

Hüttner, Frank; Boyacı, Tamer; Akçay, Yalçın

Working Paper

Consumer choice under limited attention when options have different information costs

ESMT Working Paper, No. 16-04 (R1)

Provided in Cooperation with:

ESMT European School of Management and Technology, Berlin

Suggested Citation: Hüttner, Frank; Boyacı, Tamer; Akçay, Yalçın (2016) : Consumer choice under limited attention when options have different information costs, ESMT Working Paper, No. 16-04 (R1), European School of Management and Technology (ESMT), Berlin

This Version is available at:

<https://hdl.handle.net/10419/147027>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

October 04, 2016

ESMT Working Paper 16-04 (R1)

Consumer choice under limited attention when options have different information costs

Frank Huettner, ESMT European School of Management and Technology

Tamer Boyacı, ESMT European School of Management and Technology

Yalçın Akçay, College of Administrative Sciences and Economics, Koç University

Revised version

Copyright 2016 by ESMT European School of Management and Technology GmbH, Berlin, Germany, www.esmt.org.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, used in a spreadsheet, or transmitted in any form or by any means - electronic, mechanical, photocopying, recording, or otherwise - without the permission of ESMT.

Find more ESMT working papers at [ESMT faculty publications](#), [SSRN](#), [RePEc](#), and [EconStor](#).

Consumer Choice Under Limited Attention When Options Have Different Information Costs

Frank Huettner, Tamer Boyacı

ESMT European School of Management and Technology,
Schlossplatz 1, 10178, Berlin, Germany
frank.huettner@esmt.org, tamer.boyaci@esmt.org

Yalçın Akçay

College of Administrative Sciences and Economics, Koç University,
Rumeli Feneri Yolu, Istanbul, 34450, Turkey
yakcay@ku.edu.tr

Consumers often do not have complete information about the choices they face and therefore have to spend time and effort in acquiring information. Since information acquisition is costly, consumers trade-off the value of better information against its cost, and make their final product choices based on imperfect information. We model this decision using the rational inattention approach and describe the rationally inattentive consumer's choice behavior when she faces options with different information costs. To this end, we introduce an information cost function that distinguishes between direct and implied information. We then analytically describe the optimal behavior and derive the choice probabilities in closed-form. We find that non-uniform information costs can have a strong impact on product choice, which gets particularly conspicuous when the product alternatives are otherwise very similar. It can also lead to situations where it is disadvantageous for the seller to provide easier access to information for a particular product. Furthermore, it provides a new explanation for strong failure of regularity of consumer behaviour, which occurs if the addition of an inferior (never chosen) product to the choice set increases the market share of another existing product.

Key words: discrete choice, rational inattention, information acquisition, non-uniform information costs, strong failure of regularity.

1. Introduction

Facing an abundance of product choices and related information, but with only limited time and attention to evaluate them, consumers have to come to grips with how much and what type of information to acquire and to pay attention to (and what to ignore), and make product choice and purchase decisions based on this partial information. It is therefore quite possible that consumers make “wrong” choices, but this does not necessarily imply that they are irrational. Since the works of Simon (1955, 1979), *bounded rationality* acknowledges the fact that individuals make rational

decisions, but subject to constraints. In an information driven world, attention that can be allocated to a specific choice task is limited, which puts constraints on the amount and type of information that can be acquired. As information is “costly”, rational consumers have to trade-off the value of better information against its cost. *Rational inattention*¹ theory offers a compelling approach to capture this trade-off by *endogenizing* the information acquisition process. Specifically, the pioneering works of Sims (1998, 2003, 2006) propose a framework in which information is quantified as reduction in Shannon entropy, and utility-maximizing decision makers optimally select the type and quantity of information they need, and ignore the information that is not worth obtaining and processing.

Rational inattention theory has been applied to a broad spectrum of economic problems and has been a powerful construct in providing explanations to some observed market and macroeconomic phenomena such as price stickiness, business cycles and contractions, consumption (Sims 2003, Maćkowiak and Wiederholt 2009, 2015, and Tutino 2013). It is increasingly applied to microeconomics topics as well, especially pricing (e.g., Matějka and McKay 2012, Matějka 2015 and Boyaci and Akçay 2016). A fundamental driver of these applications is the evolving understanding of how rational inattention influences choice behavior. In a recent paper, Matějka and McKay (2015) study the choice behavior of rationally inattentive consumers facing discrete choices with stochastic (pay-off) values, assuming that the costs of acquiring and processing information is *identical* across choice alternatives. They establish that the optimal information processing strategy leads to a choice behaviour that can be characterized as *generalized* multinomial logit (GMNL). In particular, the choice probabilities depend not only on the true realizations of the choices as in the standard multinomial logit (MNL), but also on the prior beliefs of the consumer and the cost of information. In this paper, we generalize the GMNL model by describing the rationally inattentive consumer’s choice behavior when she faces options with *different* information costs.

There are three key reasons as to why rational inattention to discrete choice with different information costs is significant. These also constitute the cuneate contributions of our paper:

REALISM & APPLICABILITY. The uniform information cost assumption underpinning the GMNL characterization can be interpreted as the consumer acquiring and processing information through a common channel with a certain associated cost. Effectively, it means that the amount of effort (and hence cost) spent to obtain and process 1-byte of information about each option is the same. Clearly, this is not necessarily the case in reality. It is often times easier to obtain information about some products than about others, by the very nature of the product. Or sometimes information is obtained from different sources with different levels of time-and-attention-efficiency (online, catalog, direct sales force etc). It can also depend on the assortment that is offered – it is easier to obtain

¹ Throughout the paper, we use the terms “limited attention” and “rational inattention” interchangeably.

information about products that are readily available for “touch and feel” compared to others that are not available for such an experience and require extra effort to garner information.² These realities call for a choice model that allows the information cost to vary among the alternatives considered by the consumer. Such a choice model would form the essential building block for a variety of operations/marketing applications involving consumers with limited attention.

RELATION TO OTHER CHOICE MODELS. Discrete choice models under rational inattention are particularly promising because of their close relation to MNL choice models. By generalizing the GMNL model, we extend this relationship. The connection with MNL is particularly relevant in our context because a rather common approach to model bounded rationality of customers is to adopt the quantal choice model of Luce (1959), which leads to the MNL (see McKelvey and Palfrey 1995). We refer to Anderson et al. (1992) for a comprehensive coverage of MNL models in general and to Wierenga (2008) for their use in marketing science. MNL and its variations (e.g., nested logit) have been extensively used to model consumer behavior in the operations management literature as well, in particular in the context of pricing, revenue management and assortment planning (Hanson and Martin 1996, van Ryzin and Mahajan 1999, Dong et al. 2009, Zhang and Adelman 2009, Davis et al. 2014, among many others). In this stream, the need for richer and more general choice models has also been recognized and some propositions have been made, such as Talluri and van Ryzin (2004), Alptekinoglu and Semple (2015), Blanchet et al. (2016), Srikanth and Rusmevichientong (2016). Our paper complements this literature by offering a new, general and versatile choice model that is derived from an analysis of the optimal behaviour of an individual consumer.

INSIGHTS ON CHOICE BEHAVIOR. The detailed assessment of information costs in our framework allows insight into the attention allocation strategy of the consumer, which drives the ultimate choice behavior. By comparison with the case of uniform costs, we show that cost differences among the alternatives have substantial impact on the optimal choice of consumers. Naturally, the consumer pays more attention on and processes more information about options with lower costs. If the options are otherwise identical, this implies a strict preference for the one with the lowest cost. However, in general, the information obtained from the “cheaper” options can increase or decrease the likelihood of choosing other options. In this sense, reducing the cost of an alternative does not always mean it will be selected more often, nor does it imply overall better choices for the consumer. As a matter of fact, it can lead to new decision biases. We establish that our choice model does not suffer from IIA (independence of irrelevant alternatives); duplicate (identical) options are jointly processed as one. Furthermore, dominated options are never selected. However, their presence can

² S. J. Hoch (1986) and Hamilton and Thompson (2007) stress the importance of consumers’ experience at the point of sale and consumers’ struggle to judge on the value of a product through abstract product description compared with direct user experience.

influence choice behavior. Specifically, we show that if a dominated option has a lower information cost, its addition to a set can increase choice probability of another option, providing a new plausible reason for the failure of regularity in choice behavior. By providing a precise description of rational choice behavior under limited attention and costly information, our model has the potential to guide product assortment and information provisioning strategies of firms.

The central and novel element of our model is the derivation of the total cost associated with the consumer's information processing strategy. Quantifying the amount of information the consumer acquires when evaluating a particular option and accounting for its cost is an intricate task in the presence of a non-uniform information cost structure, and this gets even more pronounced when there are similarities (i.e. correlations) between the products. This is because, as the consumer learns about a product, she may also learn about another product (and vice versa). Accordingly, there are two forms of information acquired by the consumer: (i) *direct information* that the consumer obtains by studying the particular option, and (ii) *implied information* that the consumer acquires about the option by studying another option. Since the unit costs of these sources might differ, it becomes important to glean from the consumer's information processing strategy the amount of information acquired from each source. The consumer should prioritize cheaper sources of information and should not attempt to obtain information about an option directly if that information can already be inferred from previously studied options. To exemplify the above, suppose there are two products, PRODUCT 1 and 2. PRODUCT 1 is available at a retail store the consumer is currently visiting, while PRODUCT 2 is not available there but sold elsewhere (e.g. online store) requiring extra travel/search effort for the consumer. The consumer knows that these products are very similar (in theory they can even be identical) and in line with the above description, the information cost is lower at the retail store. She should inquire about PRODUCT 1 only at the retail store (never at the online store), and this would also reveal significant information about PRODUCT 2. For PRODUCT 2, she should acquire (direct) information at the online store only to obtain information beyond the implied information that she can infer from PRODUCT 1 (already obtained at the retail store). We develop an information cost function that quantifies separately the amount of implied and direct information, and generalizes the Shannon entropy based cost functions utilized in the rational inattention literature. We derive this cost function from first principles, using an axiomatic approach.

We formulate the consumer's discrete choice problem based on this information cost function, and then determine the structure of the optimal solution. We show that the optimal choice behaviour can be described analytically in closed-form. We generalize the GMNL model in the sense that our choice model reduces to GMNL when the cost of information is uniform across all options. After establishing this result, we concentrate on a number of limiting cases involving infinite information costs, zero information costs, duplicate options, with the additional objective of highlighting some

features of the optimal choice. Subsequently, we focus on an auxiliary example, the classical red-bus/blue-bus problem, and illustrate the impact of multiple information channels and different costs on the choice behavior of consumers with limited attention.

2. Overview of literature

As noted earlier, the theory of rational inattention belongs to the literature of bounded rationality and receives significant interest in economics (Gabaix 2014) as well as psychology (Todd and Gigerenzer 2000). Models can differ in the way information is acquired. In Reis (2006) the consumer either pays a fee to become fully informed or not, while in other models the consumer may decide on the degree of the precision to which she receives information (e.g. in the models of Verrecchia 1982, the consumer decides on the variance of the signals). The models conceived by Sims (1998) and later adopted by many other researchers generalize this approach, as they offer the consumer the opportunity to receive signals of any type and to improve her prior in every desirable way. Our paper follows this prominent stream of modeling rational inattention. The common feature is the modeling of the cost of information as a reduction of uncertainty with respect to the prior, where uncertainty is measured as Shannon entropy (Shannon 1948). For a motivation of Shannon entropy as a measure of uncertainty and information, we refer the reader to the axiomatic treatment in Csiszár (2008) and to the excellent introduction by Cover and Thomas (2006). To our knowledge, Matějka and McKay (2015) is the only application of rational inattention to discrete choice. We expand this literature by incorporating non-uniform information cost structure to the choice decision and characterizing the resulting optimal choice.

The fact that customers need to exercise differing levels of effort in order to get informed about different alternatives has been studied for multiple purposes in the context of both parallel and sequential search models. In parallel search, consumers form a fixed set (referred to as the consideration set) of options to evaluate, and make a choice from this set (e.g. Manzini and Mariotti 2014). In contrast, all options are kept on the table in rational inattention models. In sequential search models, consumers gather information about a particular product one-by-one (and possibly one attribute at a time), and purchase once they decide to stop collecting more information optimally (e.g. Weitzman 1979, Branco et al. 2012, Ke et al. 2016). Rational inattention models differ in that no assumption is made on the process by which the consumer gets informed nor on the type or quantity of information acquired (the information strategy is fully endogenized).

3. Choice Model Formulation

In this section, we develop the choice model for a rationally inattentive consumer with different information costs across alternatives. The consumer can choose from the finite set of alternatives

$\mathcal{A} = \{1, \dots, n\}$. The state of nature is a random variable \mathbf{V} taking values $\mathbf{v} \in \mathbb{R}^n$, where v_i denotes the value of alternative $i \in \mathcal{A}$. The consumer has the prior belief $p \in \Delta(\mathbb{R}^n)$ where $\Delta(\mathbb{R}^n)$ denotes the set of all probability distributions on \mathbb{R}^n . In order to improve her decision, the consumer can process information with the goal of sharpening her belief about the state of nature. Let $\mathbf{S} = \mathbb{R}^n$ denote the signal space that is available to the consumer. Consistent with the theory of rational inattention, we assume that the consumer can set up any joint distribution $f_{\mathbf{S}, \mathbf{V}}$ of signals and values, given that it is consistent with her prior belief, i.e., the marginal of f with respect to \mathbf{V} equals p ,

$$f_{\mathbf{V}}(\mathbf{v}) = \int_{\mathbf{s} \in \mathbb{R}^n} f_{\mathbf{S}, \mathbf{V}}(d\mathbf{s}, \mathbf{v}) = p(\mathbf{v}) \quad \text{for all } \mathbf{v} \in \mathbb{R}^n.$$

Clearly, depending on the joint distribution of signals and values, the signal can be more or less beneficial. In particular, given the reception of signal $\mathbf{s} \in \mathbb{R}^n$, the consumer creates an updated belief $f_{\mathbf{V}|\mathbf{s}}$ over the values of the alternatives and chooses the alternative that gives the highest expected value under the updated belief, yielding the expected payoff

$$U(f_{\mathbf{V}|\mathbf{s}}) = \max_{i \in \mathcal{A}} \mathbb{E}_{f_{\mathbf{V}|\mathbf{s}}} [v_i]. \quad (1)$$

The less noise that remains in the updated belief, the more promising this choice becomes. Signals that reveal the true state and eliminate all uncertainty allow the consumer to make a perfect choice. This would maximize the payoff

$$R(f) = \int_{\mathbf{v} \in \mathbb{R}^n} \int_{\mathbf{s} \in \mathbb{R}^n} U(f_{\mathbf{V}|\mathbf{s}}) f_{\mathbf{S}|\mathbf{v}}(d\mathbf{s}) p(d\mathbf{v}). \quad (2)$$

The consumer has no restrictions in choosing her information strategy as long as it is consistent with her priors, but information processing is costly. Specifically, information costs depend on the extent of the reduction of uncertainty, measured by entropy \mathcal{H} . The a-priori entropy of the state of nature is given by

$$\mathcal{H}(p) = \sum_{\mathbf{v} \in \mathbb{R}^n} p(\mathbf{v}) \cdot (-\log p(\mathbf{v})).$$

Receiving signal \mathbf{s} yields a reduction of this entropy by $\mathcal{H}(p) - \mathcal{H}(f_{\mathbf{V}|\mathbf{s}})$ to the level of the entropy of the posterior belief $f_{\mathbf{V}|\mathbf{s}}$. The expectation of this reduction over all signals is called mutual information between \mathbf{S} and \mathbf{V} under the joint distribution f , and is denoted by $\mathcal{I}_f(\mathbf{V}, \mathbf{S})$

$$\mathcal{I}_f(\mathbf{V}, \mathbf{S}) = \mathcal{H}(p) - \sum_{\mathbf{s} \in \mathbf{S}} f_{\mathbf{S}}(\mathbf{s}) \mathcal{H}(f_{\mathbf{V}|\mathbf{s}}). \quad (3)$$

In particular, $\mathcal{I}_f(\mathbf{V}, \mathbf{S})$ quantifies the extent of what the consumer expects to learn about the state of nature from the signals.

Extant literature assumes that the cost per unit of mutual information λ is uniform across all alternatives, and accordingly defines the total cost of information as $C(f) = \lambda \cdot \mathcal{I}_f(\mathbf{V}, \mathbf{S})$. Matějka and McKay (2015) show that the information strategy optimizing the net pay-off $R(f) - C(f)$ results in a choice behavior that can be characterized as Generalized Multinomial Logit (GMNL). Specifically, the *conditional* probability $p(i | \mathbf{v})$ of choosing alternative $i \in A$ when the state is \mathbf{v} is given as

$$p(i | \mathbf{v}) = \frac{e^{\frac{v_i}{\lambda}} \cdot p(i)}{\sum_{j=1}^n e^{\frac{v_j}{\lambda}} \cdot p(j)} \quad (4)$$

Here $p(i) := \int_{\mathbf{v} \in \mathbb{R}^n} p(i | \mathbf{v}) p(d\mathbf{v})$ represents the *unconditional* probability of selecting alternative $i \in \mathcal{A}$. Note that $p(i)$ are not exogenous parameters; they are part of the consumer's decision making strategy, capturing the effects of prior beliefs. When the consumer is a-priori indifferent to all alternatives (i.e., $p(\mathbf{v})$ is invariant to all permutations of the elements of \mathbf{v}), then $p(i) = \frac{1}{n}$ and (4) reduces to the standard MNL formula.

In this paper, we relax the uniform information cost assumption and let λ_i denote the unit cost of acquiring information for alternative i . Without loss of generality, suppose that the alternatives are ordered such that $\lambda_1 \leq \lambda_2 \dots \leq \lambda_n$. At this point, one could speculate that the resulting conditional choice probabilities should resemble (4) with λ simply replaced with λ_i for each alternative. This conjecture, however, turns out to be incorrect. As we show later, the conditional choice probabilities can still be characterized in closed-form similar to the GMNL formula, but the “attraction” term associated with each alternative is more involved.

