Rational Inattention
and Non-Compensatory Choice

David Walker-Jones

University of Toronto

david.walker.jones@mail.utoronto.ca

Keywords: rational inattention, Shannon Entropy, perceptual distance, non-compensatory choice, consideration set

JEL Classification: D83, D91

October 30, 2020

Abstract

This note studies the implications of perceptual distance for choice behavior in models of rational inattention. Using a measure for the cost of information that is more flexible than Shannon's standard measure of entropy, this note creates a new foundation for 'non-compensatory' behavior, whereby increasing the value of an option can result in a lower chance of it being selected, and demonstrates novel predictions for the formation of consideration sets. This note thus connects the literatures on rational inattention and heuristic choice rules and presents new challenges for revealed preference analysis.[1]

[1]Special thanks to Rahul Deb for all of the support. I would also like to thank Yoram Halevy and Colin Stewart for their helpful advice.

1 Introduction

Caveat emptor: this note builds upon the work done in (Walker-Jones, 2020), pointing out some interesting features of a rational inattention model with Multi-source Shannon Entropy through its simple but illustrative examples.

In the marketing literature it has long been established that agents do not consider all of the options that are available to them, and the determination of the set of options that they do consider has been the focus of an extensive body of research (Hauser & Wernerfelt, 1990). Firms, however, are interested in understanding the characteristics that lead to their product being selected, and a theory of choice typically seeks to explain the option selected by the agent. Why then is so much effort committed to understanding consideration sets, which are not necessarily observable, and are simply an intermediate stage in the choice process?

The formation of the consideration set is viewed by the marketing literature as an important process to understand in and of itself because it is thought that the factors that determine the consideration set may be different than the factors that determine which option is selected from the consideration set. This distinction can have important counterfactual implications. A firm that is trying to get their product selected must ensure that it has both the characteristics required for acceptance into the consideration set, and the characteristics required for it to be selected from the consideration set. Increasing the value of a product to the agent does not increase the chance of it being selected if it is not being considered, and if a product is not being considered, then it is crucial to identify the characteristics that determine consideration.

In the marketing literature the determination of the agent's consideration set is frequently modelled as the outcome of a heuristic style 'non-compensatory' choice rule, which have been said to fit closely with agents' self reports of how their consideration sets are formed (Hauser, Toubia, Evgeniou, Befurt, & Dzyabura, 2010; Payne, 1976), and have basis in the psychology literature (Gigerenzer & Todd, 1999). Though it is hard to find a standard definition for what 'non-compensatory' means, this note argues that the unifying feature of non-compensatory choice rules is that it is possible for an option to be made more valuable to the agent, but be selected with a lower frequency as a result.

This disparity in the marketing literature between the value of an option and its inclusion in the consideration set, which is epitomized by non-compensatory behavior, is mirrored in the economics literature where it is often assumed that the option selected from the consideration set is that which has the highest value to the agent (Manzini & Mariotti, 2014; Masatlioglu, Nakajima, & Ozbay, 2012). If the consideration set simply consisted of the options with the highest values to the agent, a theory of consideration sets would be vacuous. Thus, even if it is not stated explicitly, it seems that the economic literature's interest in consideration sets stems from their generation of non-compensatory behavior. This makes sense as non-compensatory behavior challenges the logic of revealed preference analysis, which makes welfare and counterfactual analysis more difficult.

So far, the economic literature has predominately focused on the identification of consideration sets, and their implications for choice behavior, generally assuming they are formed through an exogenous process (Manzini & Mariotti, 2014; Masatlioglu et al., 2012). This means the economic literature on consideration sets has done little to explain the foundation of non-compensatory behavior, and why it may be so prevalent.

If consideration sets are a response to a costly learning environment, as is frequently postulated in both the marketing and economic literatures, and are associated with non-compensatory choice, then a model of costly learning may seem like the perfect setting in which to study the foundation of non-compensatory choice.

Caplin, Dean, and Leahy (2018) do demonstrate that a Shannon Entropy model of rational inattention (RI) results in the endogenous formation of consideration sets based on the prior beliefs of the agent. RI with Shannon Entropy thus provides a testable framework that can be used to predict and further study the formation of consideration sets.

RI paired with Shannon Entropy also results in a dichotomy between value and choice: the most valuable option may not be selected with the highest probability if the agent thinks *a priori* that it is unlikely that the option is valuable. RI paired with Shannon Entropy, however, is incapable of explaining the behavior that is explored in the marketing literature and is exemplified by non-compensatory choice rules.

With standard Shannon Entropy, decreasing the realization of the value of an option, *ceteris paribus*, always results in that option being selected with a weakly lower probability (strictly lower probability if the option is not selected with zero probability, see Example 2). If a change is made in the characteristics of an option, and it is selected with a lower probability, then the conclusion reached by a model of RI with Shannon Entropy is that the option's value has been reduced. This is not consistent with the predictions of a non-compensatory choice rule.

