\lambda_1}; \lambda)$ over $(\lambda_1, \tilde{\lambda}]$ that $V'_+(D, \ell; \lambda_1) > V'_+(D^{\lambda_1}, \ell^{\lambda_1}; \lambda_1)$. Moreover, by the Envelope Theorem, we have

$$-\bar{\ell}(Z^{\lambda_1}) \leq V'_+(D^{\lambda_1}, \ell^{\lambda_1}; \lambda_1) < V'_+(D, \ell; \lambda_1).$$

Moreover, since $V(D, \ell; \lambda)$ is differentiable almost everywhere, there exists $\lambda' \in [\lambda_1, \tilde{\lambda})$ such that

$$-\bar{\ell}(D, \ell, I^{\lambda'}(D, \ell)) = V'(D, \ell; \lambda') > V'_+(D, \ell; \lambda_1) > -\bar{\ell}(Z^{\lambda_1}).$$

This then implies that $\bar{\ell}(Z^{\lambda_2}) \leq \bar{\ell}(D, \ell, I^{\lambda'}(D, \ell)) < \bar{\ell}(Z^{\lambda_1})$. □

A.6. **Proof of Proposition 4.1.** Let $\lambda_2 = \sup_{x \in X} (\max_{a \in A} u(a, x) - \min_{a \in A} u(a, x))$. Then, for all $\lambda \geq \lambda_2$, the utility gain from distinguishing any actions is smaller than the cost, so the optimal consideration set D is a singleton.

Next we show that the optimal consideration set must be A^* for λ sufficiently small. By the assumption of the proposition, each action $a \in A^*$ is optimal for a positive measure of states. As a result, for any (A^*, ℓ) and (D', ℓ') in $\mathcal{F}^*$ with $D' \subsetneq A^*$,

$$\max_{(A^*, \ell) \in \mathcal{F}^*} V(\lambda; A^*, \ell) > \max_{(D', \ell') \in \mathcal{F}^* \text{ with } D' \subsetneq A^*} V(\lambda; D, \ell')$$

at $\lambda = 0$, and by continuity, there exists $\lambda_1 > 0$ such that the same inequality holds for all $\lambda \leq \lambda_1$. Thus, for any $\lambda \leq \lambda_1$, any optimal inquiry has $D = A^*$.

Next, let (D, ℓ, I) be the outcome of an optimal inquiry Q. Denote

$$\delta(a', a'') = \sup_{x \in X} |u(a', x) - u(a'', x)|.$$

Suppose by contradiction that $\delta(a', a'') < \lambda$ for some $a', a'' \in D$. There are two cases.

Case 1. Suppose that $\ell_{a'} \neq \ell_{a''}$. W.l.o.g., let $\ell_{a'} > \ell_{a''}$. By Theorem 4.1, equation (6), and assumption (A_2), $a' \in D$ implies that the set

$$I_{a'} = \{x \in X : u(a', x) > u(a, x) + \lambda(\ell_{a'} - \ell_a) \text{ for all } a \in D - \{a'\}\} \tag{29}$$

has nonempty interior. Therefore, because $a'' \in D$, we must have

$$u(a',x) > u(a'',x) + \lambda(\ell_{a'} - \ell_{a''}) \geq u(a'',x) + \lambda \text{ for each } x \in I_{a'},$$

where the first inequality is by (29), and the second inequality is because $\ell_{a'} > \ell_{a''}$ and both $\ell_{a'}$ and $\ell_{a''}$ are integers. This contradicts the assumption that $\delta(a',a'') < \lambda$.

Case 2. Suppose that $\ell_{a'} = \ell_{a''}$. Consider an alternative outcome $(\hat{D},\hat{\ell},\hat{I})$ given by $\hat{D} = D - \{a''\}$, $\hat{\ell}_{a'} = \ell_{a'} - 1$, $\hat{\ell}_a = \ell_a$ for all $a \in D - \{a'\}$, $\hat{I}_{a'} = I_{a'} \cup I_{a''}$, and $\hat{I}_a = I_a$ for all $a \in D - \{a'\}$. In words, this outcome is the same as (D,ℓ,I) except that this outcome merges actions a' and a'' and the two categories that distinguishes these actions into one. Because