The most crucial element of the rationally inattentive consumer's choice framework with non-uniform information costs is the development of a total information cost function based on mutual information that accounts for the sources of information. This is because the mutual information $\mathcal{I}_f(\mathbf{V}, \mathbf{S})$ generated from the information processing strategy of the consumer contains information about different alternatives, but these alternatives carry different costs. In contrast to the uniform cost case, it becomes necessary to be more precise about the way the consumer acquires information. Conceptually, $\mathcal{I}_f(\mathbf{V}, \mathbf{S})$ is generated from a series of queries and their responses. Practically, this is tantamount to consumers studying the alternatives in some order, asking questions and updating beliefs accordingly. We do not specify the exact process by which information is acquired, but as noted earlier, we provide an axiomatic derivation of our cost function. This derivation is not limited to the case where information is measured as mutual information, i.e., as reduction of Shannon entropy. Thus, we provide a more general form in the next section before turning back to the entropy-based model in Section 3.2.

3.1. Axiomatic derivation of the total information cost function

For an information strategy f and any set of alternatives $D \subseteq \mathcal{A}$, let $I(f, D)$ denote the information that is obtained from the information strategy if the scope of its application is limited to the alternatives in D . We adopt the convention $I(f, \emptyset) = 0$ and assume that the information measure $I: \Delta(\mathbb{R}^{2n}) \times 2^{\mathcal{A}} \rightarrow \mathbb{R}_+$ satisfies the following property:

$$D \subseteq E \subseteq \mathcal{A} \Rightarrow I(f, D) \leq I(f, E) \text{ for all } f \in \Delta(\mathbb{R}^{2n}).$$

This intuitive requirement states that for any information strategy, the information obtained does not diminish when applied to a larger set of alternatives.

For a given vector of information prices λ and a given information measure I , the cost function $C_\lambda^I: \Delta(\mathbb{R}^{2n}) \rightarrow \mathbb{R}_+$ associates a cost to any information strategy $f \in \Delta(\mathbb{R}^{2n})$. We assume that an information strategy that does not provide any information costs nothing, i.e., denoting the no-information strategy by f^0 , we have $I(f^0, D) = 0$ for all $D \subseteq \mathcal{A}$ and $C_\lambda^I(f^0) = 0$. At the heart of the axiomatic derivation is the specification of how the cost of two information strategies differ, due to the differences in the amount of information obtained as well as its scope. By scope, we refer to the fact that information strategy change can be related to one particular alternative or to a combination of alternatives. In what follows, we cover each scenario separately.

We first deal with the scenario where the change in information strategy from f_1 to f_2 involves a particular alternative $i \in \mathcal{A}$. Suppose that the information that is obtained about i increases by $c \geq 0$, i.e., $I(f_2, \{i\}) = I(f_1, \{i\}) + c$. Moreover, suppose that the information that is obtained about i together with other alternatives increases by the same amount, $I(f_2, E) = I(f_1, E) + c$ for all $E \ni i$, while there is no change in information about other alternatives or sets of alternatives that do not contain i , i.e., $I(f_2, E) = I(f_1, E)$ for all $E \subseteq \mathcal{A} \setminus \{i\}$. This means that the additional information is directly related to this particular alternative i . We therefore postulate that it is acquired at the unit cost of that alternative λ_i . Formally, we have the following.

SINGULAR INFORMATION COSTING. Let $f_1, f_2 \in \Delta(\mathbb{R}^{2n})$, $c \in \mathbb{R}$, and $i \in \mathcal{A}$. If the information obtained through f_1 and f_2 for every $E \subseteq \mathcal{A}$ is such that

$$I(f_2, E) = \begin{cases} I(f_1, E) + c & \text{whenever } E \ni i \\ I(f_1, E) & \text{whenever } E \subseteq \mathcal{A} \setminus \{i\}, \end{cases}$$

then we have $C_\lambda^I(f_1) = C_\lambda^I(f_2) + c \cdot \lambda_i$.

We now deal with the scenario where the change in information strategy from f_1 to f_2 involves a set of alternatives $D \subseteq \mathcal{A}$, $|D| > 1$. Suppose that the information that is obtained about D increases by $c \geq 0$, $I(f_2, D) = I(f_1, D) + c$, while the information about proper subsets of D remains unchanged, $I(f_2, E) = I(f_1, E)$ for all $E \subsetneq D$. In this case, the additional information cannot be obtained by

considering (subsets of) alternatives in D in isolation, but only if the information is applied to *all* alternatives in D *together*. Moreover, suppose that the information that is obtained about D together with other alternatives increases by the same amount, $I(f_2, E) = I(f_1, E) + c$ for all $E \supseteq D$, while there is no change in information about other alternatives or sets of alternatives that do not contain D . This means that the gain in information is precisely the additional information about the *combination* of alternatives in D . We postulate that this information is acquired at the highest unit cost associated with the alternatives in D , $\max_{i \in D} \{\lambda_i\}$.

JOINT INFORMATION COSTING. Let $f_1, f_2 \in \Delta(\mathbb{R}^{2n})$, $c \in \mathbb{R}$, and $D \subseteq \mathcal{A}$, $|D| > 1$. If the information obtained through f_1 and f_2 for every $E \subseteq \mathcal{A}$ is such that

$$I(f_2, E) = \begin{cases} I(f_1, E) + c & \text{whenever } D \subseteq E \\ I(f_1, E) & \text{whenever } D \not\subseteq E, \end{cases}$$

then we have $C_{\lambda}^I(f_1) = C_{\lambda}^I(f_2) + c \cdot \max_{i \in D} \{\lambda_i\}$.

This condition reflects a “conservative” but reasonable view on the costs of jointly inferred information. To see this more clearly, consider a bipartition of $D = D_1 \cup D_2$ with $D_1 \cap D_2 = \emptyset$, $D_1, D_2 \neq \emptyset$. The *joint information* that is obtained from pooling the insights obtained about the alternatives D_1 with the insights about the alternatives D_2 is given by

$$I(f_1, D) - I(f_1, D_1) - I(f_1, D_2).$$

We call this joint information because this information can only be obtained when looking at the alternatives *together*. Then, whenever the condition of joint information costing holds, we have

$$I(f_1, D) - I(f_1, D_1) - I(f_1, D_2) + c = I(f_2, D) - I(f_2, D_1) - I(f_2, D_2),$$

since the information about proper subsets of D remains unchanged. According to the postulation, this additional joint information is acquired at unit cost $\max_{i \in D} \{\lambda_i\}$. This is motivated by the understanding that the information channel associated with the most “costly” alternative is required in order to obtain this joint information.³

The following result characterizes the unique cost function satisfying these properties.

THEOREM 1. *Let $\lambda \in \mathbb{R}_+^n$, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$, denote the vector specifying the unit costs for information and I denote an information measure. The unique cost function that satisfies singular information costing and joint information costing is given by*

$$C_{\lambda}^I(f) = \sum_{k=1}^n \lambda_k \cdot [I(f, \{1, \dots, k\}) - I(f, \{1, \dots, k-1\})] \quad \text{for } f \in \Delta(\mathbb{R}^{2n}). \quad (5)$$

³ Note that we do not rule out $I(f, D) - I(f, D_1) - I(f, D_2)$ to be negative. This could be the case if a priori, the consumer could easily infer information about alternatives D_2 from knowing D_1 but due to her information strategy, this opportunity to infer is intentionally ceded. This could be the case if an information strategy is designed to mainly inform about a particular alternative. Then, the joint information is negative in the sense that it is harder to infer something from D_1 on D_2 given that the signal was received, than it was before information acquisition.

Theorem 1 establishes that the two principles of information costing are characteristic to our cost function.⁴ Strictly speaking, expanding the scope of joint information costing definition by including $|D| = 1$ would contain singular information costing as a special case. Effectively, the information cost function (5) captures the idea that the consumer does not try to obtain information that is already implied by previously studied options, and concentrates on obtaining direct information beyond what she already knows. Furthermore, she prioritizes cheaper alternatives (and information channels) in acquiring information.

3.2. Non-uniform cost function when information is reduction in Shannon entropy

As is standard in the literature, we now use mutual information as the measure of information. With $I(f, D) = \mathcal{I}_f((V_i)_{i \in D}, \mathbf{S})$, the cost function (5) defined in Theorem 1 becomes

$$C_{\lambda}^{\mathcal{I}}(f) = \sum_{i=1}^N \lambda_i \cdot [\mathcal{I}_f(\mathbf{V}_{1..i}, \mathbf{S}) - \mathcal{I}_f(\mathbf{V}_{1..i-1}, \mathbf{S})]$$

where $\mathbf{V}_{1..i-1}$ denotes (V_1, \dots, V_{i-1}) . The application of the chain rule of mutual information

$$\mathcal{I}_f(V_i, \mathbf{S} | \mathbf{V}_{1..i-1}) = \mathcal{I}_f(\mathbf{V}_{1..i}, \mathbf{S}) - \sum_{k=1}^{i-1} \mathcal{I}_f(V_k, \mathbf{S} | \mathbf{V}_{1..k-1}), \quad (6)$$

where $\mathcal{I}_f(V_i, \mathbf{S} | \mathbf{V}_{1..i-1})$ is the *conditional mutual information* between V_i and the signal \mathbf{S} given $\mathbf{V}_{1..i-1}$, results in the following corollary.⁵

COROLLARY 1. *When information is measured by mutual information (3) and unit cost of information for each alternative is λ_i , where $\lambda_1 \leq \lambda_2 \dots \leq \lambda_n$, the cost of information $C(f)$ of an information strategy according to (5) is given as*

$$C(f) := C_{\lambda}^{\mathcal{I}}(f) = \sum_{i=1}^N \lambda_i \cdot \mathcal{I}_f(V_i, \mathbf{S} | \mathbf{V}_{1..i-1}). \quad (7)$$

Furthermore, $\lambda_i = \lambda$ for all $i \in \mathcal{A}$, implies $C(f) = \lambda \cdot \mathcal{I}_f(\mathbf{V}, \mathbf{S})$ and the total information cost function reduces to that in extant literature.

We now illustrate that the cost function derived in Corollary 1 represents a consumer who first learns from the cheapest source, then acquires additional information from the second cheapest source beyond what she already knows, and so on. In order to facilitate the explanation and to simplify the exposition, suppose that there are only two alternatives, PRODUCT 1 and PRODUCT 2,

⁴ Technically, for any given information strategy f , the cost function $C_{\lambda}^{\mathcal{I}}$ as a function on the domain of unit cost vectors \mathbb{R}_+^n is in fact the Lovász extension (Lovász 1983) of the information measure I as a function on $2^{\mathcal{A}}$.

⁵ The conditional mutual information $\mathcal{I}_f(V_i, \mathbf{S} | \mathbf{V}_{1..i-1})$ is often defined as the expected mutual information between V_i and \mathbf{S} conditional on $\mathbf{V}_{1..i-1}$, i.e., $\mathcal{I}_f(V_i, \mathbf{S} | \mathbf{V}_{1..i-1}) = \sum_{\mathbf{v}_{1..i-1} \in \mathbf{V}_{1..i-1}} f(\mathbf{v}_{1..i-1}) \cdot \mathcal{I}_{f_{V_i, \mathbf{S} | \mathbf{v}_{1..i-1}}}(V_i, \mathbf{S})$. The chain rule of mutual information guarantees that this definition is equivalent to the recursive definition in (6).

for which $\mathbf{V} = (V_1, V_2)$ with $0 < \lambda_1 \leq \lambda_2 < \infty$. As PRODUCT 1 has a lower information cost, let us first focus on the information learned about PRODUCT 1 from the signals. To this end, let p_{V_1} and $f_{V_1|\mathbf{s}}$ denote the marginals of the prior and the posterior with respect to the value of PRODUCT 1. By asking questions, the consumer receives signals that reduce the uncertainty of her knowledge of V_1 . This reduction is the difference between the prior and posterior entropies, i.e., the mutual information between \mathbf{S} and V_1 ,

$$\mathcal{I}_f(V_1, \mathbf{S}) = \mathcal{H}(p_{V_1}) - \sum_{\mathbf{s} \in \mathbf{S}} f_{\mathbf{S}}(\mathbf{s}) \cdot \mathcal{H}(f_{V_1|\mathbf{s}}).$$

Figure 1 illustrates this relationship on a Venn Diagram. The entropy of V_1 is reduced by $\mathcal{I}_f(V_1, \mathbf{S})$, i.e., by the intersection of $\mathcal{H}(p_{V_1})$ and $\mathcal{H}(f_{\mathbf{S}})$. Note that in a similar vein, the mutual information $\mathcal{I}_f(V_2, \mathbf{S})$ between \mathbf{S} and V_2 is given by the intersection of $\mathcal{H}(p_{V_2})$ and $\mathcal{H}(f_{\mathbf{S}})$, while the mutual information $\mathcal{I}(V_1, V_2)$ between V_1 and V_2 is given by the intersection of $\mathcal{H}(p_{V_1})$ and $\mathcal{H}(p_{V_2})$.

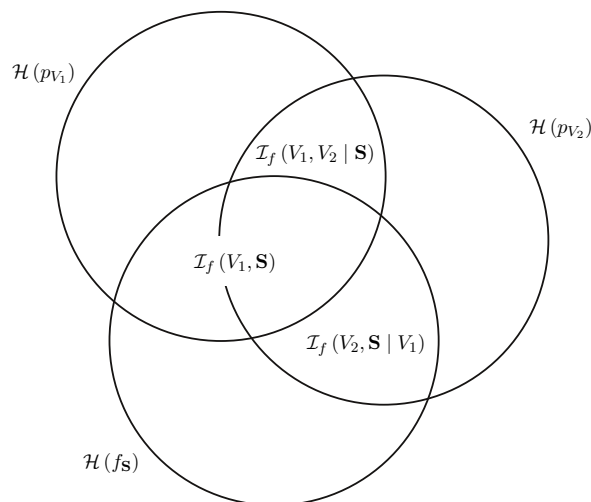


Figure 1 Relationship between entropy, mutual information and conditional mutual information

Since the signals that inform about PRODUCT 1 are processed through a channel that cost λ_1 per information unit⁶, the cost incurred by this activity is $\lambda_1 \mathcal{I}_f(V_1, \mathbf{S})$. When the consumer processes information through this channel and learns about PRODUCT 1, she typically also learns about PRODUCT 2. In other words, a part of what is learned from the signal about PRODUCT 2 is already contained in the mutual information of the signal and PRODUCT 1. This implied information about PRODUCT 2 can be quantified as the intersection of $\mathcal{I}_f(V_1, \mathbf{S})$ and $\mathcal{I}_f(V_2, \mathbf{S})$ in Figure 1. The consumer does *not* need to acquire this information by directly inquiring about PRODUCT 2 – all direct information acquisition efforts can be concentrated on learning what she needs to learn beyond

⁶ When the base of the logarithm is 2, this unit is bytes.

PRODUCT 1. This additional information is the mutual information between V_2 and \mathbf{S} conditional on V_1 , $\mathcal{I}_f(V_2, \mathbf{S} | V_1)$. To quantify this information, we need to first determine the information learned about the entire state variable (V_1, V_2) from the signal, i.e. $\mathcal{I}_f((V_1, V_2), \mathbf{S})$, which is given as the union of the intersection of $\mathcal{H}(p_{V_1})$ and $\mathcal{H}(p_{V_2})$ with $\mathcal{H}(f_{\mathbf{S}})$. Subtracting from this the information $\mathcal{I}_f(V_1, \mathbf{S})$ that the consumer already knows from PRODUCT 1, we get the mutual information between V_2 and \mathbf{S} conditional on V_1 :

$$\mathcal{I}_f(V_2, \mathbf{S} | V_1) = \mathcal{I}_f((V_1, V_2), \mathbf{S}) - \mathcal{I}_f(V_1, \mathbf{S}) \quad (8)$$

This additional information is acquired at a unit cost λ_2 , and the total cost of information becomes

$$C(f) = \lambda_1 \cdot \mathcal{I}_f(V_1, \mathbf{S}) + \lambda_2 \cdot \mathcal{I}_f(V_2, \mathbf{S} | V_1). \quad (9)$$

The Venn diagram in Figure 1 aids in visualizing the relationship between mutual information and conditional mutual information among alternatives and signals and facilitates description of our cost function. It is important to recognize, however, that this relationship is more intricate than it may seem from the Venn diagram, even for the case with two alternatives. For example, consider the intersection of $\mathcal{H}(p_{V_1})$, $\mathcal{H}(p_{V_2})$ and $\mathcal{H}(f_{\mathbf{S}})$. This area can actually represent a *negative* quantity.⁷ Hence, it is possible that $\mathcal{I}_f(V_2, \mathbf{S} | V_1) > \mathcal{I}_f(V_2, \mathbf{S})$. To see this more lucidly, consider the case where V_1 and V_2 are independent. Then, there is no mutual information between V_1 and V_2 , i.e., $\mathcal{I}(V_1, V_2) = 0$ and knowing V_1 does not reveal anything about V_2 . It seems quite plausible then that $\mathcal{I}_f(V_2, \mathbf{S} | V_1) = \mathcal{I}_f(V_2, \mathbf{S})$. However, this is *not* true. Even though V_1 and V_2 are a-priori independent, they typically do “become” dependent because of the signals received (this can be formalized via a contradiction argument for an optimal information strategy). That is, V_1 and V_2 are dependent conditional on \mathbf{S} such that the mutual information of V_1 and V_2 conditional on \mathbf{S} , $\mathcal{I}_f(V_1, V_2 | \mathbf{S})$, is non-zero. Then, the *joint information* is positive:

$$\mathcal{I}_f((V_1, V_2), \mathbf{S}) - \mathcal{I}_f(V_1, \mathbf{S}) - \mathcal{I}_f(V_2, \mathbf{S}) = \mathcal{I}_f(V_1, V_2 | \mathbf{S}) - \mathcal{I}_f(V_1, V_2) > 0, \quad (10)$$

i.e., the information that only emerges from learning from the signal about both alternatives together is positive. In contrast, if $\mathcal{I}_f(V_1, V_2 | \mathbf{S}) - \mathcal{I}_f(V_1, V_2) < 0$, the joint information is negative as evaluating both alternatives together is less informative after receiving the signal than before.