When perceptual distance is incorporated into a model of RI, however, non-compensatory behavior can be generated, as is shown in this note (see Example 2). Multisource Shannon Entropy (MSSE) is a generalization of standard Shannon Entropy that allows for multiple perceptual distances so that some events can be harder to differentiate between than others (Walker-Jones, 2020). As a result, MSSE predicts non-compensatory behavior, and further predicts novel patterns for the formation of consideration sets (see Example 1).

The intuition for how multiple perceptual distances can lead to non-compensatory behavior is quite straightforward. Multiple perceptual distances being present in a learning environment essentially means that some things are harder to learn about than others. When the characteristics of an option are changed, it is possible that easier to observe characteristics seem to indicate that the option has been made worse, while harder to observe characteristics in fact indicate that the option has been improved. But, because agents face a trade-off between the quality and cost of information in a model of RI, their rational focus on less costly to observe pieces of information can lead them to select the changed option less, even if it is in fact more valuable after the change.

While MSSE does not cause consideration sets to be formed in a non-compensatory fashion in a rigorous sense, it is approximately true (ostensibly true for empirical applications) that MSSE can causes consideration sets to be formed in a non-compensatory way, as is demonstrated by Example 2. This means that finite data generated by a rationally inattentive agent who pays for information according to MSSE may appear to indicate that they form their consideration set in a non-compensatory fashion.

2 Rational Inattention

Economic agents frequently make decisions in environments with uncertainty about payoffs. In these settings agents must decide how much and what to learn before they choose between available options. Models of Rational Inattention (RI) study how the trade-off agents face between the quality of information and the cost of learning impact their choice behavior, and can provide important caveats for revealed preference analysis.

In the RI literature the learning done by the agent is typically modelled as the choice of a signal structure. Receiving a signal provides the agent with information about the state of the world, but more informative signal structures are more costly.

Suppose that the agent is learning about the probability space $(\Omega, \mathcal{F}, \mu)$, where Ω is the finite set of possible **states of the world** (the outcome space), $\mathcal{F}$ is the set of **events** (the power set of Ω), and μ is their **prior** probability measure over the states of the world, $\mu : \mathcal{F} \to [0, 1]$.

After an agent has stopped learning they make a selection from a set of **options**, denoted $\mathcal{N} = \{1, \ldots, N\}$. In any state of the world, $\omega \in \Omega$, each option, $n \in \mathcal{N}$, has a (finite) **value** to the agent of $\mathbf{v}_n(\omega)$. The agent's problem is to maximize the expected value of the option they select less the cost of learning. The agent does this by selecting an **information strategy**, which is a probability measure $F(s|\omega) : \mathbb{R} \rightarrow \mathbb{R}_+$ for each $\omega \in \Omega$. The information strategy combined with μ determines both the joint distribution $F(s,\, \omega)$, between the **signal** and the potential states of the world, and the posterior $F(\omega|s)$. In this note, as is typical in the RI literature, if $F(s, \omega)$ is optimal then the agent is done learning after a single signal s. After a signal is realized, the agent thus simply picks the action with the highest expected value:

$$a(s|F) = \arg\max_{n\in\mathcal{N}} \mathbb{E}_{F(\omega|s)}[\mathbf{v}_n(\omega)].$$

If we ignore the cost of learning momentarily, the value to the agent of receiving a signal s, which induces posterior $F(\omega|s)$, can then be written:

$$V(s|F) = \max_{n\in\mathcal{N}} \mathbb{E}_{F(\omega|s)}[\mathbf{v}_n(\omega)].$$

Let the expected cost of an information strategy, given the agent's prior, be denoted $\mathbf{C}(F(s,\, \omega),\, \mu)$. The agent's problem is thus:

$$\max_F \sum_{\omega\in\Omega} \int_s V(s|F)F(ds|\omega)\mu(\omega) - \mathbf{C}(F(s,\, \omega),\, \mu),$$

$$\text{such that } \forall\omega \in \Omega : \int_s F(ds,\, \omega) = \mu(\omega).$$

The choice behavior the agent exhibits depends on the form of $\mathbf{C}$. Frequently $\mathbf{C}$ is taken to be the expected reduction in Shannon Entropy caused by the information strategy (Sims, 2003; Matějka & McKay, 2015). This note instead measures the cost of an information strategy with the reduction in Multisource Shannon Entropy (MSSE), which is a generalisation of standard Shannon Entropy developed by Walker-Jones (2020). MSSE allows for different perceptual distances in a single learning environment. The idea behind perceptual distance is that events that are more similar should be more difficult for the agent to differentiate between. When events are more similar it is said that they have less perceptual distance between them, and it is thus more costly for the agent to differentiate between them. MSSE represents different perceptual distances with different partitions of the state space, where a partition $\mathcal{P} = \{A_1, \ldots, A_m\}$ is defined to be a set of more than one disjoint events whose union is Ω. For notational simplicity, if $\omega \in \Omega$ is the state of the world, let the **realized event** of the partition $\mathcal{P} = \{A_1, \ldots, A_m\}$ be denoted by $\mathcal{P}(\omega)$, that is $\mathcal{P}(\omega) = A_i \in \{A_1, \ldots, A_m\}$ iff $\omega \in A_i$.