$$\ell_{a'} = \ell_{a''} = \hat{\ell}_{a'} + 1, \tag{30}$$

we obtain $2^{-\ell_{a'}} + 2^{-\ell_{a''}} = 2^{-\hat{\ell}_{a'}}$. Since $\sum_{d\in D} 2^{-\ell_d} = 1$, we obtain that

$$\sum_{d\in\hat{D}} 2^{-\hat{\ell}_d} = \left(\sum_{d\in\hat{D}-\{a'\}} 2^{-\hat{\ell}_d}\right) + 2^{-\hat{\ell}_{a'}} = \left(\sum_{d\in D-\{a',a''\}} 2^{-\ell_d}\right) + 2^{-\ell_{a'}} + 2^{-\ell_{a''}} = 1.$$

Thus, by Lemma 3.1, there exists an inquiry $\hat{Q}$ with outcome $(\hat{D},\hat{\ell},\hat{I})$. As Q and $\hat{Q}$ differ only for $x \in I_{a'} \cup I_{a''}$, we obtain

$$\begin{aligned}
W(\hat{Q};\lambda) - W(Q;\lambda) &= \int_{I_{a'}} \Big((u(a',x) - \lambda\hat{\ell}_{a'}) - (u(a',x) - \lambda\ell_{a'})\Big)G(\mathrm{d}x) \\
&\quad + \int_{I_{a''}} \Big((u(a',x) - \lambda\hat{\ell}_{a'}) - (u(a'',x) - \lambda\ell_{a''})\Big)G(\mathrm{d}x) \\
&= \int_{I_{a'}} \lambda G(\mathrm{d}x) + \int_{x\in I_{a''}} \Big(u(a',x) - u(a'',x) + \lambda\Big)G(\mathrm{d}x) > 0,
\end{aligned}$$

where the first equality is by (3), the second equality is by (30), and the inequality is because $\delta(a',a'') < \lambda$ and $I_{a'} \cup I_{a''}$ has nonempty interior. We thus obtain a contradiction to the optimality of Q. □

A.7. **Proof of Theorem 4.4.** Let Q be an optimal inquiry. Because Q induces an information partition, so $p_Q(d_1,...,d_k|x)$ is always either 0 or 1, for the purpose of the proof it is convenient to use the notation introduced in Section A.4. Let (D,ℓ,I) be an outcome of Q and let E_k^λ be given by (15). Then, the definition of confirmation bias can be equivalently expressed as follows. An inquiry Q with outcome (D, ℓ, I) has confirmation bias if for every order $(d_k)_{k=1}^{K}$ that satisfies (10),

$$E_k^0 \subseteq E_k^\lambda \text{ for all } k = 1, 2, ..., K-1. \tag{31}$$

It has strict confirmation bias if (31) holds, and there exists $k \in \{1, ..., K-1\}$ such that

$$E_k^\lambda - E_k^0 \text{ has a non-empty interior.} \tag{32}$$

Then, Theorem 4.4 is immediate by Lemma A.1 with $\lambda_1 = 0$ and $\lambda_2 = \lambda > 0$. □

A.8. **Proof of Proposition 4.2.** Because $\mathcal{F}^*$ is finite, for every cost $\lambda > 0$ there exists a form $(D, \ell) \in \mathcal{F}^*$ and an interval $[\lambda', \lambda'']$, with $\lambda' < \lambda''$ and $\lambda \in [\lambda', \lambda'']$, such that (D, ℓ) is an optimal form for any cost in $[\lambda', \lambda'']$. Now, consider arbitrary $\lambda_1, \lambda_2 \in [\lambda', \lambda'']$ with $\lambda_1 < \lambda_2$. By Lemma A.1, it is immediate that, for each $k = 1, ..., K-1$,

$$E_k^{\lambda_1} \subseteq E_k^{\lambda_2}. \tag{33}$$

Moreover, if ℓ is not uniform, the inclusion (33) is strict for some k with the difference having a non-empty interior. We thus obtain that confirmation bias is increasing, and strictly so whenever ℓ is not uniform. □

References

Bloedel, A. W. and W. Zhong (2024): "The cost of optimally acquired information," *Working Paper.*

Caplin, A., M. Dean, and J. Leahy (2019): "Rational inattention, optimal consideration sets, and stochastic choice," *Review of Economic Studies*, 86, 1061–1094.