Corollary 1 generalizes the Shannon entropy based information cost functions utilized in the literature to non-uniform information cost structures. With this total cost function, the consumer is able to optimize on her information channels and to learn about each alternative up to an optimal extent. She does so by selecting her information strategy f which solves the following problem.

⁷ For this reason, there is no notion of mutual information between three or more variables (Cover and Thomas 2006).

OPTIMIZATION PROBLEM OF THE CONSUMER. Find a joint distribution f on \mathbf{S} and \mathbf{V} that solves the following:

$$\max_{f \in \Delta(\mathbb{R}^{2n})} R(f) - C(f) \quad (11)$$

$$\text{s.t.} \quad \int_{\mathbf{s} \in \mathbb{R}^n} f_{\mathbf{S}}(d\mathbf{s}, \mathbf{v}) = f(\mathbf{v}) \quad \text{for all } \mathbf{v} \in \mathbb{R}^n \quad (12)$$

where $R(f)$ and $C(f)$ are given in equations (2) and (7), respectively.⁸

4. Optimal Choice

In this section, we solve the optimization model given in (11)-(12), and describe the ensuing optimal choice probabilities. To this end, we follow the standard approach in rational inattention literature. We first show that an optimal information strategy would generate only one posterior belief for each alternative. In other words, if an information strategy were to lead to the choice of a particular alternative under distinct posteriors, the consumer would have processed “unnecessary” information, and hence such a strategy would not be optimal. This enables us to restate the consumer’s problem in terms of conditional choice probabilities directly. Solving this problem yields the characterization of the optimal choice behavior.

For a given information strategy f , we define the set of signals that lead to the choice of alternative i as

$$S_i(f) := \left\{ \mathbf{s} \in \mathbb{R}^n \mid i \in \arg \max_{i \in \mathcal{A}} \mathbb{E}_{f_{\mathbf{V}|\mathbf{s}}} [v_i] \right. \\ \left. \text{and } i \text{ is chosen according to the tie-breaking rule if } \# \arg \max_{i \in \mathcal{A}} \mathbb{E}_{f_{\mathbf{V}|\mathbf{s}}} [v_i] > 1. \right\}$$

Accordingly, we calculate the conditional choice probability for alternative i given state \mathbf{v} as

$$p_f(i | \mathbf{v}) := \int_{\mathbf{s} \in S_i} f_{\mathbf{S}|\mathbf{v}}(d\mathbf{s}),$$

and the unconditional probability of choosing i as

$$p_f(i) := \int_{\mathbf{v} \in \mathbb{R}^n} p_f(i | \mathbf{v}) p(d\mathbf{v}).$$

Suppose that two distinct signals \mathbf{s}' and \mathbf{s}'' result in the choice of the same alternative i , i.e., $\mathbf{s}', \mathbf{s}'' \in S_i$, with different posteriors under an optimal information strategy f^* , i.e., $f_{\mathbf{V}|\mathbf{s}'}^* \neq f_{\mathbf{V}|\mathbf{s}''}^*$. Now consider another information strategy, say \tilde{f} , which is identical to f^* for all $\mathbf{s} \notin \{\mathbf{s}', \mathbf{s}''\}$. Under \tilde{f} , however, the consumer receives signal \mathbf{s}' whenever she would have received either \mathbf{s}' or \mathbf{s}'' under f^* . Note that if the consumer chooses alternative i with strategy f^* , she would also make the same choice with \tilde{f} . In Lemma 1, we formally prove that such an f^* cannot be an optimal information strategy, since \tilde{f} pays off at least as much as f^* , i.e., $R(\tilde{f}) \geq R(f^*)$, and costs less, i.e., $C(\tilde{f}) \leq C(f^*)$.

⁸ Here it is implicit that the consumer makes the optimal choice based on her posterior belief. In case the consumer is indifferent between two alternatives, she uses an arbitrary rule to break any such ties.

LEMMA 1. *Let f^* be optimal and let $i \in A$ be such that $p_{f^*}(i | \mathbf{v}) > 0$. Then, for all signals $\mathbf{s}', \mathbf{s}'' \in S_i$ the posterior beliefs are identical, i.e., $f_{\mathbf{V}|\mathbf{s}'}^* = f_{\mathbf{V}|\mathbf{s}''}^*$, up to a set of measure zero under f^* .*

Lemma 1 suggests that the consumer processes information *efficiently* under the optimal strategy, and essentially needs only a *single* signal per choice alternative. Based on this result, we can express the consumer's objective given in (11) directly as a function of the choice probabilities, without referring to the signal space \mathbf{S} and other intricate details of the information strategy f . Particularly, we focus our attention on those strategies with n distinct signals s_1, \dots, s_n ⁹. Since receiving signal s_i is equivalent to choosing alternative i , i.e., $f_{\mathbf{S}|\mathbf{v}}(s_i) = p_f(i | \mathbf{v})$, the payoff $R(f)$, previously defined in (2), can now be expressed as

$$R(f) = \sum_{s_i \in \mathbf{S}} U(f_{\mathbf{V}|s_i}) \cdot \int_{\mathbf{v} \in \mathbb{R}^n} p_f(i | \mathbf{v}) p(d\mathbf{v}) = \sum_{s_i \in \mathbf{S}} \int_{\mathbf{v} \in \mathbb{R}^n} v_i f_{\mathbf{V}|s_i}(\mathbf{v}) d\mathbf{v} \cdot p_f(i)$$

Note that the term $p_f(i) \cdot f_{\mathbf{V}|s_i}(\mathbf{v})$ specifies the joint probability that the state of nature \mathbf{V} assumes the value \mathbf{v} and that the consumer chooses alternative i . Hence, we also have $p_f(i) \cdot f_{\mathbf{V}|s_i}(\mathbf{v}) = p_f(i | \mathbf{v}) \cdot p(\mathbf{v})$. Then, we can rewrite $R(f)$ as just a function of the conditional choice probabilities

$$R(f) = \sum_{i \in \mathcal{A}} \int_{\mathbf{v} \in \mathbb{R}^n} v_i p_f(i | \mathbf{v}) p(d\mathbf{v}). \quad (13)$$

Similarly, we can restate the cost $C(f)$ given in (7) as

$$C(f) = \sum_{i \in \mathcal{A}} \lambda_i \cdot \mathcal{I}_f(V_i, \mathbf{A} | \mathbf{V}_{1..i-1}) \quad (14)$$

where \mathbf{A} denotes the random variable that takes value $i \in \mathcal{A}$ with probability $p_f(i)$. Here, we once again recognize that the consumer's choice behaviour, captured by the random variable \mathbf{A} , is as informative about the state of nature \mathbf{V} , as the signal space \mathbf{S} leading to the consumer's choices, i.e., $\mathcal{I}_f(V_i, \mathbf{S} | \mathbf{V}_{1..i-1}) = \mathcal{I}_f(V_i, \mathbf{A} | \mathbf{V}_{1..i-1})$. As a result, using (13) and (14), we can express the objective function (11) directly as a function of the choice probabilities without making any implicit reference to the information strategy (even in the form of a subscript). The next proposition presents the resulting formulation to the consumer's choice problem.

PROPOSITION 1. *The set of conditional choice probabilities $\{p(i | \mathbf{v})\}_{i \in \mathcal{A}}$ leads to the optimal information strategy of an inattentive consumer (optimally solves the problem in (11)-(12)), if and only if it is a solution to the following problem*

$$\max_{\{p(i|\mathbf{v})\}_{i \in \mathcal{A}}} \sum_{i \in \mathcal{A}} \int_{\mathbf{v} \in \mathbb{R}^n} v_i p(i | \mathbf{v}) p(d\mathbf{v}) - \sum_{j \in \mathcal{A}} \lambda_j \cdot \mathcal{I}_p(V_j, \mathbf{A} | \mathbf{V}_{1..j-1}) \quad (15)$$

⁹ The exact argument does allow for multiple signals leading to i , $|S_i| \geq 1$, and is given in the appendix.

subject to

$$\begin{aligned} p(i | \mathbf{v}) &\geq 0 \quad \text{for all } i \in \mathcal{A} \text{ and } \mathbf{v} \in \mathbb{R}^n \\ \sum_{i \in \mathcal{A}} p(i | \mathbf{v}) &= 1, \end{aligned}$$

where

$$\mathcal{I}_p(V_j, \mathbf{A} | \mathbf{V}_{1..j-1}) = - \sum_{i \in \mathcal{A}} \int_{\mathbf{v}_{1..j} \in \mathbb{R}^j} p(i | \mathbf{v}_{1..j}) (\log p(i | \mathbf{v}_{1..j-1}) - \log p(i | \mathbf{v}_{1..j})) p(d\mathbf{v}_{1..j}) \quad (16)$$

and

$$p(i | \mathbf{v}_{1..j}) = \int_{\mathbf{v}_{j+1..n} \in \mathbb{R}^{n-j}} p(i | \mathbf{v}_{1..j} \mathbf{v}_{j+1..n}) p(\mathbf{v}_{j+1..n} | \mathbf{v}_{1..j}) d\mathbf{v}_{j+1..n}.$$

Next, we consider how one can go about solving this alternative formulation given in Proposition 1. First note that this formulation is a concave maximization problem on a compact set. The objective function in (15) is concave in $\{p(i | \mathbf{v})\}_{i \in \mathcal{A}}$, since $\mathcal{I}_p(V_j, \mathbf{A} | \mathbf{V}_{1..j-1}) = \mathcal{D}(p(i, \mathbf{v}_{1..j}) \| p(i, \mathbf{v}_{1..j-1}) \cdot p(\mathbf{v}_{1..j}))$ is convex in $\{p(i | \mathbf{v})\}_{i \in \mathcal{A}}$. Here, \mathcal{D} denotes the relative entropy, which is convex in the pair $(p(i, \mathbf{v}_{1..j}), p(i, \mathbf{v}_{1..j-1}) \cdot p(\mathbf{v}_{1..j}))$ and hence in $p(i, \mathbf{v})$ (c.f. Cover and Thomas 2006, Theorem 2.7.2). Moreover, as we show in the Appendix, $p(i | \mathbf{v}_{1..j}) = 0$ only if $p(i) = 0$. Treating this case separately, the optimization problem essentially has one equality constraint, which allows us to obtain the structure of the optimal solution from the first order conditions of the Lagrangian. We present this result next in Theorem 2.

THEOREM 2. *For any information cost structure $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n < \infty$, the consumer forms her information strategy such that the optimal conditional choice probabilities satisfy*

$$p(i | \mathbf{v}) = \frac{e^{\frac{v_i}{\lambda_n}} \cdot p(i)^{\frac{\lambda_1}{\lambda_n}} \cdot \prod_{k=1}^{n-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_n}}}{\sum_{j=1}^n e^{\frac{v_j}{\lambda_n}} \cdot p(j)^{\frac{\lambda_1}{\lambda_n}} \cdot \prod_{k=1}^{n-1} p(j | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_n}}} \quad (17)$$

almost surely.

We remark that (17) is derived for alternatives with a positive probability of being chosen, i.e., $p(i) > 0$, but it trivially holds for alternatives never chosen, i.e., $p(i) = 0$, as well. Notice also that the conditional choice probability in (17) resembles the GMNL formula – in fact, when information costs are identical, (18) becomes the exact GMNL expression in (4). It is more general in the sense that when information costs differ, the conditional choice probabilities are not only driven by state \mathbf{v} , prior beliefs $p(\mathbf{v})$ and unconditional probabilities $p(i)$, but also by the *partially* conditional choice probabilities $p(j | \mathbf{v}_{1..k})$'s of selecting each alternative. To gain a better understanding of the implications, it is useful to rewrite (17) as

$$p(i | \mathbf{v}) = \frac{e^{\frac{v_i}{\lambda_n} + \alpha_i}}{\sum_{j \in \mathcal{A}} e^{\frac{v_j}{\lambda_n} + \alpha_j}}, \quad (18)$$

where we define α_i as

$$\alpha_i = \frac{\lambda_1}{\lambda_n} \cdot \log p(i) + \sum_{k=1}^{n-1} \frac{\lambda_{k+1} - \lambda_k}{\lambda_n} \cdot \log p(i | \mathbf{v}_{1..k}).$$

Written this way, the conditional probabilities follow a formula similar to the standard MNL, with the pay-off of each alternative shifted by the term α_i . For the GMNL, α_i simply equals $\log p(i)$, implying that if an alternative is in general attractive, i.e., $p(i)$ is relatively high, it can still be chosen with high probability even if its true value is low (Matějka and McKay 2015). When the information costs are different, the consumer will typically know more about the cheaper alternatives, and this is reflected into the computation of how “attractive” the alternative is. Specifically, the shift term is a weighted average of the log transformations of the unconditional and partially conditional choice probabilities. Hence, a generally attractive alternative (with a relatively high $p(i)$) can be chosen with a low probability if the information obtained from studying an alternative with low cost (say Alternative 1) implies a low selection probability $p(i | \mathbf{v}_1)$, even if the true value is high.

The next proposition restates the consumer’s choice problem in Proposition 1 using the structure of the conditional choice probabilities presented in Theorem 2.

PROPOSITION 2. *The set of conditional choice probabilities $\{p(i | \mathbf{v}_{1..n-1})\}_{i \in \mathcal{A}}$ solves the problem in Proposition 1 if and only if it is a solution to the following problem*

$$\max_{\{p(i | \mathbf{v}_{1..n-1})\}_{i \in \mathcal{A}}} \lambda_n \int_{\mathbf{v}} \log \left(\sum_{j=1}^n e^{\frac{v_j}{\lambda_n}} \cdot p(j)^{\frac{\lambda_1}{\lambda_n}} \cdot \prod_{k=1}^{n-1} p(j | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_n}} \right) p(d\mathbf{v}) \quad s.t.$$

$$p(i | \mathbf{v}_{1..n-1}) \geq 0 \quad \text{for all } i \in \mathcal{A} \text{ and } \mathbf{v}_{1..n-1} \in \mathbb{R}^{n-1},$$

$$\sum_{i \in \mathcal{A}} p(i | \mathbf{v}_{1..n-1}) = 1.$$

We should remark that the formulation in Proposition 2 offers a significant simplification over the initial formulation in (11)-(12) for finding the optimal information strategy. Nevertheless, solving for the optimal $\{p(i | \mathbf{v}_{1..n-1})\}_{i \in \mathcal{A}}$ can still be quite challenging when there are many alternatives and possible realizations of the values. Therefore, in what follows, we first examine some limiting cases and then present some illustrative examples to better understand the inattentive consumer’s choice behavior facing alternatives with non-uniform information costs.

5. Limiting Scenarios and Choice Behavior

In the previous section, we characterized the optimal behaviour of customers for the most general case involving distinct alternatives with non-zero and finite information costs. There are some limiting scenarios that do not directly follow from the conditional choice probability equation (17) in

Theorem 2. In this section, we focus on four such scenarios – infinite and zero information cost for some alternatives, and duplicate and dominated alternatives. Delving into these limiting cases also sheds some light on how non-uniform information costs impact the choices of inattentive consumers. To keep the exposition simple, we assume that the consumer chooses one of three alternative products – PRODUCT 1, PRODUCT 2, and PRODUCT 3 – with information costs of $\lambda_1 \leq \lambda_2 \leq \lambda_3$, respectively (we provide the formulae for the general case in Appendix B). The conditional choice probabilities for these three alternatives can be expressed as:

$$p(i | v_1, v_2, v_3) = \frac{e^{\frac{v_i}{\lambda_3}} \cdot p(i)^{\frac{\lambda_1}{\lambda_3}} \cdot p(i | v_1)^{\frac{\lambda_2 - \lambda_1}{\lambda_3}} \cdot p(i | v_1, v_2)^{\frac{\lambda_3 - \lambda_2}{\lambda_3}}}{\sum_{j \in \{1, 2, 3\}} e^{\frac{v_j}{\lambda_3}} \cdot p(j)^{\frac{\lambda_1}{\lambda_3}} \cdot p(j | v_1)^{\frac{\lambda_2 - \lambda_1}{\lambda_3}} \cdot p(j | v_1, v_2)^{\frac{\lambda_3 - \lambda_2}{\lambda_2}}} \quad (19)$$

for all $i \in \mathcal{A} = \{1, 2, 3\}$ and all $\mathbf{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$.