Because there are a finite number of events in $\mathcal{F}$, we can order them by how difficult it is for the agent to differentiate them from their complements, and use these events to generate different partitions. Let $\mathcal{P}_{\lambda_1}$ be the (unique coarsest) partition whose elements can be used to form all of the events that are easiest to differentiate from their complements. Let $\mathcal{P}_{\lambda_2}$ be the (unique coarsest) partition whose elements can be used to form all of the events that are second easiest to differentiate from their complements. Continue in this fashion until we let $\mathcal{P}_{\lambda_M}$ be the (unique coarsest) partition whose elements can be used to form all of the events that are the hardest to differentiate from their complements. For a more complete discussion of the formation of $\mathcal{P}_{\lambda_1}, \cdots, \mathcal{P}_{\lambda_M}$, please see (Walker-Jones, 2020).

MSSE uses strictly positive constants $\lambda_1 < \ldots < \lambda_M$ to measure the total uncertainty[2] for any probability measure μ on $\mathcal{F}$ as follows:

$$C^M(\Omega, \mu) = \lambda_1 \mathcal{H}\Big(\mathcal{P}_{\lambda_1}, \mu\Big) + \mathbb{E}\Bigg[\lambda_2 \mathcal{H}\Big(\mathcal{P}_{\lambda_2}, \mu(\cdot|\mathcal{P}_{\lambda_1}(\omega))\Big) + \cdots + \lambda_M \mathcal{H}\Big(\mathcal{P}_{\lambda_M}, \mu(\cdot|\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\omega))\Big)\Bigg],$$

where $\mathcal{H}$ is Shannon Entropy as defined as in equation (2) in the Appendix, as is shown by (Walker-Jones, 2020).

We can then define $\mathbf{C}(F(s, \omega), \mu)$, the cost of an information strategy F given

[2]Each constant λ_i is associated with the partition $\mathcal{P}_{\lambda_i}$, with a larger constant corresponding to a greater difficulty of differentiating between events in $\mathcal{P}_{\lambda_i}$.

μ, as the expected reduction is MSSE it causes:

$$\mathbf{C}(F(s,\,\omega),\,\mu) = \mathbb{E}[C^M(\Omega,\,\mu) - C^M(\Omega,\,\mu(\cdot|s))].$$

This is equivalent to what is done in RI models with standard Shannon Entropy, except we use MSSE instead of Shannon Entropy to measure total uncertainty.

What we are concerned with is the choice behavior that results from the rational inattention of the agent when MSSE is used to measure the cost of information. We stop to define the notation we use to discuss this behavior before moving onto this note's results and examples.

Define the chance of option n being selected conditional on the state of the world being $\omega \in \Omega$, given the agent's optimal information strategy, to be $\Pr(n|\omega)$. For event $A \in \mathcal{F}$, define the chance of n being selected conditional on A being realized to be:

$$\Pr(n|A) = \sum_{\omega \in A} \Pr(n|\omega)\mu(\omega|A).$$

Define the **unconditional choice probability** of option n to be:

$$\Pr(n) = \sum_{\omega \in \Omega} \Pr(n|\omega)\mu(\omega).$$

3 Consideration and Non-Compensatory Choice

We now turn to our results about consideration sets and non-compensatory choice. In different papers the role of the consideration set and how it is assumed to be formed differs. Generally in the marketing literature the the formation of the consideration set is the first stage of a two stage process Gensch (1987). In the first stage of this process the agent selects a subset of all of their available options to be part of their consideration set in the second stage, with the consideration set typically being selected with a non-compensatory process, which we define shortly.

In the second stage the agent picks one option from their consideration set, but it is generally assumed all options in the consideration set have a positive probability of being selected. In the economics literature in contrast, the consideration set is frequently formed exogenously, and it is simply the option with the highest value that is selected from the consideration set. The behavior that is predicted by a model of RI with MSSE is closer in nature to the behvior assumed in the marketing literature.

Suppose that when the agent faces a set of options, $\mathcal{N}$, that an optimal information strategy is chosen and results in conditional choice probabilities $\mathbb{P} = \{\Pr(n|\omega)\}_{n=1}^{N}$, we define the **consideration set in the state of the world** ω to be $\mathcal{C}(\omega) \equiv \{n \in \mathcal{N} | \Pr(n|\omega) > 0\}$, and define the **consideration set** to be $\mathcal{C} \equiv \{n \in \mathcal{N} | \Pr(n) > 0\}$. We say an option n is **considered in the state of the world** ω if $n \in \mathcal{C}(\omega)$, and we say an option n is **considered** if $\Pr(n) > 0$, i.e. $\exists \omega$ such that $n \in \mathcal{C}(\omega)$. Our definition of a consideration set has the advantage that it can be observed in the data, fits closely with the marketing literature, and further fits with the definition given by Caplin et al. (2018).

As for 'non-compensatory' choice rules, while they are ubiquitous in the marketing literature on consideration sets, and have been touched upon in the economics literature, a standard definition of what qualifies a choice rule as 'non-compensatory' is hard to come by. Manzini and Mariotti (2012), for instance, quote the work of Tversky (1969), and in the quote Tversky refers to lexiographic semiorders as a "non-compensatory principle" (Tversky, 1969, p. 40), but neither work provides a definition of non-compensatory. Supreme Court Justice Potter Stewart's adage about pornography, "I know it when I see it" (*Jacobellis v. Ohio*, 1964), is applicable to 'non-compensatory' choice rules in most circumstances, however, and most people familiar with the phrase would agree that the definition given in this note is fitting.