——— (2022): "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy," *Journal of Political Economy*, 130, 1676–1715.

Cover, T. M. and J. A. Thomas (2006): *Elements of Information Theory*, Wiley-Interscience, 2nd ed.

Cremer, J., L. Garicano, and A. Prat (2007): "Language and the Theory of the Firm," *The Quarterly Journal of Economics*, 122, 373–407.

Croskerry, P., G. Singhal, and S. Mamede (2013): "Cognitive debiasing 1: origins of bias and theory of debiasing," *BMJ Quality & Safety*, 22, 58–64.

Denti, T., M. Marinacci, and A. Rustichini (2022): "Experimental cost of information," *American Economic Review*, 112, 3106–3123.

Dilmé, F. (2024): "Precision vs Complexity: Efficient Communication and Optimal Codes," Working paper.

Gould, J. B. and R. A. Leo (2010): "One hundred years later: Wrongful convictions after a century of research," *The Journal of Criminal Law and Criminology*, 825–868.

Hébert, B. and M. Woodford (2021): "Neighborhood-based information costs," *American Economic Review*, 111, 3225–3255.

Huffman, D. A. (1952): "A Method for the Construction of Minimum-Redundancy Codes," *Proceedings of the IRE*, 40, 1098–1101.

Ishak, K. J., N. Kreif, A. Benedict, and N. Muszbek (2013): "Overview of parametric survival analysis for health-economic applications," *Pharmacoeconomics*, 31, 663–675.

Jackson, C. H., L. D. Sharples, and S. G. Thompson (2010): "Survival models in health economic evaluations: balancing fit and parsimony to improve prediction," *International Journal of Biostatistics*, 6.

Jehiel, P. and J. Steiner (2020): "Selective sampling with information-storage constraints," *The Economic Journal*, 130, 1753–1781.

Jung, J., J. H. Kim, F. Matějka, and C. A. Sims (2019): "Discrete actions in information-constrained decision problems," *The Review of Economic Studies*, 86, 2643–2667.

KRAFT, L. G. (1949): "A device for quantizing, grouping, and coding amplitude-modulated pulses," Ph.D. thesis, Massachusetts Institute of Technology.

MAĆKOWIAK, B., F. MATĚJKA, AND M. WIEDERHOLT (2023): "Rational inattention: A review," *Journal of Economic Literature*, 61, 226–273.

MANDLER, M. (2024): "Increasing returns and the efficient acquisition of information," *Journal of Economic Theory*, 105815.

MANZINI, P. AND M. MARIOTTI (2014): "Stochastic choice and consideration sets," *Econometrica*, 82, 1153–1176.

MASATLIOGLU, Y., D. NAKAJIMA, AND E. Y. OZBAY (2012): "Revealed attention," *American Economic Review*, 102, 2183–2205.

MATĚJKA, F. AND A. MCKAY (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105, 272–298.

MORRIS, S. AND P. STRACK (2019): "The Wald problem and the equivalence of sequential sampling and static information costs," Available at SSRN: https://ssrn.com/abstract=2991567.

NICKERSON, R. S. (1998): "Confirmation bias: A ubiquitous phenomenon in many guises," *Review of general psychology*, 2, 175–220.

NIMARK, K. P. AND S. SUNDARESAN (2019): "Inattention and belief polarization," *Journal of Economic Theory*, 180, 203–228.

POMATTO, L., P. STRACK, AND O. TAMUZ (2023): "The cost of information: The case of constant marginal costs," *American Economic Review*, 113, 1360–1393.

SIMS, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50, 665–690.

SINGH, H., A. N. MEYER, AND E. J. THOMAS (2014): "The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving US adult populations," *BMJ quality & safety*, 23, 727–731.

SINGH, H., G. D. SCHIFF, M. L. GRABER, I. ONAKPOYA, AND M. J. THOMPSON (2017): "The global burden of diagnostic errors in primary care," *BMJ quality & safety*, 26, 484–494.

STEINER, J., C. STEWART, AND F. MATĚJKA (2017): "Rational inattention dynamics: Inertia and delay in decision-making," *Econometrica*, 85, 521–553.

WALD, A. (1945): "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, 16, 117 – 186.

WILSON, A. (2014): "Bounded memory and biases in information processing," *Econometrica*, 82, 2257–2294.