INFINITE INFORMATION COST. Suppose that it is infinitely costly (or prohibitively expensive) for the consumer to process information about PRODUCT 3. This could represent a product for which the customer is not willing to acquire any information, or for which such information is not obtainable (e.g. product is not offered/available). Since, $\lambda_3 = \infty$ and

$$C(f) = \lambda_1 \mathcal{I}_f(V_1, \mathbf{S}) + \lambda_2 \mathcal{I}_f(V_2, \mathbf{S} | V_1) + \lambda_3 \mathcal{I}_f(V_3, \mathbf{S} | V_1, V_2),$$

the consumer would have to set $\mathcal{I}_f(V_3, \mathbf{S} | V_1, V_2) = 0$ under her optimal information strategy (to avoid an infinite information processing cost). Accordingly, the consumer does *not* update her priors beyond the information about PRODUCT 2. Then, from (16),

$$\begin{aligned} \mathcal{I}_f(V_3, \mathbf{S} | V_1, V_2) = 0 &\Rightarrow \log p(i | v_1, v_2) - \log p(i | v_1, v_2, v_3) = 0 \\ &\Rightarrow p(i | v_1, v_2, v_3) = p(i | v_1, v_2). \end{aligned}$$

The above expression implies that the conditional choice probabilities do *not* depend on v_3 . However this does not mean that she makes decisions about PRODUCT 3 based on prior beliefs only. In fact, the customer updates her *expectation* of the value v_3 on the basis of information learned about PRODUCT 1 and PRODUCT 2 values (v_1, v_2) . Defining this conditional expectation as $\bar{v}_3(v_1, v_2) = \mathbb{E}_{p_{V_3|v_1, v_2}}[v_3]$, the conditional choice probabilities in (19) can be rewritten as

$$p(i | v_1, v_2, v_3) = \frac{q(i | v_1, v_2)}{\sum_{j \in \{1, 2, 3\}} q(j | v_1, v_2)} \quad \text{for all } i \in \{1, 2, 3\},$$

where

$$\begin{aligned} q(1 | v_1, v_2) &= e^{\frac{v_1}{\lambda_2}} \cdot p(1)^{\frac{\lambda_1}{\lambda_2}} \cdot p(1 | v_1)^{\frac{\lambda_2 - \lambda_1}{\lambda_2}}, \\ q(2 | v_1, v_2) &= e^{\frac{v_2}{\lambda_2}} \cdot p(2)^{\frac{\lambda_1}{\lambda_2}} \cdot p(2 | v_1)^{\frac{\lambda_2 - \lambda_1}{\lambda_2}}, \\ q(3 | v_1, v_2) &= e^{\frac{\bar{v}_3(v_1, v_2)}{\lambda_2}} \cdot p(3)^{\frac{\lambda_1}{\lambda_2}} \cdot p(3 | v_1)^{\frac{\lambda_2 - \lambda_1}{\lambda_2}}. \end{aligned}$$

ZERO INFORMATION COST (AND DETERMINISTIC ALTERNATIVES). Suppose that the consumer can freely process all information for PRODUCT 1, i.e., $\lambda_1 = 0$. This could represent a product for which the customer can assign a true value very easily (e.g. a simple search good). Then, taking the limit of (19) with $\lambda_1 \rightarrow 0$, we get

$$p(i | v_1, v_2, v_3) = \frac{e^{\frac{v_i}{\lambda_3}} \cdot p(i | v_1)^{\frac{\lambda_2}{\lambda_3}} \cdot p(i | v_1, v_2)^{\frac{\lambda_3 - \lambda_2}{\lambda_3}}}{\sum_{j \in \{1,2,3\}} e^{\frac{v_j}{\lambda_3}} \cdot p(j | v_1)^{\frac{\lambda_2}{\lambda_3}} \cdot p(j | v_1, v_2)^{\frac{\lambda_3 - \lambda_2}{\lambda_3}}} \quad \text{for all } i \in \{1, 2, 3\}.$$

Note that if the value of PRODUCT 1 is deterministic, i.e., $V_1 = v_1$ with probability 1, then $p(i) = p(i | v_1)$ even for $\lambda_1 > 0$. Accordingly, (19) also reduces to the above expression for deterministic alternatives. This could represent a product about which the consumer is well-informed due to past experience, or the no-purchase alternative (reservation value). We remark however that only the functional form of the conditional choice probabilities are the same; the actual values of the conditional probabilities and resulting behavior are likely to differ. This is because learning about a deterministic alternative reveals no information about other alternatives. In contrast learning the exact value of an alternative can reveal significant information about other alternatives. More precisely, for a deterministic alternative, V_1 takes only one value, whereas alternatives with zero information cost can potentially take an arbitrary number of values. Hence the partially conditional purchase probabilities involving V_1 (e.g. $p(i | v_1, v_2)$) would average over all these potential values.

Next, we provide an example involving the above limiting scenarios to demonstrate the impact on consumer choice.

Example 1 The consumer is choosing from PRODUCT 1, PRODUCT 2 and PRODUCT 3. There are 3 possible states, and let $\mathbf{v} \in \{(1, 0, 0), (0, 1, 0), (0, 0, \frac{1}{2})\}$ and $p(\mathbf{v}) = \frac{1}{3}$ for all \mathbf{v} . Table 1 depicts the unconditional choice probability of choosing each alternative for different levels of information costs.

Table 1 Information costs and unconditional choice probabilities

Case #	λ_1	λ_2	λ_3	$p(1)$	$p(2)$	$p(3)$
0	0	0	0	1/3	1/3	1/3
1	∞	∞	∞	0.5	0.5	0
2	0.4	∞	∞	0.41	0.59	0
3	0.4	0.4	∞	0.44	0.44	0.12
4	0	0.4	∞	1/3	0.44	0.23
5	0	0	∞	1/3	1/3	1/3
6	0.4	0.4	0.4	0.44	0.44	0.12

Observe that each alternative is the best only in one of the three states, so under full information (Case 0), $p(i) = \frac{1}{3}$ for $i = 1, 2, 3$. In contrast, if the consumer does not process *any* information at all

(Case 1), then she would choose either PRODUCT 1 or PRODUCT 2, but never PRODUCT 3 since its expected value is lower than the other two (actually she is indifferent to PRODUCT 1 and 2, and here we assume she chooses each one with equal probability). At a first glance, it might seem intuitively appealing that reducing the information cost of either PRODUCT 1 or PRODUCT 2 should increase its choice probability, since the consumer would be able to more confidently assess it as the better of the two alternatives. However, this is not correct, as seen in Case 2. When $\lambda_1 = 0.4$, $\lambda_2 = \lambda_3 = \infty$, the consumer chooses PRODUCT 1 less ($p(1) < p(2)$). This is because with reduced information cost, the consumer is also able to learn with more confidence the states in which PRODUCT 1 is *not* the best alternative. Note that PRODUCT 3 is still not chosen since in expectation it is inferior to PRODUCT 2. However, this changes when the information cost of PRODUCT 2 is also reduced, as in Case 3 with $\lambda_1 = \lambda_2 = 0.4$ and $\lambda_3 = \infty$. Now, the consumer selects PRODUCT 1 and 2 with equal probability. More interestingly, even though she does *not* process any information about PRODUCT 3 *directly*, from the *implied* information obtained through PRODUCT 1 and PRODUCT 2, she is able to identify cases where it is likely to be the best alternative. As a result, $p(3)$ increases to 0.125.

Let us now explore the impact of freely available information. As seen in Case 4, when the consumer knows the exact value of PRODUCT 1, she chooses it only when $v_1 = 1$ (hence $p(1)$ further reduces to $1/3$). On the other hand, conditional on $v_1 = 0$, the expected value of PRODUCT 2 is higher than that of PRODUCT 3 and given that she processes information on PRODUCT 2, she chooses it with higher probability. Interestingly, the fact that she has perfect information about PRODUCT 1 mainly benefits PRODUCT 3. This is because in this example there are three products and only three states. Hence, it is sufficient for the DM to evaluate PRODUCT 1 and PRODUCT 2 independently (form opinions about each taking a non-zero value). Given that there are only 3 states possible, she need not process additional information to learn about PRODUCT 3. Consequently, knowing more confidently that she is facing states 2 or 3 does not significantly alter how she will process information about PRODUCT 2 (λ_2 is the same in Cases 3 and 4) - she still has to evaluate PRODUCT 2 and form opinion about its value being 1 or 0. Of course, if the cost of information for PRODUCT 2 reduces, she will be able to make this assessment better and the conditional choice probability for PRODUCT 2 would decrease, as confirmed by Case 5. The fact that the consumer does not need to process information about PRODUCT 3 beyond what she learns and infers through PRODUCT 1 and 2 is substantiated by Case 6. Comparing it with Case 3 confirms that λ_3 has no impact on the choice probability for this example (as long as it is equal or higher than λ_2).

DUPLICATE ALTERNATIVES. It has been shown that when the information costs are uniform, the resulting choice behavior of rationally inattentive consumers (i.e. GMNL) does not suffer from the IIA property. Specifically, Matějka and McKay (2015) establish that duplicate alternatives are

jointly treated as one alternative. Two alternatives are referred to as “duplicates” if in the prior beliefs of the customers, the probability of the two alternatives taking different values is zero in all states of the world. Recognizing that information costs of duplicate alternatives can indeed be different, we now explore the extension of this result to the case of differentiated information costs.

To this end, we investigate what happens when a duplicate alternative is added to the choice set. Suppose that a new product, say PRODUCT $\hat{2}$, is added to the existing set of products $\mathcal{A} = \{1, 2, 3\}$. Let $\hat{\mathcal{A}}$ denote the choice set which includes Product $\hat{2}$, i.e., $\hat{\mathcal{A}} = \mathcal{A} \cup \{\hat{2}\}$. Moreover, define $\hat{p}(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^4$, as the prior belief of the inattentive consumer choosing a product in $\hat{\mathcal{A}}$. The prior $\hat{p}(\mathbf{u})$ is obtained from $p(\mathbf{v})$, $\mathbf{v} \in \mathbb{R}^3$, by simply duplicating PRODUCT $2 \in \mathcal{A}$, i.e., $p(\mathbf{v})$ is the marginal distribution of $\hat{p}(\mathbf{u})$ with respect to $u_{\hat{2}}$. Note that since the values of the duplicate products in $\hat{\mathcal{A}}$ are fully correlated, the consumer will only process information about the cheaper cost duplicate and this would yield the exact same information about the other duplicate. Hence, even if the individual information costs might differ, the consumer remains indifferent between PRODUCT 2 and PRODUCT $\hat{2}$ in $\hat{\mathcal{A}}$. Consequently, the probability of choosing PRODUCT 2 *or* PRODUCT $\hat{2}$ among available alternatives in $\hat{\mathcal{A}}$ *exactly equals* the choice probability for PRODUCT 2 among alternatives in \mathcal{A} , provided that it is *not* cheaper for the consumer to process information about PRODUCT $\hat{2}$ than about PRODUCT 2, i.e., if $\lambda_{\hat{2}} \geq \lambda_2$. Proposition 3 formally presents this result for the general case with n alternatives.

PROPOSITION 3. *Let $\{p(i | \mathbf{v})\}_{i \in \mathcal{A}}$ for all $\mathbf{v} \in \mathbb{R}^n$ be the set of optimal conditional choice probabilities for an inattentive consumer choosing from available alternatives in \mathcal{A} . If a new alternative \hat{j} , which is the duplicate of an existing alternative $j \in \mathcal{A}$, with information cost $\lambda_{\hat{j}} \geq \lambda_j$ is also offered to the consumer, then the set of optimal conditional choice probabilities, $\{\hat{p}(i | \mathbf{u})\}_{i \in \mathcal{A} \cup \hat{j}}$ for all $\mathbf{u} \in \mathbb{R}^{n+1}$, satisfies*

$$\hat{p}(i | \mathbf{u}) = p(i | \mathbf{v}) \quad \text{for all } i \notin \{j, \hat{j}\}$$

and

$$\hat{p}(j | \mathbf{u}) + \hat{p}(\hat{j} | \mathbf{u}) = p(j | \mathbf{v}),$$

whenever $v_i = u_i$ for all $i \notin \{j, \hat{j}\}$.

We should mention that analogous equalities also hold for the unconditional choice probabilities in the above proposition. Further, the requirement that $\lambda_{\hat{j}} \geq \lambda_j$ is critical – if processing information about the duplicate is cheaper than about the original alternative, the optimal choice may differ since more information is likely to be processed due to the availability of a cheaper information source. The above also implies that when duplicates exist in the choice set, it is possible to drop the ones with higher information costs from the set and determine the choice probability for the

cheapest cost alternative of the duplicates. This choice probability can then be allocated over all duplicate alternatives arbitrarily.

DOMINATED ALTERNATIVES. A closely related notion to duplicates is dominated alternatives. A dominated alternative is one whose value is lower than another alternative in all states of the world. Such alternatives are never selected when the cost of information is uniform across alternatives (Matějka and McKay 2015). It can be easily verified that this extends to the more general case of differentiated information costs (shifting the choice probability to the dominating alternative would increase the consumer’s objective function). This does not mean information is not processed about a dominated alternative. As a matter of fact, whether a dominated alternative is available or not in the choice set can become relevant in the case of non-uniform information costs since it might serve as a cheap channel to learn about other alternatives, as is illustrated by the following example.

Example 2 (Strong Failure of Regularity) Consider a consumer visiting a retail store to buy a certain product (e.g. a smart phone) and choosing between either PRODUCT 1 or PRODUCT 2. She is well-informed about PRODUCT 1 based on past experience, which leads to a safe (deterministic) valuation of 1.0 as given in Table 2. On the other hand, PRODUCT 2 might be better or worse than PRODUCT 1 such that there are two equally likely states (*State 1* and *State 2*) corresponding to a higher and a lower value for PRODUCT 2 (either 1.2 or 0.6). Suppose that exploring PRODUCT 2 in the store is difficult (e.g. lack of sample product on display). Accordingly, let the information cost be $\lambda_2 = 1$. At this cost, she will not acquire information about PRODUCT 2 and simply buy PRODUCT 1 (refer to Table 2). Now suppose that the store adds PRODUCT 3 to its assortment. Although PRODUCT 3 has most of the key features of PRODUCT 2 (e.g. operating system), it is inferior in some other elements (e.g. build quality, camera resolution). Overall, PRODUCT 2 dominates PRODUCT 3 in both *State 1* and *State 2*. Assume that a sample for PRODUCT 3 is readily available in the store for customers to inspect, hence the consumer can learn about it at cost $\lambda_3 = 0.2$. Subsequently, the consumer also processes information about PRODUCT 2 by mainly inferring through PRODUCT 3. Although the consumer never chooses the dominated PRODUCT 3, the unconditional purchase probability of PRODUCT 2 increases from 0 to 0.29.

Table 2 Values and unconditional choice probabilities when an inferior product is added

PRODUCT	Value in		Unconditional Choice Probabilities when	
	<i>State 1</i>	<i>State 2</i>	$\mathcal{A} = \{1, 2\}$	$\mathcal{A} = \{1, 2, 3\}$
1	1	1	1.00	0.71
2	0.6	1.2	0.00	0.29
3	0.5	0.9	–	0.00

The above example highlights failure of the regularity condition put forth by Luce and Suppes (1965), which requires that adding a product to the choice set does *not* increase the market share of another product. Matějka and McKay (2015) show that a rationally inattentive consumer facing uniform information costs might fail the regularity condition. This is because introducing a new product can set incentives for the consumer to get information about the new product in a way that she is also informed about a previously “uninteresting” product. With this additional information, she might identify cases where she buys the previously uninteresting product. If the new product is inferior (i.e., dominated), however, the consumer would completely disregard the new product and also would not process any information about it. Hence, there is *no failure of regularity* under uniform information costs when the added product is dominated. In contrast, we show that non-uniform information costs can induce failure of regularity even if the inclusion is an inferior, never-selected alternative (hence our usage of the term *strong* failure of regularity).

There is a long standing discussion about the concept of failure of regularity. One explanation is the asymmetric dominance effect (Huber et al. 1982), which requires that the inferior product is only dominated by one of the alternative products. Note that this is not the case in our example; PRODUCT 3 is dominated by both PRODUCT 2 and PRODUCT 1. Another popular explanation is extremeness aversion (Simonson and Tversky 1992), where it is assumed that the consumer buys from the middle. Also menu/halo effects, where the addition of a product signals the quality of the producer, can induce strong failure of regularity (Luce and Raiffa 1957). Such explanations suggest that the consumer’s preferences somehow “change” by the introduction of further alternatives. In stark contrast, in our model, the consumer updates her belief about the value of the alternatives in a way that is completely consistent with her prior.¹⁰ Rather, it is the rational inattentiveness of the customer, coupled with asymmetric information costs that leads to the strong failure of regularity. To the best of our knowledge, this is a new explanation for this well-documented phenomenon.

6. General Choice Behavior: RED-BUS/Blue-Bus

We now demonstrate the choice behavior of rationally inattentive customers under a general setting with different degrees of correlations among the products and non-uniform information costs. For this purpose we consider the classic red-bus/blue-bus problem, and adopt the primary setup in Matějka and McKay (2015). The inattentive consumer, whom we refer to as the decision-maker (DM) in this particular context, faces three alternatives – she may take the TRAIN (T), the BLUE BUS (B), or the RED BUS (R). Table 3 gives the four possible states of nature $\mathbf{v} = (v_T, v_R, v_B)$, and prior belief of the decision-maker about each state, where ρ denotes the correlation between

Table 3 Possible states and prior beliefs

	State 1	State 2	State 3	State 4
TRAIN	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
RED BUS	0	1	0	1
BLUE BUS	0	0	1	1
$p(\mathbf{v})$	$\frac{1}{4}(1+\rho)$	$\frac{1}{4}(1-\rho)$	$\frac{1}{4}(1-\rho)$	$\frac{1}{4}(1+\rho)$

the values of the two buses (values of 0 and 1 indicate that the particular bus is “slow” or “fast”, respectively).

Note that the speed of the TRAIN is deterministic ($v_T = \frac{1}{2}$), and the expected values of all three alternatives are equal to $\frac{1}{2}$. If the decision-maker were to choose an alternative *without* processing any information, she would be indifferent between the three alternatives, i.e., $p(T) = p(R) = p(B) = \frac{1}{3}$. On the other hand, if the decision-maker could process information freely, she would always choose the fastest option, i.e., $p(T) = \frac{1}{4}(1+\rho)$ and $p(R) = p(B) = \frac{1}{8}(3-\rho)$ (see Figure 2a). Clearly, since the buses are symmetric, they are chosen with equal probability, and this probability decreases as ρ increases (the DM believes increasingly that the buses are similar in speed).