We define the behavior of the agent to be **non-compensatory** if there are states of the world ω_i, $\omega_j \in \Omega$, and an option $n \in \mathcal{N}$, such that: $\mathbf{v}_m(\omega_i) = \mathbf{v}_m(\omega_j)$ $\forall m \in$

$\{\mathcal{N}\backslash n\}$, $\mathbf{v}_n(\omega_i) > \mathbf{v}_n(\omega_j)$, and under optimal behavior: $\Pr(n|\omega_i) < \Pr(n|\omega_j)$. In plainer language, the behavior of the agent is non-compensatory if, all else equal, increasing the value of an option to the agent can result in it being selected with a lower frequency.

We say the agent's **consideration set formation is non-compensatory** if there are states of the world ω_i, $\omega_j \in \Omega$, and an option $n \in \mathcal{N}$, such that: $\mathbf{v}_m(\omega_i) = \mathbf{v}_m(\omega_j)\ \forall m \in \{\mathcal{N}\backslash n\}$, $\mathbf{v}_n(\omega_i) > \mathbf{v}_n(\omega_j)$, and under optimal behavior: $n \in \mathcal{C}(\omega_j)$, while $n \notin \mathcal{C}(\omega_i)$. In plainer language, the agent forms their consideration set in a non-compensatory fashion if, all else equal, increasing the value of an option to the agent can result in it being removed from the agents consideration set in the resultant state of the world.

For those that are not familiar with non-compensatory choice rules, we now discuss two common examples that are used to form consideration sets in the marketing literature, disjunctive rules and conjunctive rules. The agent uses a **disjunctive rule** to form their consideration set if an option is considered when it features at least one of a predetermined list of characteristics: the agent considers a car when it gets 30 or more miles per gallon or it has a safety rating of at least four out of five. The agent uses a **conjunctive rule** to form their consideration set if an option is considered when it features all of a predetermined list of characteristics: the agent considers a car when it gets 30 or more miles per gallon and it has a safety rating of at least four out of five.

It is easy to see that if the agent's valuation of the car increases continuously with the characteristics of the car, then the agent forming their consideration set based on the disjunctive choice rule outlined in the previous paragraph could result in non-compensatory behavior.[3] Suppose option n gets 30 miles per gallon and has a safety rating of three out of five. Option n is considered with these attributes, but if the miles per gallon of option n is reduced by any amount, then a full point increase

[3] As was mentioned above, typically in the marketing literature, and in some economics papers, all options in the consideration set are selected with a positive probability.

in the safety rating is needed to maintain option n's membership in the consideration set. Thus, for small reductions of miles per gallon, and an increase in the safety rating of half a point, option n is more valuable to the agent, but is no longer included in the consideration set. A similar logic applies to the conjunctive choice rule outlined in the previous paragraph.

When does the agent consider a newly introduced option according to a RI model with MSSE? Suppose that the agent's problem is changed so that there is a new option x available. Is this new option considered? Let the new set of options be denoted by $\tilde{\mathcal{N}} = \mathcal{N} \cup x$, let the new optimal consideration set be denoted $\tilde{\mathcal{C}}$, and let the new optimal consideration set in the state of the world ω be denoted $\tilde{\mathcal{C}}(\omega)$.

To determine if the new option x is considered, x needs to be compared to a representative value of the options currently being considered, and given a score in each state of the world. x is considered if it scores well enough across all states of the world.

To describe when x is considered, we first need to establish what x is compared to in each state of the world. In state ω, option x is compared to a weighted average of the options currently being considered, but the weight assigned to an option n that is currently being considered depends on the unconditional probability of it being selected, as well as each of the conditional probabilities of it being selected $\Pr(n| \cap_{i=1}^{j} \mathcal{P}_{\lambda_i}(\omega))$ for $j \in \{1, \ldots M-1\}$. If $\mathbb{P}$ denotes optimal agent behavior when the set of options is $\mathcal{N}$, we define the **representative value** in any state ω of the options being considered before the introduction of x to be:

$$R(\omega) = \sum_{n \in \mathcal{C}} \Pr(n)^{\frac{\lambda_1}{\lambda_M}} (\Pr(n|\mathcal{P}_{\lambda_1}(\omega)))^{\frac{\lambda_2 - \lambda_1}{\lambda_M}} \ldots (\Pr(n| \cap_{i=1}^{M-1} \mathcal{P}_{\lambda_i}(\omega)))^{\frac{\lambda_M - \lambda_{M-1}}{\lambda_M}} e^{\frac{\mathbf{v}_n(\omega)}{\lambda_M}}.$$

In the case of Shannon Entropy (when $\lambda_1 = \ldots = \lambda_M = \lambda$), this representative value is much simpler since the weight used for each option is just its unconditional probability of being selected.