Figure 2b illustrates the subtle changes in the DM’s choice behaviour when processing information is costly but the costs are *symmetric* (uniform), i.e., $\lambda_R = \lambda_B = 0.4$. As expected, the two buses are still always selected with equal probabilities, which are non-increasing in ρ . Furthermore, both buses are selected more by the DM, compared to the case with freely available information. In particular, when the DM has sufficiently strong belief that *one* of the buses must be fast, she does not consider the TRAIN as an option (this happens when $\rho \leq -0.31$). Given that information processing is costly, due to her beliefs that the TRAIN is unlikely to be the best alternative, she instead allocates all her time and attention to understand which bus is faster (since the buses are symmetric, each bus is chosen with 50% chance).

Now consider the scenario with *asymmetric* (non-uniform) information costs. Specifically, suppose that acquiring and processing information about the RED BUS is less expensive than about the BLUE BUS for the DM, i.e., $\lambda_R \leq \lambda_B$. Based on Proposition 2, the DM’s choice is determined by the solution to the following optimization problem

$$\max_{\{p(i|v_T, v_B)\}_{i \in A}} \lambda_B \sum_{\mathbf{v}} p(\mathbf{v}) \log \left(\sum_{i \in \{T, B, R\}} e^{\frac{v_i}{\lambda_B}} p(i | \frac{1}{2})^{\frac{\lambda_R}{\lambda_B}} p(i | (\frac{1}{2}, v_T))^{\frac{\lambda_B - \lambda_R}{\lambda_B}} \right) \quad s.t.$$

$$p(i | v_T, v_B) \geq 0 \quad \text{for } i \in \{T, B, R\} \text{ and } (v_T, v_B) \in \{(\frac{1}{2}, 0), (\frac{1}{2}, 1)\},$$

$$\sum_{i \in A} p(i | v_T, v_B) = 1 \quad \text{for } (v_T, v_B) \in \{(\frac{1}{2}, 0), (\frac{1}{2}, 1)\},$$

¹⁰ Even if the information cost of PRODUCT 3 approaches zero, PRODUCT 2 is not chosen more than 50% of the time.

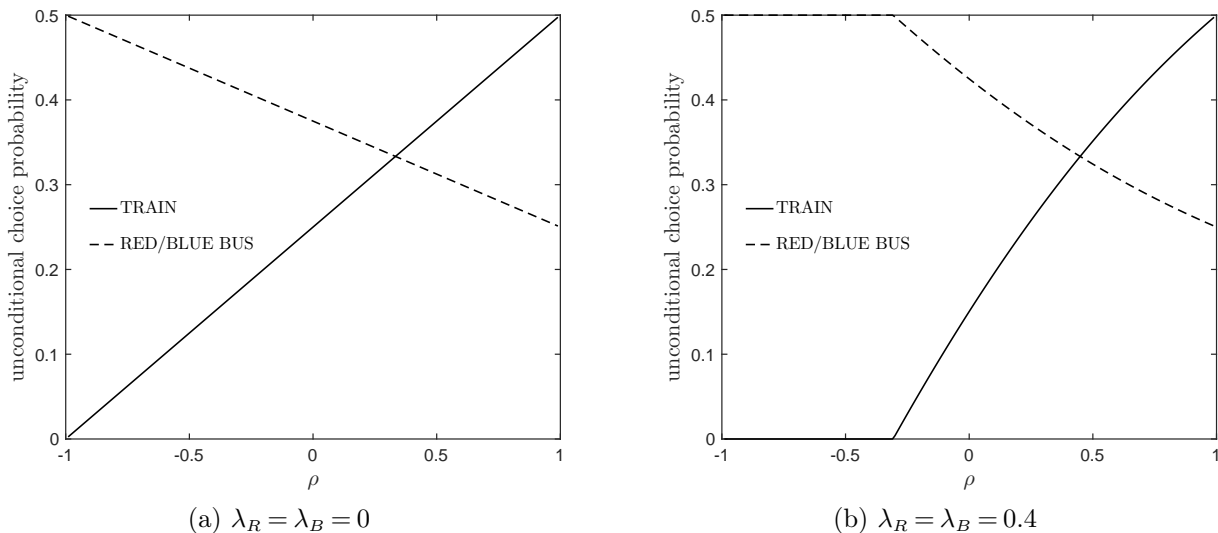


Figure 2 RED BUS/BLUE BUS when information costs are symmetric

which can be solved easily by numerical search.

Figure 3a shows the unconditional choice probabilities when $\lambda_R = 0.2$ and $\lambda_B = 0.4$. We observe, as in the case of uniform information costs, that the TRAIN is never selected if the DM has sufficient belief that the two bus speeds are negatively correlated (i.e., if one of the buses must be fast). Note that this occurs for a smaller range of beliefs/correlations compared to the case of uniform costs ($\rho \leq -0.47$ versus $\rho \leq -0.31$). Nevertheless, as the DM's prior belief that the two buses are similar gets stronger with ρ , she also starts selecting the TRAIN. Interestingly, in this range the DM builds a stronger preference for the “cheap” RED BUS over the BLUE BUS. This is because she acquires more information about the RED BUS and has more confidence about its speed compared to the BLUE BUS. In particular, as ρ approaches 1, the DM believes that the buses have identical speed. Consequently, whenever she decides to take a bus, she takes the RED BUS, on which she has more information. This signifies the importance of information provision for a seller in forming its product choice set. When the alternatives are very similar in the eyes of the DM, even a slight improvement in the provision of information for one product can significantly shift demand towards it. This is particularly stark considering that when $\rho = 1$, the DM treats duplicate alternatives jointly as one.

We can delineate the driving forces of this behavior by directly investigating the attention allocation and information acquisition strategies of the DM. For this purpose, Figure 3b depicts the mutual information and the conditional mutual information that is processed by the DM. Naturally, since the information is more readily available about the RED BUS, the DM learns more about it; i.e., $\mathcal{I}(V_R, \mathbf{S}) > \mathcal{I}(V_B, \mathbf{S})$. Furthermore, part of what the DM learns about the BLUE BUS is implied information, while the rest is direct information. The latter, which quantifies the amount of effort the DM spends to learn about the BLUE BUS beyond what she can infer from studying the RED

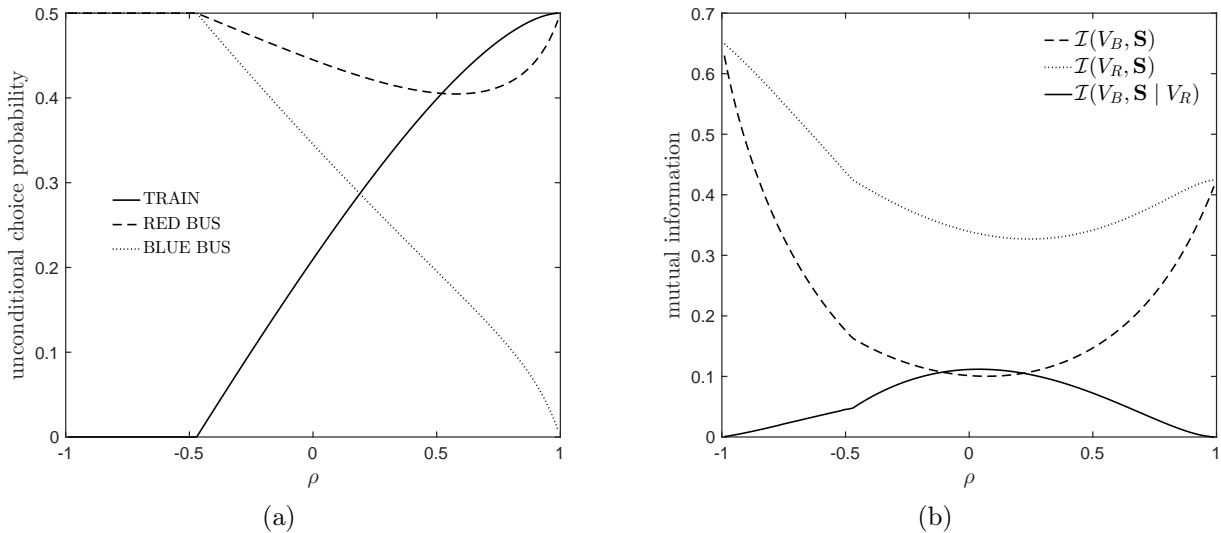


Figure 3 RED BUS/BLUE BUS when information costs are asymmetric ($\lambda_R = 0.2$ and $\lambda_B = 0.4$)

BUS, is the conditional mutual information $\mathcal{I}(V_B, \mathbf{S} | V_R)$. Observe that $\mathcal{I}(V_B, \mathbf{S} | V_R)$ approaches zero as $\rho \downarrow -1$ and $\rho \uparrow 1$. This is because, when $\rho \downarrow -1$, the DM closely studies the RED BUS and if she concludes that it must be fast, she simply takes it, whereas if she concludes that it must be slow, she automatically takes the BLUE BUS (since she simply infers and then strongly believes that it must be fast). On the other extreme, when $\rho \uparrow 1$, the DM closely studies the RED BUS and if she concludes that it is most probably fast, she simply takes it, whereas if she concludes that it must be slow, she automatically takes the TRAIN (since she infers that the BLUE BUS must be slow too). In contrast, when ρ is around zero, the decision-maker cannot infer sufficient information about the BLUE BUS through the information processed about the RED BUS, and decides to pay additional attention and acquire *direct* information about the BLUE BUS to make the optimal choice.

In order to deepen our understanding of the DM's choice behaviour when information costs are asymmetric, we next focus on the conditional choice probabilities given in Figures 4 and 5. As seen in Figure 4a, even when the TRAIN is the best alternative, it is not selected by the DM for sufficiently negative ρ , $\rho \leq -0.47$ (as previously explained). Moreover, in this range, her conditional belief for the RED BUS being slow yet the BLUE BUS being fast is decreasing in ρ . Therefore, the conditional choice probability of RED BUS (resp. BLUE BUS) increases (resp. decreases) in ρ . On the other hand, for $\rho > -0.47$, the TRAIN is also chosen and the decision-maker increasingly prefers the TRAIN and avoids the buses as ρ increases. Further she also learns that both buses are more likely to be slow mainly by processing direct information about the (cheaper) RED BUS. Hence, as long as $\rho \leq 0.60$, the DM takes the RED BUS less often than the BLUE BUS. Interestingly, for high levels of ρ , when the DM erroneously believes that the RED BUS is fast, she also infers that the BLUE BUS must also be fast but since she is more informed about the RED BUS, she takes it. In this case, the

BLUE BUS is rarely chosen. In contrast, when both buses are fast, i.e., $v_R = v_B = 1$, as in Figure 4b, the DM always prefers the RED BUS over the BLUE BUS due to its information advantage. In this case, she rarely makes the wrong decision by taking the TRAIN. Further, only if the correlation ρ is around zero, the DM processes direct information about both buses and may follow a positive signal concerning the BLUE BUS when receiving an erroneous negative signal about the RED BUS. The selection probability of the BLUE BUS is highest in this range, i.e., the information cost disadvantage for a fast BLUE BUS is low if little can be inferred about it from knowing the RED BUS.

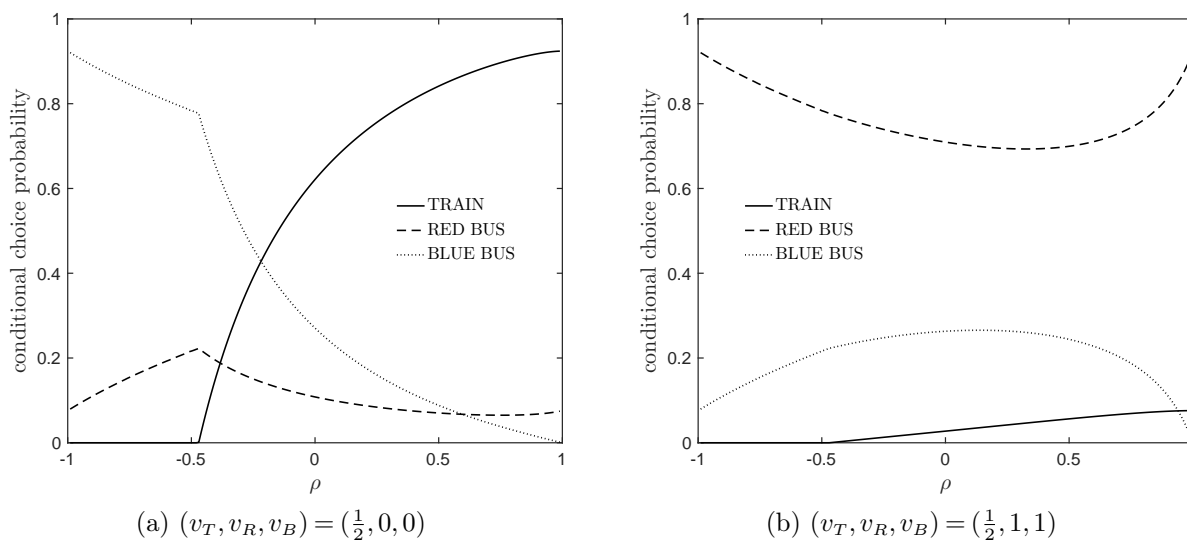


Figure 4 Conditional choice probabilities for RED BUS/BLUE BUS when the two buses are *equally fast*

Figure 5a confirms that the DM most often makes the right choice when information processing is cheap for the fast bus (RED BUS) and expensive for the slow bus (BLUE BUS), i.e., $v_R = 1$ and $v_B = 0$. As shown in Figure 5b, this remains the case even if it is harder to acquire information about the fast bus $v_R = 0$ and $v_B = 1$, provided that the DM has negatively correlated beliefs. When the DM increasingly believes the buses to be similar (ρ increases), however, the likelihood of taking the BLUE BUS decreases sharply. This is because the DM has more information about the slow RED BUS, and since the buses are very similar according to her beliefs, she draws the inference that the BLUE BUS must also be slow. She instead increasingly chooses the TRAIN (and makes the wrong decision). In the extreme case, $\rho \geq 0.95$ the likelihood of her taking the correct BLUE BUS is even less than the RED BUS.

From the above discussion, it is clear that reducing the information cost of an alternative results in more information to be processed by the DM. This means that the net-payoff to the DM also improves after optimal processing of information. However, it is also evident that this does not mean that the DM makes more correct choices all the time. To highlight this fact, we depict in Figure 6, the

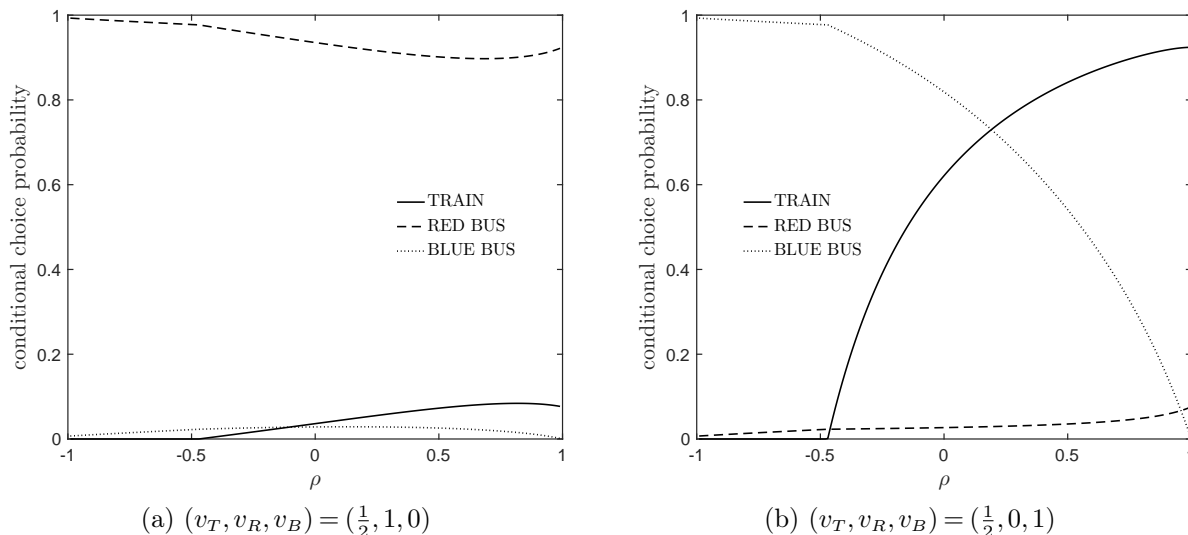


Figure 5 Conditional choice probabilities for RED BUS/BLUE BUS when the two buses are *different*

conditional probability of taking “a bus” (red or blue). We know that when the decision maker faces uniform information costs ($\lambda_R = \lambda_B = 0.4$), she takes the bus too often (see also Figure 2 and the related discussion). One could conjecture that reducing the information cost of even one alternative increases the amount of information processed, so it should bring this probability close to the perfect information case ($\lambda_R = \lambda_B = 0$). From Figure 6, it is evident that this is only partially correct. For negative correlation levels, the DM more correctly identifies the TRAIN as the fastest alternative. However, at positive correlation levels a new decision bias is created. Since the DM knows more about the bus with the lower cost, she starts drawing strong (and wrong) inferences about the other bus, and this time she ends up taking the TRAIN too often. Reducing the information cost of the BLUE BUS so that information costs are uniform again ($\lambda_R = \lambda_B = 0.2$) eliminates this decision bias and brings the conditional choice probability of choosing either bus closer to the perfect information case for all levels of ρ . This highlights the benefits that a seller can potentially earn from presenting information about different choice alternatives in a rather similar and uniform manner.

7. Concluding Remarks

In this paper, we develop a consumer choice model where rationally inattentive customers choose among a given set of alternatives. Our novel contribution is the incorporation of information costs that differ among the alternatives. This captures the notion that it might be inherently (or by seller design) more difficult to learn about some alternatives than about others. We axiomatically derive an information cost function that distinguishes between direct and implied information obtained by the consumer from studying each alternative, and that prioritizes the use of cheaper sources in the acquisition and processing of information. This conditional mutual information based function generalizes the Shannon cost functions commonly utilized in the rational inattention literature.

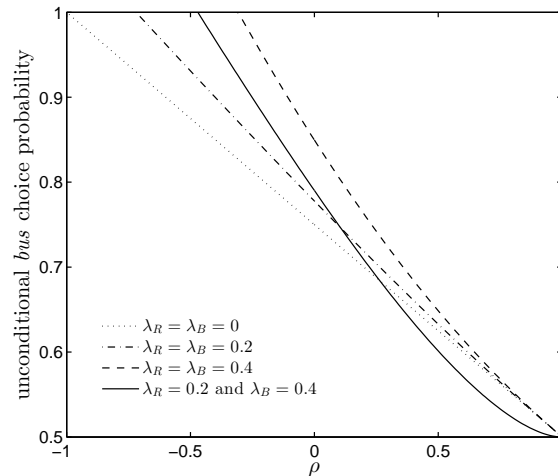


Figure 6 Unconditional choice probabilities for choosing either the RED BUS or the BLUE BUS

We analyze the choice problem of the consumer and show that the optimal choice behavior can be characterized analytically. When the unit cost of acquiring information is the same across all alternatives, the choice behavior reduces to the GMNL choice studied by Matějka and McKay (2015). According to the optimal choice behavior, the conditional choice probability associated with each alternative depends on realized values of the alternatives, their information costs, and prior beliefs. Although the exact relationship is non-trivial, essentially the relative “attractiveness” of each alternative is adjusted by the fact that the consumer learns more about the alternatives with lower information costs. Accordingly, if the information obtained by these alternatives imply a higher (or lower) likelihood of selecting a particular alternative, it is weighed into the attractiveness of that alternative appropriately.

We study a number of limiting scenarios and typical examples to illustrate the optimal choice behavior, and show that non-uniform information costs can induce complex consumer behaviour. Accordingly, the consequences for the seller depend on the particular situation. Although an asymmetric reduction of information costs yields an overall better-informed consumer, the consumer’s beliefs can become strongly biased by focusing on a particular information channel. Perhaps surprisingly, there are situations where the market share of a product may increase when it becomes harder to learn about it. Our characterization enables us to verify if such changes (perhaps due to alterations in the information provision strategy of the seller) would lead to more correct (or incorrect) choices for the consumer, and can be used to evaluate the benefits (or losses) to the seller. We also identify that if two products are very similar in nature, a difference in information costs typically leads to a striking change in relative market share. This is because the consumer then mainly relies on information about the product with low information cost and forms her belief about the product with high information cost based on implied information. As both products are

similar in quality, she prefers the product of which she is more confident about. In addition to above, our model provides an explanation for strong failure of regularity in consumer behavior, a well-documented phenomenon that occurs if the addition of an inferior – dominated, never selected – product increases the market share of another product. This may occur if the newly introduced inferior product facilitates an easier access to information about existing products. To the best of our knowledge, this explanation for the strong failure of regularity has not been discussed in the extant literature.

Most of the decisions that consumers have to make require time, attention and cognitive effort, all of which are limited resources. Our model offers a micro-founded description of how such choices are made when the consumers trade off the value of better information against the costs, in a context where information can be acquired about the alternatives with different rates of time-and-attention-efficiency. As noted earlier, this choice behavior and the resulting description of demand is a crucial input to many practical operational problems. As a concluding example, consider an online firm like airbnb.com or booking.com. When consumers search for a particular accommodation, there are usually a large number of potential hits. It is well-known (e.g., De los Santos et al. 2012) that people do not have the time and attention span to go through all pages. What is often displayed on the first page (or even a subset of this page) is where most attention is directed, while choices listed on the following pages require additional effort to evaluate. From the seller’s revenue management perspective, it is extremely important to decide on the order at which alternatives are displayed. Determining this requires a consumer demand model that describes how choices are going to be made when the cost of information differs among the alternatives and the consumer is rational and efficient when evaluating her alternatives. Going a step of further, such sellers face the trade-off between displaying more alternatives on the same page with less related information (high information costs) versus less alternatives but with more available information (low information costs). Our choice model has the potential to serve as the building block of such product assortment, ordering and strategic information provisioning decisions.

It is worth pointing out that real-life practical applications of our consumer choice model would benefit from two key developments. The first one pertains to the empirical validation of rational inattention and estimation of the choice model. Fortunately, there is a fast growing recent interest in the economics literature on rational inattention and breakthroughs are being made in both theoretical and empirical directions. In recent work, Caplin and Dean (2015) describe a method to identify whether a decision maker is rational inattentive or not from state dependent choice data. In a similar vein, Oliveira et al. (2016) introduce a method to elicit preferences and to estimate the information cost function of a rationally inattentive decision maker. Their setup works for what they call canonical information cost functions, to which our generalized Shannon information cost

function belongs. The second development required pertains to solution methods. It is well-known that assortment optimization problems, even under simple variations of the multinomial logit choice, are very difficult to solve (they are often NP-hard or NP-complete). The generalized GMNL choice model we develop adds additional computational challenges. Therefore, in order to solve realistically-sized practical problems, it is necessary to develop an efficient algorithm to first solve the consumer choice model and then embed it in pricing and/or assortment optimization. This is subject of our ongoing research.

References

- Alptekinoglu, A., J. H. Semple. 2015. The exponential choice model: A new alternative for assortment and price optimization. *Operations Research* **64**(1) 79–93.
- Anderson, S. P., A. de Palma, J-F. Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge.
- Blanchet, J. H., G. Gallego, V. Goyal. 2016. A markov chain approximation to choice modeling. *Operations Research* **64**(4) 886–905.
- Boyacı, T., Y. Akçay. 2016. Pricing when customers have limited attention. Working paper, ESMT Berlin.
- Branco, F., M. Sun, J. M. Villas-Boas. 2012. Optimal search for product information. *Management Science* **58**(11) 2037–2056.
- Caplin, A., M. Dean. 2015. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review* **105**(7) 2183–2203.
- Cover, T. M., J. A. Thomas. 2006. *Elements of Information Theory*. 2nd ed. Wiley, New York.
- Csiszár, I. 2008. Axiomatic characterizations of information measures. *Entropy* **10**(3) 261–273.
- Davis, A. M., E. Katok, A. M. Kwasnica. 2014. Should sellers prefer auctions? a laboratory comparison of auctions and sequential mechanisms. *Management Science* **60**(4) 990–1008.
- De los Santos, B., A. Hortaçsu, M.R. Wildenbeest. 2012. Testing models of consumer search using data on web browsing and purchasing behavior. *The American Economic Review* **102**(6) 2955–2980.
- Dong, L., P. Kouvelis, Z. Tian. 2009. Dynamic pricing and inventory control of substitute products. *Manufacturing & Service Operations Management* **11**(2) 317–339.
- Gabaix, X. 2014. A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics* **129**(4) 1661–1710.
- Hamilton, R. W., D. V. Thompson. 2007. Is there a substitute for direct experience? Comparing consumers' preferences after direct and indirect product experiences. *Journal of Consumer Research* **34**(4) 546–555.
- Hanson, W., K. Martin. 1996. Optimizing multinomial logit profit functions. *Management Science* **42**(7) 992–1003.

- Huber, J., J. W. Payne, C. Puto. 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research* **9**(1) 90–98.
- Ke, T. T., Z-J M. Shen, J. M. Villas-Boas. 2016. Search for information on multiple products. *Management Science*. Forthcoming.
- Lovász, L. 1983. Submodular functions and convexity. Achim Bachem, Bernhard Korte, Martin Grötschel, eds., *Mathematical Programming The State of the Art*. Springer, 235–257.
- Luce, R. D., H. Raiffa. 1957. *Games and Decisions*. Wiley, New York.
- Luce, R. D., P. Suppes. 1965. *Preference, utility, and subjective probability*, *Handbook of Mathematical Psychology*, vol. 3. Wiley, New York.
- Luce, R.D. 1959. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York.
- Maćkowiak, B., M. Wiederholt. 2009. Optimal sticky prices under rational inattention. *American Economic Review* **99**(3) 769–803.
- Maćkowiak, B., M. Wiederholt. 2015. Business cycle dynamics under rational inattention. *The Review of Economic Studies* **82**(4) 1502 – 1532.
- Manzini, P., M. Mariotti. 2014. Stochastic choice and consideration sets. *Econometrica* **82**(3) 1153–1176.
- Matějka, F. 2015. Rigid pricing and rationally inattentive consumer. *Journal of Economic Theory* **158** 656–678.
- Matějka, F., A. McKay. 2012. Simple market equilibria with rationally inattentive consumers. *American Economic Review* **102**(3) 24–29.
- Matějka, F., A. McKay. 2015. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review* **105**(1) 272–298.
- McKelvey, R. D., T. R. Palfrey. 1995. Quantal response equilibria for normal form games. *Games and economic behavior* **10**(1) 6–38.
- Oliveira, H., T. Denti, M. Mihm, K. Ozbek. 2016. Rationally inattentive preferences and hidden information costs. *Theoretical Economics* forthcoming.
- Reis, R. 2006. Inattentive producers. *The Review of Economic Studies* **73**(3) 793–821.
- S. J. Hoch, Y-W. Ha. 1986. Consumer learning: Advertising and the ambiguity of product experience. *Journal of Consumer Research* **13**(2) 221–233.
- Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* **27**(3) 379–423.
- Simon, H. A. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* **69**(1) 99–118.
- Simon, H. A. 1979. Information processing models of cognition. *Annual review of psychology* **30**(1) 363–396.

- Simonson, I., A. Tversky. 1992. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research* **29**(3) 281–295.
- Sims, C. A. 1998. Stickiness. *Carnegie-Rochester Conference Series on Public Policy* **49**(1) 317–356.
- Sims, C. A. 2003. Implications of rational inattention. *Journal of Monetary Economics* **50**(3) 665–690.
- Sims, C. A. 2006. Rational inattention: Beyond the linear-quadratic case. *The American Economic Review* **96**(2) 158–163.
- Srikanth, J., P. Rusmevichientong. 2016. A nonparametric joint assortment and price choice model. *Management Science*. Forthcoming.
- Talluri, K., G. van Ryzin. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science* **50**(1) 15–33.
- Todd, P. M., G. Gigerenzer. 2000. Précis of simple heuristics that make us smart. *Behavioral and brain sciences* **23**(05) 727–741.
- Tutino, A. 2013. Rationally inattentive consumption choices. *Review of Economic Dynamics* **16**(3) 421–439.
- van Ryzin, G., S. Mahajan. 1999. On the relationship between inventory costs and variety benefits in retail assortments. *Management Science* **45**(11) 1496–1509.
- Verrecchia, R. E. 1982. Information acquisition in a noisy rational expectations economy. *Econometrica* 1415–1430.
- Weitzman, M. L. 1979. Optimal search for the best alternative. *Econometrica* **47**(3) 641–654.
- Wierenga, B. 2008. *Handbook of marketing decision models*. Springer, New York.
- Zhang, D., D. Adelman. 2009. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science* **43**(3) 381–394.

Appendix A: Proofs

Proof of Theorem 1

We first verify that the cost function C_λ^I given in (5) satisfies singular information pricing (SIP) and joint information pricing (JIP). Let $f_1, f_2 \in \Delta(\mathbb{R}^{2n})$, $c \in \mathbb{R}$, and $D \subseteq \mathcal{A}$ be such that

$$I(f_2, E) = \begin{cases} I(f_1, E) + c & \text{whenever } D \subseteq E \\ I(f_1, E) & \text{whenever } D \not\subseteq E. \end{cases}$$

Let $i_D = \max_{i \in D} \{i\}$ denote the alternative with highest index in D . Note that $i_D \in \arg \max_{i \in D} \{\lambda_i\}$. Then,

$$\begin{aligned} C_\lambda^I(f_1) - C_\lambda^I(f_2) &= \sum_{k=1}^n \lambda_k \cdot [I(f_1, \{1, \dots, k\}) - I(f_2, \{1, \dots, k\})] \\ &\quad - \sum_{k=1}^n \lambda_k \cdot [I(f_1, \{1, \dots, k-1\}) - I(f_2, \{1, \dots, k-1\})] \\ &= \sum_{k=i_D}^n \lambda_k \cdot c - \sum_{k=i_D+1}^n \lambda_k \cdot c = c \cdot \lambda_{i_D}. \end{aligned}$$

Concerning uniqueness, let ϕ_λ^I denote a cost function that satisfies SIP and JIP. For $f \in \Delta(\mathbb{R}^{2n})$ and $D \subseteq \mathcal{A}$ define $\mu(f, D) \in \mathbb{R}$ by

$$\mu(f, D) = I(f, D) - \sum_{E \subsetneq D} \mu(f, E),$$

where $\mu(f, \emptyset) = 0$. Then, $I(f, D) = \sum_{E \subseteq D} \mu(f, E)$ for all $D \subseteq \mathcal{A}$.

Given an information strategy \hat{f} , let $\hat{f} \dot{+} \mu(f, E)$ be a strategy with

$$I(\hat{f} \dot{+} \mu(f, E), D) = \begin{cases} I(\hat{f}, D) + \mu(f, E), & \text{if } E \subseteq D \\ I(\hat{f}, D), & \text{else.} \end{cases}$$

By SIP and JIP, $\phi_\lambda^I(\hat{f} \dot{+} \mu(f, E)) = \phi_\lambda^I(\hat{f}) + \mu(f, E) \cdot \max_{i \in E} \lambda_i$. Since we can obtain f from f^0 by stepwise $\dot{+}$ -addition of $\mu(f, D)$, $D \subseteq \mathcal{A}$, we obtain a uniquely defined cost $\phi_\lambda^I(f)$ starting from $\phi_\lambda^I(f^0) = 0$. Hence, $\phi_\lambda^I(f) = C_\lambda^I(f)$ for all $f \in \Delta(\mathbb{R}^{2n})$.

Proof of Lemma 1

The proof is by contradiction. Suppose that $i \in \mathcal{A}$ is such that $p_{f^*}(i, \mathbf{v}) > 0$, and that there exist $S_i^1, S_i^2 \subseteq S_i$, satisfying $\int_{\mathbf{v} \in \mathbb{R}^n} \int_{\mathbf{s} \in S_i^1} f^*(d\mathbf{s}, d\mathbf{v}) > 0$, $\int_{\mathbf{v} \in \mathbb{R}^n} \int_{\mathbf{s} \in S_i^2} f^*(d\mathbf{s}, d\mathbf{v}) > 0$, and $f_{\mathbf{v}|\mathbf{s}^1}^* \neq f_{\mathbf{v}|\mathbf{s}^2}^*$ for all $\mathbf{s}^1 \in S_i^1, \mathbf{s}^2 \in S_i^2$. We can construct a better information strategy h as follows. Pick some $\hat{\mathbf{s}} \in S_i^1 \cup S_i^2$. Define h by setting for all \mathbf{v} :

$$h(\mathbf{s}, \mathbf{v}) := f^*(\mathbf{s}, \mathbf{v}) \text{ for all } \mathbf{s} \notin (S_i^1 \cup S_i^2), \quad (20)$$

$$h_{\mathbf{S}|\mathbf{v}}(\hat{\mathbf{s}}) := \int_{\mathbf{s} \in S_i^1} f_{\mathbf{S}|\mathbf{v}}^*(d\mathbf{s}) + \int_{\mathbf{s} \in S_i^2} f_{\mathbf{S}|\mathbf{v}}^*(d\mathbf{s}), \text{ and} \quad (21)$$

$$h_{\mathbf{S}|\mathbf{v}}(\mathbf{s}) := 0 \text{ for all } \mathbf{s} \in (S_i^1 \cup S_i^2) \setminus \{\hat{\mathbf{s}}\}. \quad (22)$$

Note that h is consistent with p . Matějka and McKay (2015) establish that h yields a revenue at least as high as the revenue of f^* , i.e., $\int_{\mathbf{v} \in \mathbb{R}^n} \int_{\mathbf{s} \in \mathbb{R}^n} U(h_{\mathbf{v}|\mathbf{s}}) h_{\mathbf{S}|\mathbf{v}}(d\mathbf{s}) p(d\mathbf{v}) \geq \int_{\mathbf{v} \in \mathbb{R}^n} \int_{\mathbf{s} \in \mathbb{R}^n} U(f_{\mathbf{v}|\mathbf{s}}) f_{\mathbf{S}|\mathbf{v}}^*(d\mathbf{s}) g(d\mathbf{v})$.

We next compare the information costs of h and f^* . Since the difference between the mutual information of h and f^* stems from where the distributions differ, it is helpful to make use of the the probability distributions restricted to this domain. More precisely, we construct a probability distribution $g|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}$ on the restricted domain $S_i^1 \cup S_i^2 \times \mathbf{V}_{1..l}$ from g by rescaling to $\int_{\mathbf{s} \in S_i^1 \cup S_i^2, \mathbf{v} \in \mathbf{V}_{1..l}} g(\mathbf{s}, \mathbf{v}) d\mathbf{s} d\mathbf{v}$. Formally, let $f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}$ denote the probability distribution obtain from f^* as

$$f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}, \mathbf{v}) = \frac{f^*(\mathbf{s}, \mathbf{v})}{\int_{\mathbf{s} \in S_i^1 \cup S_i^2, \mathbf{v} \in \mathbf{V}_{1..l}} f^*(d\mathbf{s}, d\mathbf{v})} \quad \text{for all } \mathbf{s} \in S_i^1 \cup S_i^2, \mathbf{v} \in \mathbf{V}_{1..l}.$$

Analogously, we define $h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}$.