We next need to assign a score to x in each state of the world. This is easier to do. We define the **score** of option x in state ω to be:

$$s_x(\omega) = \frac{e^{\frac{\mathbf{v}_x(\omega)}{\lambda_M}}}{R(\omega)}$$

Using our new notation, we are ready to state the results that describes when x is considered in the case of standard Shannon Entropy.

Proposition 1:

In the case of Shannon Entropy:

$$\textbf{i:} \text{ if } \sum_{\omega \in \Omega} s_x(\omega)\mu(\omega) > 1, \text{ then } x \in \tilde{\mathcal{C}}.$$

$$\textbf{ii:} \text{ if } \sum_{\omega \in \Omega} s_x(\omega)\mu(\omega) < 1, \text{ then } x \notin \tilde{\mathcal{C}}.$$

Proof. All proofs appear in the Appendix. Implied by Proposition 1 in Caplin et al. (2018).

MSSE results in non-compensatory behavior, as is shown in Example 2, but with MSSE, as is true with Shannon Entropy, if an option is considered in one state of the world, then it is considered in every state of the world. With Shannon Entropy, this is evident since Theorem 2 in (Walker-Jones, 2020) implies that optimal behavior results in:

$$\Pr(n|\omega) = \frac{\Pr(n)e^{\frac{\mathbf{v}_n(\omega)}{\lambda}}}{\sum_{\nu \in \mathcal{N}} \Pr(\nu)e^{\frac{\mathbf{v}_\nu(\omega)}{\lambda}}} \quad \forall n \in \mathcal{N}, \forall \omega \in \Omega.$$

With Shannon Entropy, it is thus trivially true that if an option is considered there is a positive probability of it being selected in each state of the world: $\Pr(n) > 0 \implies \Pr(n|\omega) > 0 \ \forall \omega \in \Omega$. This result is not as evident with MSSE. While $\Pr(n)$ may be larger than zero, if $\Pr(n|\mathcal{P}_{\lambda_1}(\omega)) = 0$, for instance, then Theorem 2 from (Walker-Jones, 2020) implies $\Pr(n|\omega) = 0$. This, however, is never the case with MSSE. With

MSSE, as with Shannon Entropy, if an option is in the consideration set in one state of the world, then it is in the consideration set in every state of the world:

Proposition 2:

If $\mathbb{P}$ describes optimal agent behavior, then:

$$\forall\, n \in \mathcal{N}: \;\; n \in \mathcal{C} \Longrightarrow n \in \mathcal{C}(\omega) \;\; \forall\, \omega \in \Omega.$$

Proposition 2 tells us that with MSSE the agent considers the same options in every state of the world, and changes in characteristics (changing from one state of the world to another) cannot change which options are considered, as could be the case if consideration sets were formed according to a disjunctive or conjunctive choice rule. Thus, MSSE cannot result in non-compensatory formation of consideration sets. MSSE, however, can generate behavior that is approximately equal to a choice rule where considerations sets are determined by a non-compensatory choice rule such as a disjunctive or conjunctive choice rule, as is demonstrated in Example 2.

Further, Proposition 2 shows that it is never optimal for the agent to select a information structure that is equivalent to a partition of the state space, as it is assumed the agent does in Ellis (2018), unless it is optimal for the agent to not learn at all.

3.1 Example 1: Unique Consideration Set Formation

Consider an environment where an agent initially has two options to choose from: option 1 and option 2. For simplicity, let us assume that both option 1 and option 2 realize a high value H with probability 1/2, realize a low value $L < H$ with probability 1/2, and that the value of option 1 is perfectly negatively correlated with the value of option 2. Suppose a firm is considering introducing a third option, option 3, into the market. Assume option 3 realizes the low value L with probability

Table 1: Example 1				
State:	ω_1	ω_2	ω_3	ω_4
Probability:	1/4	1/4	1/4	1/4
Value of selecting option 1:	H	L	H	L
Value of selecting option 2:	L	H	L	H
Value of selecting option 3:	M	M	L	L

1/2, realizes a medium value M ($L < M < H$) with probability 1/2, and that the value of option 3 is drawn independently from the values of option 1 and option 2. This environment is described in Table 1. If it is introduced, does the agent consider option 3? That is, does the agent select option 3 with some positive probability if it is introduced into the market?

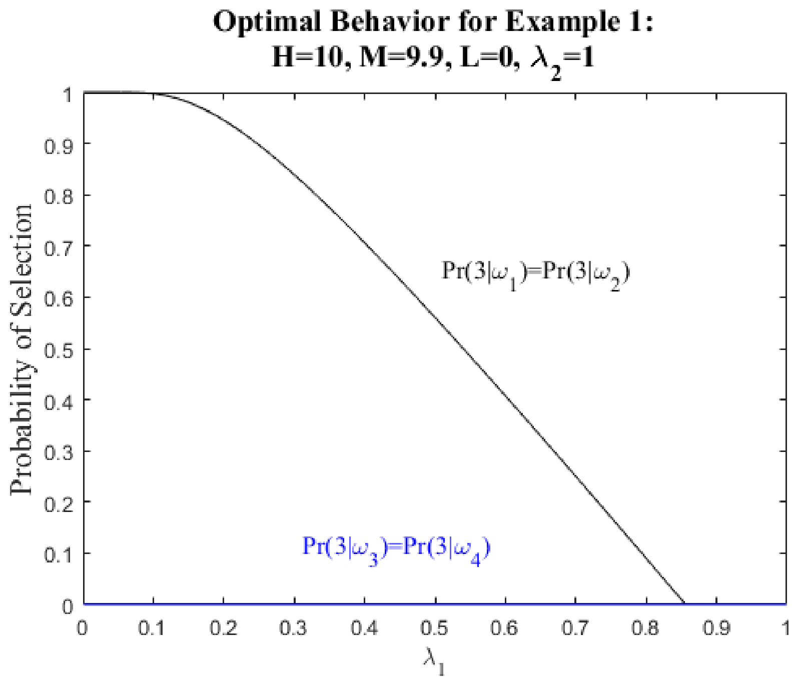
MSSE and Shannon Entropy have substantially different predictions for whether or not option 3 is included in the agent's consideration set in this environment. If the agent is assumed to do research according to Shannon Entropy, then option 3 is not considered, no matter what λ is chosen to represent the cost of research, as is shown by Proposition 1, and is discussed in the next paragraph. If the agent is instead assumed to do research according to MSSE, then there are information structures and parameter values such that option 3 is considered, as is shown later in this subsection.

If we assume that the agent pays to learn according to Shannon Entropy, then $\Pr(1) = \Pr(2)$ is optimal, since payoffs are linear, and the cost of research is shown to be strictly convex by Walker-Jones (2020). Proposition 1 then tells us option 3 is not considered since $\forall \lambda > 0$:

$$\frac{\frac{1}{2}e^{\frac{L}{\lambda}} + \frac{1}{2}e^{\frac{M}{\lambda}}}{\frac{1}{2}e^{\frac{L}{\lambda}} + \frac{1}{2}e^{\frac{H}{\lambda}}} < 1. \tag{1}$$

If we instead assume that it is easier for the agent to learn the value of option 3 than it is for them to learn the values of option 1 and option 2, then: $\mathcal{P}_{\lambda_1} = \{A_1, A_2\} = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$, and $\mathcal{P}_{\lambda_2} = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$. Since $\Pr(1|\mathcal{P}_{\lambda_1}(\omega))$ $= \Pr(2|\mathcal{P}_{\lambda_1}(\omega))$ is always optimal for each ω, as payoffs are linear and the cost of re-

Figure 1:



search is convex, if:

$$\frac{e^{\frac{M}{\lambda_2}}}{\frac{1}{2}e^{\frac{L}{\lambda_2}} + \frac{1}{2}e^{\frac{H}{\lambda_2}}} > 1,$$

which would be the case if M is close to H, then the agent considers option 3 when λ_1 is close to 0 based on its merits in A_1, as can be seen in Figure 1. This makes a lot of sense intuitively. If λ_1 is close to 0, then it is essentially free for the agent to observe the value of option 3. If they observe that option 3 takes value M, and M is close enough to H (relative to λ_2), then we should expect the agent to select option 3.

Those familiar with the work of Huettner, Boyacı, and Akçay (2019) may notice that the behavior described in Figure 1 seems to contradict their Theorem 2. When option 3 is not available, $\forall \omega : \Pr(1) = \Pr(2) = \Pr(1|\mathcal{P}_{\lambda_1}(\omega)) = \Pr(2|\mathcal{P}_{\lambda_1}(\omega)) = 1/2$ is optimal (again since payoffs are linear and the cost of research is strictly convex), and so equation (1) (with $\lambda = \lambda_2$) tells us that Theorem 2 from Huettner et al. (2019)

indicates $\Pr(3) = 0$ is part of an optimal solution when option 3 is introduced. This is not correct for small λ_1 and M close to H. If $L = 0$, $M = 9.9$, $H = 10$, and $\lambda_2 = 1$, as in the environment depicted in Figure 1, then for small λ_1, Theorem 2 from Huettner et al. (2019) contradicts both Corollary 1 from Walker-Jones (2020), and Lemma 1 from Huettner et al. (2019), since:

$$\log\left(\frac{1}{2}e^{\frac{L}{\lambda_2}} + \frac{1}{2}e^{\frac{H}{\lambda_2}}\right) < \log\left(e^{\frac{M}{\lambda_2}}\right)$$

$$\implies \log\left(\frac{1}{2}e^{\frac{L}{\lambda_2}} + \frac{1}{2}e^{\frac{H}{\lambda_2}}\right) < \frac{1}{2}\log\left(e^{\frac{M}{\lambda_2}}\right) + \frac{1}{2}\log\left(\frac{1}{2}e^{\frac{L}{\lambda_2}} + \frac{1}{2}e^{\frac{H}{\lambda_2}}\right).$$

Thus, behavior cannot be optimal if option 3 is not considered when it is introduced, and there is a problem with Theorem 2 from Huettner et al. (2019).