Dividing the following equation by $\int_{\mathbf{s} \in S_i^1 \cup S_i^2, \mathbf{v} \in \mathbf{V}_{1..l}} f^*(d\mathbf{s}, d\mathbf{v})$,

$$\begin{aligned} & \mathcal{I}_{f^*}(V_\ell, \mathbf{S} | \mathbf{V}_{1..l-1}) - \mathcal{I}_h(V_\ell, \mathbf{S} | \mathbf{V}_{1..l-1}) \\ &= \int_{\mathbf{v}_{1..l}, \mathbf{s}} f^*(\mathbf{s}, \mathbf{v}_{1..l}) \left(\log \frac{f_{\mathbf{S}|\mathbf{v}_{1..l}}^*(\mathbf{s})}{f_{\mathbf{S}|\mathbf{v}_{1..l-1}}^*(\mathbf{s})} \right) d\mathbf{s} d\mathbf{v}_{1..l} - \int_{\mathbf{v}_{1..l}, \mathbf{s}} h_{\mathbf{S}|\mathbf{v}_{1..l}}(\mathbf{s}) \left(\log \frac{h_{\mathbf{S}|\mathbf{v}_{1..l}}(\mathbf{s})}{h_{\mathbf{S}|\mathbf{v}_{1..l-1}}(\mathbf{s})} \right) d\mathbf{s} d\mathbf{v}_{1..l} \\ & \stackrel{(20)-(22)}{=} \int_{\mathbf{v}_{1..l}, \mathbf{s} \in S_i^1 \cup S_i^2} f^*(\mathbf{s}, \mathbf{v}_{1..l}) \left(\log \frac{f_{\mathbf{S}|\mathbf{v}_{1..l}}^*(\mathbf{s})}{f_{\mathbf{S}|\mathbf{v}_{1..l-1}}^*(\mathbf{s})} \right) d\mathbf{s} d\mathbf{v}_{1..l} - \int_{\mathbf{v}_{1..l}} h(\hat{\mathbf{s}}, \mathbf{v}_{1..l}) \left(\log \frac{h_{\mathbf{S}|\mathbf{v}_{1..l}}(\hat{\mathbf{s}})}{h_{\mathbf{S}|\mathbf{v}_{1..l-1}}(\hat{\mathbf{s}})} \right) d\mathbf{v}_{1..l}, \end{aligned}$$

we get

$$\begin{aligned}
X &:= \left(\int_{\mathbf{s} \in S_i^1 \cup S_i^2} f^*(d\mathbf{s}, d\mathbf{v}) \right)^{-1} \cdot [\mathcal{I}_{f^*}(V_\ell, \mathbf{S} | \mathbf{V}_{1..l-1}) - \mathcal{I}_h(V_\ell, \mathbf{S} | \mathbf{V}_{1..l-1})] \\
&= \int_{\mathbf{v}_{1..l} \in \mathbb{R}^\ell, \mathbf{s} \in S_i^1 \cup S_i^2} f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}, \mathbf{v}_{1..l}) \left(\log \frac{\left(f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}) \right)_{\mathbf{S} | \mathbf{v}_{1..l}}(\mathbf{s})}{\left(f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}) \right)_{\mathbf{S} | \mathbf{v}_{1..l-1}}(\mathbf{s})} \right) d\mathbf{s} d\mathbf{v}_{1..l} \\
&\quad - \int_{\mathbf{v}_{1..l} \in \mathbb{R}^\ell} h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\hat{\mathbf{s}}, \mathbf{v}_{1..l}) \left(\log \frac{\left(h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\hat{\mathbf{s}}) \right)_{\mathbf{S} | \mathbf{v}_{1..l}}(\hat{\mathbf{s}})}{\left(h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\hat{\mathbf{s}}) \right)_{\mathbf{S} | \mathbf{v}_{1..l-1}}(\hat{\mathbf{s}})} \right) d\mathbf{v}_{1..l}
\end{aligned}$$

since the ratio of the conditionals within the log terms remains the same when restricting the domain to $S_i^1 \cup S_i^2 \times \mathbf{V}_{1..l}$. Using $\Pr(x, y) = \Pr(x | y) \cdot \Pr(y)$, we get

$$\begin{aligned}
X &= \int_{\mathbf{v}_{1..l} \in \mathbb{R}^\ell, \mathbf{s} \in S_i^1 \cup S_i^2} f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}, \mathbf{v}_{1..l}) \left(\log \frac{f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}, \mathbf{v}_{1..l})}{f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}, \mathbf{v}_{1..l-1})} - \log \frac{f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{v}_{1..l})}{f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{v}_{1..l-1})} \right) d\mathbf{s} d\mathbf{v}_{1..l} \\
&\quad - \int_{\mathbf{v}_{1..l} \in \mathbb{R}^\ell} h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\hat{\mathbf{s}}, \mathbf{v}_{1..l}) \left(\log \frac{h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\hat{\mathbf{s}}, \mathbf{v}_{1..l})}{h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\hat{\mathbf{s}}, \mathbf{v}_{1..l-1})} - \log \frac{h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{v}_{1..l})}{h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{v}_{1..l-1})} \right) d\mathbf{v}_{1..l}.
\end{aligned}$$

Using (20)-(22), this can be further simplified to

$$\begin{aligned}
X &= \int_{\mathbf{v}_{1..l} \in \mathbb{R}^\ell, \mathbf{s} \in S_i^1 \cup S_i^2} f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}, \mathbf{v}_{1..l}) \left(\log \frac{f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}, \mathbf{v}_{1..l})}{f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\mathbf{s}, \mathbf{v}_{1..l-1})} \right) d\mathbf{s} d\mathbf{v}_{1..l} \\
&\quad - \int_{\mathbf{v}_{1..l} \in \mathbb{R}^\ell} h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\hat{\mathbf{s}}, \mathbf{v}_{1..l}) \left(\log \frac{h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\hat{\mathbf{s}}, \mathbf{v})}{h|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}(\hat{\mathbf{s}}, \mathbf{v}_{1..l-1})} \right) d\mathbf{v}_{1..l} \\
&= \mathcal{H} \left(\left(f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}} \right)_{\mathbf{S}, \mathbf{v}_{1..l-1}} \right) - \mathcal{H} \left(f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}} \right) - \left[\mathcal{H} \left(\left(f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}} \right)_{\mathbf{v}_{1..l-1}} \right) - \mathcal{H} \left(f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}} \right)_{\mathbf{v}_{1..l}} \right] \\
&= \mathcal{I}_{f^*|_{S_i^1 \cup S_i^2, \mathbf{v}_{1..l}}}(\mathbf{S}, \mathbf{V}_{1..l} | \mathbf{V}_{1..l-1}) \geq 0.
\end{aligned}$$

Since we have multiple posteriors on $S_i^1 \cup S_i^2$, it cannot be true that for all $\ell \in A$, \mathbf{S} and $\mathbf{V}_{1..l}$ are independent conditional on $\mathbf{V}_{1..l-1}$. Hence, there exists ℓ for which the last term is strictly positive. But then $\mathcal{I}_{f^*}(V_\ell, \mathbf{S} | \mathbf{V}_{1..l-1}) - \mathcal{I}_h(V_\ell, \mathbf{S} | \mathbf{V}_{1..l-1}) > 0$, implying that h is strictly cheaper than f^* ; a contradiction.

Proof of Proposition 1

According to Lemma 1, an optimal information strategy f necessitates a single posterior for all signals $\mathbf{s} \in S_i$. Thus, the revenue function becomes

$$\begin{aligned}
&\int_{\mathbf{v} \in \mathbb{R}^n} \int_{\mathbf{s} \in \mathbb{R}^n} U(f_{\mathbf{v}|\mathbf{s}}) f_{\mathbf{S}|\mathbf{v}}(d\mathbf{s}) g(d\mathbf{v}) \\
&= \int_{\mathbf{v} \in \mathbb{R}^n} \sum_{i=1}^n \int_{\mathbf{s} \in S_i} U(f_{\mathbf{v}|\mathbf{s}}) f_{\mathbf{S}|\mathbf{v}}(d\mathbf{s}) g(d\mathbf{v}) = \sum_{i=1}^n U(f_{\mathbf{v}|\mathbf{s} \in S_i}) \int_{\mathbf{v} \in \mathbb{R}^n} \int_{\mathbf{s} \in S_i} f_{\mathbf{S}|\mathbf{v}}(d\mathbf{s}) g(d\mathbf{v}) \\
&= \sum_{i=1}^n U(f_{\mathbf{v}|\mathbf{s} \in S_i}) \cdot p_f(i) = \sum_{i=1}^n \int_{\mathbf{v} \in \mathbb{R}^n} v_i \cdot f_{\mathbf{v}|\mathbf{s} \in S_i}(\mathbf{v}) \cdot p_f(i) d\mathbf{v} \\
&= \sum_{i=1}^n \int_{\mathbf{v} \in \mathbb{R}^n} v_i \cdot f(\mathbf{s} \in S_i, \mathbf{v}) d\mathbf{v} = \sum_{i=1}^n \int_{\mathbf{v} \in \mathbb{R}^n} v_i \cdot p_f(i | \mathbf{v}) p(d\mathbf{v})
\end{aligned}$$

Moreover, define for an optimal information strategy f the random variable \mathbf{A}_f that takes the value i with the probability that i is chosen under f , $\Pr(\mathbf{A}_f = i) = p_f(i)$. The distribution of \mathbf{V} conditional

on \mathbf{S} is the same as the distribution of \mathbf{V} conditional on \mathbf{A}_f , hence knowing \mathbf{A}_f is as informative about \mathbf{V} as knowing \mathbf{S} , i.e., $\mathcal{I}_{f^*}(V_\ell, \mathbf{S} | \mathbf{V}_{1..\ell-1}) = \mathcal{I}_{f^*}(V_\ell, \mathbf{A}_{f^*} | \mathbf{V}_{1..\ell-1})$. Formally, this can be shown as follows: $f_{V_\ell|i, \mathbf{v}_{1..\ell-1}} = f_{V_\ell|\mathbf{s}, \mathbf{v}_{1..\ell-1}}$ for $\mathbf{s} \in S_i$ implies $\mathcal{H}(f_{V_\ell|\mathbf{s}, \mathbf{v}_{1..\ell-1}}) = \mathcal{H}(f_{V_\ell|i, \mathbf{v}_{1..\ell-1}})$ for $\mathbf{s} \in S_i$. Thus,

$$\begin{aligned} & \mathcal{I}_{f^*}(V_\ell, \mathbf{S} | \mathbf{V}_{1..\ell-1}) \\ &= \int_{\mathbf{v}_{1..\ell-1}} f(\mathbf{v}_{1..\ell-1}) \mathcal{H}(f_{V_\ell|\mathbf{v}_{1..\ell-1}}) d\mathbf{v}_{1..\ell-1} - \int_{\mathbf{v}_{1..\ell-1}} \sum_{i=1}^n \int_{\mathbf{s} \in S_i} f(\mathbf{s}) f_{\mathbf{v}_{1..\ell-1}|\mathbf{s}}(\mathbf{v}_{1..\ell-1}) \mathcal{H}(f_{V_\ell|\mathbf{s}, \mathbf{v}_{1..\ell-1}}) ds d\mathbf{v}_{1..\ell-1} \\ &= \int_{\mathbf{v}_{1..\ell-1}} f(\mathbf{v}_{1..\ell-1}) \mathcal{H}(f_{V_\ell|\mathbf{v}_{1..\ell-1}}) d\mathbf{v}_{1..\ell-1} - \int_{\mathbf{v}_{1..\ell-1}} \sum_{i=1}^n p_f(i) f_{\mathbf{v}_{1..\ell-1}|i}(\mathbf{v}_{1..\ell-1}) \mathcal{H}(f_{V_\ell|i, \mathbf{v}_{1..\ell-1}}) ds d\mathbf{v}_{1..\ell-1} \quad (23) \\ &= \mathcal{I}_{f^*}(V_\ell, \mathbf{A}_{f^*} | \mathbf{V}_{1..\ell-1}). \end{aligned}$$

The cost of information can now be written as $\hat{c}(f) = \sum_{\ell=1}^n \lambda_\ell \cdot \mathcal{I}_f(V_\ell, \mathbf{A}_f | \mathbf{V}_{1..\ell-1})$. Using $\mathcal{H}(p_f) = -\sum_{i \in \mathcal{A}} p_f(i) \log p_f(i)$ and $\mathcal{H}(p_f(i | \mathbf{v}_{1..\ell})) = -\sum_{i \in \mathcal{A}} p_f(i | \mathbf{v}_{1..\ell}) \log p_f(i | \mathbf{v}_{1..\ell})$, we can write the objective in terms of $\{p_f(i | \mathbf{v}_{1..\ell})\}_{i \in \mathcal{A}}$. Thus, every optimal information strategy f induces choice probabilities p_f that also maximize

$$\max_{p_f} \sum_{i \in \mathcal{A}} \int_{\mathbf{v} \in \mathbb{R}^n} v_i p_f(i | \mathbf{v}) p(d\mathbf{v}) - \sum_{j \in \mathcal{A}} \lambda_j \cdot \mathcal{I}_p(V_j, \mathbf{A}_f | \mathbf{V}_{1..j-1}).$$

On the other hand, let $\{p^*(i | \mathbf{v})\}_{i \in \mathcal{A}}$ be a collection that solves the problem in the proposition. Select n distinct signals $\{\hat{\mathbf{s}}_i\}_{i \in \mathcal{A}}$. Define f^* by

$$f^*(\mathbf{s}, \mathbf{v}) = \begin{cases} p^*(i | \mathbf{v}) p(\mathbf{v}), & \text{if } \mathbf{s} = \hat{\mathbf{s}} \\ 0, & \text{otherwise.} \end{cases}$$

Then, f^* is an optimal strategy according to (11) and (12). To see this, note that f^* is consistent with the priors and induces choice probabilities that maximizes the objective in the proposition. Since every other optimal strategy induces choice probabilities that maximize objective in the proposition, there cannot be a strategy that yields a higher payoff than f^* .

Proof of Theorem 2

We already proved the general case allowing for $\lambda_{h+1} = \dots = \lambda_n = \infty$. The parameters $\lambda_{h+1}, \dots, \lambda_n$ appear in the respective mutual information terms (16) in the optimization problem of Proposition 1. Setting these terms to zero implies $p(i | \mathbf{v}_{1..j-1}) = p(i | \mathbf{v}_{1..j})$ for all i and all $j \in \{h+1, \dots, n\}$. The objective then becomes

$$\max_{\{p(i|\mathbf{v}_{1..h})\}_{i \in \mathcal{A}}} \sum_{i=1}^h \int_{\mathbf{v}_{1..h} \in \mathbb{R}^h} v_i p(i | \mathbf{v}_{1..h}) p(d\mathbf{v}_{1..h}) + \sum_{i=h+1}^n \int_{\mathbf{v} \in \mathbb{R}^n} v_i p(i | \mathbf{v}_{1..h}) p(d\mathbf{v}) - \sum_{j=1}^h \lambda_j \cdot \mathcal{I}_p(V_j, \mathbf{A} | \mathbf{V}_{1..j-1})$$

subject to

$$\begin{aligned} p(i | \mathbf{v}_{1..h}) &\geq 0 \quad \text{for all } i \in \mathcal{A} \text{ and } \mathbf{v}_{1..h} \in \mathbb{R}^h \\ \sum_{i \in \mathcal{A}} p(i | \mathbf{v}_{1..h}) &= 1. \end{aligned}$$

Using

$$\begin{aligned} \int_{\mathbf{v} \in \mathbb{R}^n} v_i p(i | \mathbf{v}_{1..h}) p(\mathbf{v}) d\mathbf{v} &= \int_{\mathbf{v}_{1..h} \in \mathbb{R}^h} \left(\int_{\mathbf{v}_{h+1..n} \in \mathbb{R}^{n-h}} v_i p(\mathbf{v}_{1..h+1} | \mathbf{v}_{1..h}) d\mathbf{v}_{h+1..n} \right) p(i | \mathbf{v}_{1..h}) p(\mathbf{v}_{1..h}) d\mathbf{v}_{1..h} \\ &= \int_{\mathbf{v}_{1..h} \in \mathbb{R}^h} \mathbb{E}_{p_{V_i|\mathbf{v}_{1..h}}} [v_i] p(i | \mathbf{v}_{1..h}) p(d\mathbf{v}_{1..h}) \end{aligned}$$

for $i > h$, and insertion of the equation

$$\mathcal{I}_p(V_j, \mathbf{A} | \mathbf{V}_{1..j-1}) = - \sum_{i \in \mathcal{A}} \int_{\mathbf{v}_{1..j} \in \mathbb{R}^j} p(i | \mathbf{v}_{1..j}) (\log p(i | \mathbf{v}_{1..j-1}) - \log p(i | \mathbf{v}_{1..j})) p(d\mathbf{v}_{1..j}),$$

yields the Lagrangian

$$\begin{aligned} \mathcal{L}(\{p(i | \mathbf{v}_{1..h})\}_{i \in \mathcal{A}}) &= \sum_{i=1}^h \int_{\mathbf{v}_{1..h} \in \mathbb{R}^h} v_i p(i | \mathbf{v}_{1..h}) p(d\mathbf{v}_{1..h}) + \sum_{i=h+1}^n \int_{\mathbf{v}_{1..h} \in \mathbb{R}^h} \mathbb{E}_{p_{V_i | \mathbf{v}_{1..h}}} [v_i] p(i | \mathbf{v}_{1..h}) p(d\mathbf{v}_{1..h}) \\ &+ \sum_{j=1}^h \lambda_j \cdot \sum_{i \in \mathcal{A}} \int_{\mathbf{v}_{1..j} \in \mathbb{R}^j} p(i | \mathbf{v}_{1..j}) (\log p(i | \mathbf{v}_{1..j-1}) - \log p(i | \mathbf{v}_{1..j})) p(d\mathbf{v}_{1..j}) \\ &+ \int_{\mathbf{v}_{1..h} \in \mathbb{R}^h} \xi_i(\mathbf{v}_{1..h}) p(i | \mathbf{v}_{1..h}) p(d\mathbf{v}_{1..h}) \\ &- \int_{\mathbf{v}_{1..h} \in \mathbb{R}^h} \mu(\mathbf{v}_{1..h}) \left(\sum_{i=1}^n p(i | \mathbf{v}_{1..h}) p(d\mathbf{v}_{1..h}) - 1 \right) p(d\mathbf{v}_{1..h}). \end{aligned}$$

If $p(i) > 0$, then $p(i | \mathbf{v}_{1..j}) > 0$ for $j < h + 1$. To see this, suppose there exists a smallest m such that $p(i | \mathbf{v}_{1..m}) = 0$ for some $\mathbf{v}_{1..m}$ and $p(i | \mathbf{v}_{1..k}) > 0$ for all $k < m$. Then, $p(i, \mathbf{v}_{1..m}) = 0$, implying $p(i, \mathbf{v}_{1..k}) = 0$ and hence $p(i | \mathbf{v}_{1..k}) = 0$ for all $k > m$ and all $\mathbf{v}_{1..k}$ with positive measure. Yet, an incremental increase in $p(i | \mathbf{v}_{1..m})$ and setting $p(i | \mathbf{v}_{1..m}) = p(i | \mathbf{v}_{1..k})$ for all $k > m$ would reduce the cost of information by

$$\lim_{p(i | \mathbf{v}_{1..m}) \searrow 0} \frac{d}{dp(i | \mathbf{v}_{1..m})} p(i | \mathbf{v}_{1..m}) (\log p(i | \mathbf{v}_{1..m-1}) - \log p(i | \mathbf{v}_{1..m})) = -\infty.$$

Thus, $p(i | \mathbf{v}_{1..m}) = 0$ cannot be optimal. Consequently, $\xi_i(\mathbf{v}_{1..h}) = 0$.