3.2 Example 2: Non-Compensatory Behavior

Consider the environment described in Table 2, where $L < M < H$, and $\epsilon > 0$. If the agent pays for learning in this environment according to Shannon Entropy, then the agent always has a higher probability of selecting option 3 in ω_4 than ω_1, unless their probability of selecting option 3 is 0 in both, since the only difference between ω_1 and ω_4 is that option 3 has a higher value in ω_4. With MSSE, in contrast, under certain information structures and parameter values, there is a probability close to one of option 3 being selected in ω_1, and a probability close to zero of option 3 being selected in ω_4, which would constitute non-compensatory behavior.

Table 2: Example 2

State:	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
Probability:	1/6	1/6	1/6	1/6	1/6	1/6
Value of selecting option 1:	M	H	L	M	H	L
Value of selecting option 2:	M	L	H	M	L	H
Value of selecting option 3:	M	M	M	$H+\epsilon$	L	L

Assume, for instance, that it is easier for the agent to learn whether or not the value of option 3 is M, and all other learning is the same level of difficulty:

$$\mathcal{P}_{\lambda_1} = \{A_1, A_2\} = \{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5, \omega_6\}\},$$

$$\text{and } \mathcal{P}_{\lambda_2} = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_5\}, \{\omega_6\}\}.$$

Note: $\Pr(1|\mathcal{P}_{\lambda_1}(\omega)) = \Pr(2|\mathcal{P}_{\lambda_1}(\omega))$ is always optimal for all ω, as payoffs are linear and the cost of research is convex. Figure 2 and Figure 3 depict optimal agent behavior in this environment. As you can see, if λ_1 is close to zero, then $\Pr(3|\omega_1) > \Pr(3|\omega_4)$, even though the only difference between ω_1 and ω_4 is that option 3 has a higher value in ω_4. MSSE can thus generate non-compensatory behavior.

Proposition 2 tells us that MSSE cannot result in consideration sets being formed in a non-compensatory fashion since the consideration set is the same in each state of the world. In some cases, however, it is approximately true that MSSE results in the consideration set being formed according to a non-compensatory choice rule. Consider the setting of Example 2, and assume that λ_1 is close to zero, the three options are cars, and that the realization of $\mathcal{P}_{\lambda_1}$ determines the safety rating and miles per gallon of the three options. Then, if in A_2 option 1 and option 2 have characteristics that satisfy the conjunctive choice rule outlined above, and option 3 does not, while in A_1 option 3 has characteristics that satisfy the conjunctive choice rule outlined above, and option 1 and option 2 do not, then it is approximately true that the agent's consideration set is formed according to the non-compensatory conjunctive choice rule outline above, as is shown in Figure 2 and Figure 3 (it is true in the limit as λ_1 goes to zero).

Further, in such a setting, if there are multiple options in the approximate consideration set, then it appears that the agent is selecting from their consideration set based on a RU model that is described by a logit regression where the values are normalized by λ_2.

Figure 2:
Optimal Behavior for Example 2:
H=10, M=9.9, L=0, ϵ=0.1, λ_2=1

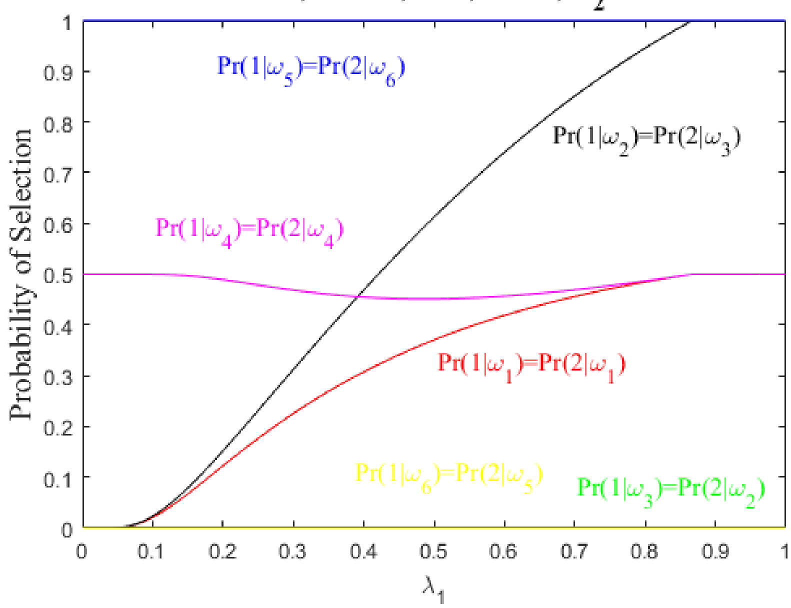
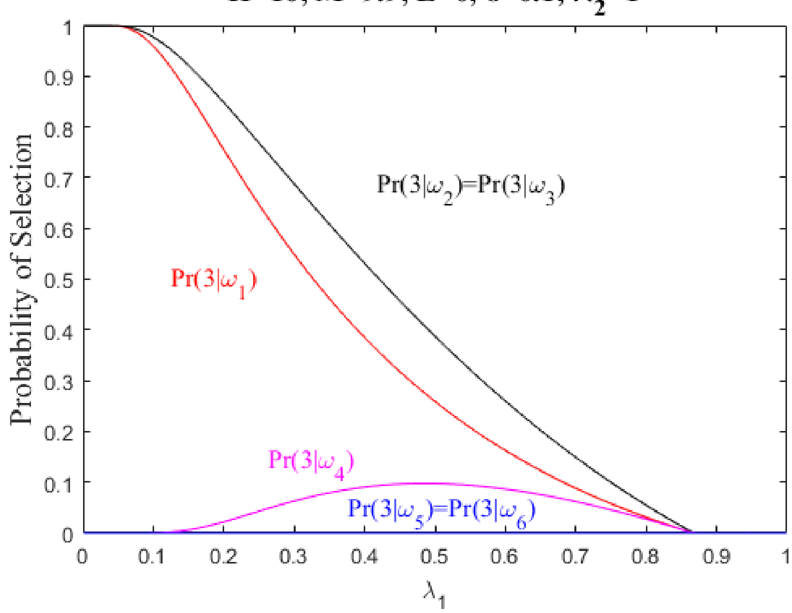