The FOCs w.r.t. $p(i | \mathbf{v}_{1..h})$ are

$$\begin{aligned} 0 &= \int_{\mathbf{v}_{1..h} \in \mathbb{R}^h} v_i - \mu(\mathbf{v}_{1..h}) p(d\mathbf{v}_{1..h}) + \lambda_1 \cdot (\log p(i) + 1) \\ &+ \sum_{j=1}^{h-1} (\lambda_{j+1} - \lambda_j) \cdot \int_{\mathbf{v}_{1..j}} (\log p(i | \mathbf{v}_{1..j}) + 1) p(d\mathbf{v}_{1..j}) - \lambda_h \cdot \int_{\mathbf{v}_{1..h}} (\log p(i | \mathbf{v}_{1..h}) + 1) p(d\mathbf{v}_{1..h}) \end{aligned}$$

for $i \leq h$, while for $i > h$ we obtain

$$\begin{aligned} 0 &= \int_{\mathbf{v}_{1..h} \in \mathbb{R}^h} \mathbb{E}_{p_{V_i | \mathbf{v}_{1..h}}} [v_i] - \mu(\mathbf{v}_{1..h}) p(d\mathbf{v}_{1..h}) + \lambda_1 \cdot (\log p(i) + 1) \\ &+ \sum_{j=1}^{h-1} (\lambda_{j+1} - \lambda_j) \cdot \int_{\mathbf{v}_{1..j}} (\log p(i | \mathbf{v}_{1..j}) + 1) p(d\mathbf{v}_{1..j}) - \lambda_h \cdot \int_{\mathbf{v}_{1..h}} (\log p(i | \mathbf{v}_{1..h}) + 1) p(d\mathbf{v}_{1..h}) \end{aligned}$$

We get

$$0 = v_i - \mu(\mathbf{v}_{1..h}) + \lambda_1 \log p(i) + \sum_{j=1}^{h-1} (\lambda_{j+1} - \lambda_j) \cdot \log p(i | \mathbf{v}_{1..j}) - \lambda_h \log p(i | \mathbf{v}_{1..h}) \quad \text{for } i \leq h \text{ and}$$

$$0 = \mathbb{E}_{p_{V_i | \mathbf{v}_{1..h}}} [v_i] - \mu(\mathbf{v}_{1..h}) + \lambda_1 \log p(i) + \sum_{j=1}^{h-1} (\lambda_{j+1} - \lambda_j) \cdot \log p(i | \mathbf{v}_{1..j}) - \lambda_h \log p(i | \mathbf{v}_{1..h}) \quad \text{for } i > h,$$

implying

$$p(i | \mathbf{v}_{1..h}) = e^{\frac{v_i - \mu(\mathbf{v}_{1..h})}{\lambda_h}} \cdot p(i)^{\frac{\lambda_1}{\lambda_h}} \cdot \prod_{k=1}^{h-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_h}} \quad \text{for } i \leq h \text{ and} \quad (24)$$

$$p(i | \mathbf{v}_{1..h}) = e^{\frac{\mathbb{E}_p[v_i] - \mu(\mathbf{v}_{1..h})}{\lambda_h}} \cdot p(i)^{\frac{\lambda_1}{\lambda_h}} \cdot \prod_{k=1}^{h-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_h}} \quad \text{for } i > h. \quad (25)$$

Summing (24) and (25) up over $j \in A$ yields

$$1 = \sum_{j=1}^h e^{\frac{v_i - \mu(\mathbf{v}_{1..h})}{\lambda_h}} \cdot p(i)^{\frac{\lambda_1}{\lambda_h}} \cdot \prod_{k=1}^{h-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_h}} + \sum_{j=h+1}^n e^{\frac{\mathbb{E}_{p_{V_i | \mathbf{v}_{1..h}}} [v_i] - \mu(\mathbf{v}_{1..h})}{\lambda_h}} \cdot p(i)^{\frac{\lambda_1}{\lambda_h}} \cdot \prod_{k=1}^{h-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_h}}$$

Solving for $\mu(\mathbf{v}_{1..h})$ gives

$$\mu(\mathbf{v}_{1..h}) = \lambda_h \ln \left(\sum_{j=1}^h e^{\frac{v_j}{\lambda_h}} \cdot p(j)^{\frac{\lambda_1}{\lambda_h}} \cdot \prod_{k=1}^{h-1} p(j | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_h}} + \sum_{j=h+1}^n e^{\frac{\mathbb{E}_{p_{V_i | \mathbf{v}_{1..h}}} [v_i]}{\lambda_h}} \cdot p(j)^{\frac{\lambda_1}{\lambda_h}} \cdot \prod_{k=1}^{h-1} p(j | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_h}} \right)$$

and plugging this into (24) and (25), respectively, yields the desired equations.

Proof of Proposition 2

Insertion of equation (17) into the objective of Proposition 1 yields

$$\begin{aligned} & \sum_{i \in \mathcal{A}} \int_{\mathbf{v} \in \mathbb{R}^n} v_i p(i | \mathbf{v}) p(d\mathbf{v}) + \sum_{j=1}^{n-1} \lambda_j \cdot \left\{ \sum_{i \in \mathcal{A}} \int_{\mathbf{v}_{1..j} \in \mathbb{R}^j} p(i | \mathbf{v}_{1..j}) (\log p(i | \mathbf{v}_{1..j-1}) - \log p(i | \mathbf{v}_{1..j})) p(d\mathbf{v}_{1..j}) \right\} \\ & + \lambda_n \cdot \left\{ \sum_{i \in \mathcal{A}} \int_{\mathbf{v} \in \mathbb{R}^n} p(i | \mathbf{v}) \left(\log p(i | \mathbf{v}_{1..n-1}) - \log \frac{e^{\frac{v_i}{\lambda_n}} \cdot p(i)^{\frac{\lambda_1}{\lambda_n}} \cdot \prod_{k=1}^{n-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_n}}}{\sum_{j=1}^n e^{\frac{v_j}{\lambda_n}} \cdot p(j)^{\frac{\lambda_1}{\lambda_n}} \cdot \prod_{k=1}^{n-1} p(j | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_n}}} \right) p(d\mathbf{v}) \right\} \\ = & \sum_{i \in \mathcal{A}} \int_{\mathbf{v} \in \mathbb{R}^n} v_i p(i | \mathbf{v}) p(d\mathbf{v}) + \sum_{j=1}^{n-1} \lambda_j \cdot \left\{ \sum_{i \in \mathcal{A}} \int_{\mathbf{v}_{1..j} \in \mathbb{R}^j} p(i | \mathbf{v}_{1..j}) (\log p(i | \mathbf{v}_{1..j-1}) - \log p(i | \mathbf{v}_{1..j})) p(d\mathbf{v}_{1..j}) \right\} \\ & + \lambda_n \cdot \left\{ \sum_{i \in \mathcal{A}} \int_{\mathbf{v} \in \mathbb{R}^n} p(i | \mathbf{v}) \log p(i | \mathbf{v}_{1..n-1}) p(d\mathbf{v}) \right\} \\ & - \left\{ \sum_{i \in \mathcal{A}} \int_{\mathbf{v} \in \mathbb{R}^n} p(i | \mathbf{v}) \left(v_i + \lambda_1 \log p(i) + \sum_{k=1}^{n-1} \lambda_{k+1} \log p(i | \mathbf{v}_{1..k}) - \sum_{k=1}^{n-1} \lambda_k \log p(i | \mathbf{v}_{1..k}) \right) p(d\mathbf{v}) \right\} \\ & + \lambda_n \cdot \left\{ \sum_{i \in \mathcal{A}} \int_{\mathbf{v} \in \mathbb{R}^n} p(i | \mathbf{v}) \left(\log \left(\sum_{j=1}^n e^{\frac{v_j}{\lambda_n}} \cdot p(j)^{\frac{\lambda_1}{\lambda_n}} \cdot \prod_{k=1}^{n-1} p(j | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_n}} \right) \right) p(d\mathbf{v}) \right\} \\ = & \lambda_n \int_{\mathbf{v}} \log \left(\sum_{j \in \mathcal{A}} e^{\frac{v_j}{\lambda_n}} \cdot p(j)^{\frac{\lambda_1}{\lambda_n}} \cdot \prod_{k=1}^{n-1} p(j | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1} - \lambda_k}{\lambda_n}} \right) p(d\mathbf{v}). \end{aligned}$$

Proof of Proposition 3

For duplicates with $\lambda_{\bar{m}} > \lambda_m$, we have $\mathcal{I}_p(V_{\bar{m}}, \hat{\mathbf{A}} | \mathbf{V}_{1..\bar{m}-1}) = 0$, i.e., the cost of information $\lambda_{\bar{m}}$ is irrelevant. Moreover, learning about \bar{m} beyond m is not possible, hence $\hat{p}(i | \mathbf{u}_{1..\bar{m}}) = \hat{p}(i | \mathbf{u}_{1..\bar{m}-1})$. Thus, the objectives of the problem with and without the duplicate are identical if $\hat{p}(\bar{m} | \mathbf{u}_{1..\ell}) = 0$. Let the optimal conditional choice probabilities be given by $\hat{p}(i | \mathbf{u}_{1..\ell})$ for $i \in \mathcal{A}$. Inspection of the cost function in (23) clarifies that one can shift conditional choice probability from m to \bar{m} without losing optimality. In particular, shifting all the probability of $\hat{p}(\bar{m} | \mathbf{u}_{1..\ell})$ to $\hat{p}(m | \mathbf{u}_{1..\ell})$ yields an optimal solution that is also a feasible solution of the of the problem without the duplicate. Hence, $p^*(i | \mathbf{v}_{1..\ell}) := \hat{p}(i | \mathbf{u}_{1..\ell})$ for all $i \in \hat{\mathcal{A}} \setminus \{m, \bar{m}\}$ and $p^*(m | \mathbf{v}_{1..\ell}) := \hat{p}(m | \mathbf{u}_{1..\bar{m}}) + \hat{p}(\bar{m} | \mathbf{u}_{1..\bar{m}-1})$ solves the problem without the duplicate. On the other hand, setting $\hat{p}^*(i | \mathbf{u}_{1..\ell}) := p(i | \mathbf{v}_{1..\ell})$ for all $i \in \hat{\mathcal{A}} \setminus \{\bar{m}\}$ and $\hat{p}^*(\bar{m} | \mathbf{u}_{1..\ell}) = 0$ yields a solution of the problem with the duplicate whenever $p(i | \mathbf{v}_{1..\ell})$, $i \in \hat{\mathcal{A}} \setminus \{\bar{m}\}$, solves the problem without the duplicate.

Appendix B: General formulae for the limiting scenarios

In this section, we present the formula for $p(i | \mathbf{v})$ for special cases. Using the notation

$$p(i | \mathbf{v}) = \frac{q(i | \mathbf{v})}{\sum_{j \in \mathcal{A}} q(j | \mathbf{v})},$$

it suffices to specify $q(j | \mathbf{v})$ for $j \in \mathcal{A}$.

Identical search cost for some alternatives

If $\lambda_h = \dots = \lambda_{h+\ell}$, then

$$q(i | \mathbf{v}) = e^{\frac{v_i}{\lambda_n}} \cdot p(i)^{\frac{\lambda_1}{\lambda_n}} \cdot \prod_{k=1}^{h-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1}-\lambda_k}{\lambda_n}} \cdot \dots \cdot 1 \cdot \dots \cdot \prod_{k=h+\ell}^{n-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1}-\lambda_k}{\lambda_n}}.$$

Zero search cost for some alternatives

Consider the case where $\lambda_i = 0$ for $i \leq \ell$. Then

$$q(i | \mathbf{v}) = e^{\frac{v_i}{\lambda_n}} \cdot p(i | \mathbf{v}_{1..\ell})^{\frac{\lambda_{\ell+1}}{\lambda_n}} \cdot \prod_{k=\ell+1}^{n-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1}-\lambda_k}{\lambda_n}}.$$

Note that the DM's conditional choice probabilities are the same as if she knew the state $\mathbf{v}_{1..\ell}$ and would not distinguish between $p(i), p(i | \mathbf{v}_1), \dots$, and $p(i | \mathbf{v}_{1..\ell})$: We come to the same result if we replace $p(i | \mathbf{v}_{1..k})$ by $p(i | \mathbf{v}_{1..\ell})$ for all $k \leq \ell$ in (17).

Infinite search cost for some alternatives

If $\lambda_{h+1} = \dots = \lambda_n = \infty$, we obtain

$$q(i | \mathbf{v}) = e^{\frac{v_i}{\lambda_h}} \cdot p(i)^{\frac{\lambda_1}{\lambda_h}} \cdot \prod_{k=1}^{h-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1}-\lambda_k}{\lambda_h}}$$

for all $i \leq h$ and

$$q(i | \mathbf{v}) = e^{\frac{\bar{v}_i}{\lambda_h}} \cdot p(i)^{\frac{\lambda_1}{\lambda_h}} \cdot \prod_{k=1}^{h-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1}-\lambda_k}{\lambda_h}}$$

for all $i > h$, where $\bar{v}_i = \mathbb{E}_{p_{V_i | \mathbf{v}_{1..\ell}}} [v_i]$.

Zero and infinite search cost for some alternatives

$0 = \lambda_1 = \dots = \lambda_\ell$, $0 < \lambda_{\ell+1}, \dots, \lambda_h < \infty$, and $\lambda_{h+1} = \dots = \lambda_n = \infty$, gives

$$q(i | \mathbf{v}_{1..h}) = \begin{cases} e^{\frac{v_i}{\lambda_h}} \cdot p(i | \mathbf{v}_{1..\ell})^{\frac{\lambda_{\ell+1}}{\lambda_h}} \cdot \prod_{k=\ell+1}^{h-1} p(i | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1}-\lambda_k}{\lambda_h}} & \text{if } i \leq h, \\ e^{\frac{\mathbb{E}_{p_{V_j | \mathbf{v}_{1..\ell}}} [v_j]}{\lambda_h}} \cdot p(j | \mathbf{v}_{1..\ell})^{\frac{\lambda_{\ell+1}}{\lambda_h}} \cdot \prod_{k=\ell+1}^{h-1} p(j | \mathbf{v}_{1..k})^{\frac{\lambda_{k+1}-\lambda_k}{\lambda_h}} & \text{if } i > h. \end{cases}$$

If further $\lambda_\ell = \dots = \lambda_h =: \lambda$, this reduces to

$$q(i | \mathbf{v}_{1..h}) = \begin{cases} e^{\frac{v_i}{\lambda}} \cdot p(i | \mathbf{v}_{1..\ell}), & \text{if } i \leq h, \\ e^{\frac{\mathbb{E}_{p_{V_i | \mathbf{v}_{1..\ell}}} [v_i]}{\lambda}} \cdot p(i | \mathbf{v}_{1..\ell}), & \text{if } i > h. \end{cases}$$

Recent ESMT Working Papers

	ESMT No.
Are level 3 fair values reflected in firm value? Evidence from European banks Katja Kisseleva, ESMT European School of Management and Technology Daniela Lorenz, Free University, Berlin	16-03
Design for reusability and product reuse under radical innovation Tamer Boyacı, ESMT European School of Management and Technology Vedat Verter, Desautels Faculty of Management, McGill University Michael R. Galbreth, Moore School of Business, University of South Carolina	16-02
Pricing when customers have limited attention Tamer Boyacı, ESMT European School of Management and Technology Yalçın Akçay, College of Administrative Sciences and Economics, Koç University	16-01 (R1)
Regional state aid control in Europe: A legal and economic assessment Hans W. Friederiszick, ESMT European School of Management and Technology Massimo Merola, Bonelli Erede, College of Europe of Bruges	15-05
Fund flows inducing mispricing of risk in competitive financial markets Axel Stahmer, ESMT European School of Management and Technology	15-04
LeChatelier-Samuelson principle in games and pass-through of shocks Alexei Alexandrov, Consumer Financial Protection Bureau Özlem Bedre-Defolie, ESMT European School of Management and Technology	15-03
Contracts as a barrier to entry when buyers are non-pivotal Özlem Bedre-Defolie, ESMT European School of Management and Technology Gary Biglaiser, University of North Carolina, Chapel Hill	15-02