Figure 3:
Optimal Behavior for Example 2:
H=10, M=9.9, L=0, ϵ=0.1, λ_2=1



This description of behavior, a combination of a non-compensatory choice rule and a logit regression in a two-stage process, is very similar to fitting that is carried out in the marketing literature. For example, Gensch (1987) carries out exactly this kind of fitting, first narrowing down the agent's options in a first stage with a maximum-likelihood-hierarch model, then fitting a logit model in the second stage. Gensch (1987) further notes that the maximum-likelihood-hierarch model in the first stage could be replaced by a conjunctive or disjunctive rule.

4 Conclusion

This note studies the implications of perceptual distance for choice behavior in models of rational inattention. Using Multisource Shannon Entropy (MSSE), a measure for the cost of information developed by Walker-Jones (2020) that is more flexible than Shannon's standard measure of entropy, this note creates a new foundation for 'non-compensatory' behavior, whereby increasing the value of an option can result in a lower chance of it being selected. Further, the new measure for information is shown to create novel predictions for the formation of consideration sets. This note thus connects the literatures on rational inattention and heuristic choice rules and presents new challenges for revealed preference analysis. The behavior generated by a model of RI with MSSE is also shown to be approximately the same as the behavior assumed by multi-staged models in the marketing literature.

Appendix

Shannon Entropy is a measure of uncertainty. If we are given a partition of the states of the world $\mathcal{P} = \{A_1, \ldots, A_m\}$, and probability measure μ over these events, the uncertainty about which event has occurred, as measured by **Shannon Entropy**, is defined as:[4]

$$\mathcal{H}(\mathcal{P}, \mu) = -\sum_{i=1}^{m} \mu(A_i) \log(\mu(A_i)). \tag{2}$$

When applying Shannon Entropy the convention is to set $0 \log(0) = 0$.

A partition can generate a sigma algebra. Let $\sigma(\mathcal{P})$ denote the sigma algebra generated by the partition $\mathcal{P}$, which is the smallest sigma algebra that contains all the elements of $\mathcal{P}$ (intersection of all sigma algebras containing all of the elements of $\mathcal{P}$). Several partitions can also generate a sigma algebra. Let $\sigma(\mathcal{P}_1, \ldots, \mathcal{P}_n)$ denote the sigma algebra generated by the partitions $\mathcal{P}_1, \ldots, \mathcal{P}_n$, which is the smallest sigma algebra that contains all the elements of each of $\mathcal{P}_1, \ldots, \mathcal{P}_n$ (intersection of all sigma algebras containing all of the elements of each of $\mathcal{P}_1, \ldots, \mathcal{P}_n$). We can also use several partitions to create a finer partition. Let $\times\{\mathcal{P}_i\}_{i=1}^n$ denote the unique partition such that $\sigma(\times\{\mathcal{P}_i\}_{i=1}^n) = \sigma(\mathcal{P}_1, \ldots, \mathcal{P}_n)$.

Proof of Proposition 1. Proposition 1 in this note is implied by Proposition 1 in Caplin et al. (2018). Proof follows for those that are interested. The Lagrangian for the problem described in Corollary 1 in Walker-Jones (2020) when the set of alternatives is $\mathcal{N}$ is:

$$\mathcal{L} = \left(\sum_{\omega \in \Omega} \log \left(\sum_{n \in \mathcal{N}} \Pr(n) e^{\frac{\mathbf{v}_n(\omega)}{\lambda}} \right) \mu(\omega) \right) + \sum_{n \in \mathcal{N}} \xi_n \Pr(n) - \gamma \left(\sum_{n \in \mathcal{N}} \Pr(n) - 1 \right)$$

The objective is concave since payoffs are linear and research costs are convex. The objective is also differentiable. The Karush-Khun Tucker conditions are thus necessary and sufficient for an optimal solution (Lange, 2013). Taking first order conditions

[4] This measure is only unique up to a positive multiplier.

with respect to $\Pr(n)$ gives:

$$\sum_{\omega\in\Omega}\left(\frac{e^{\frac{\mathbf{v}_n(\omega)}{\lambda}}}{\sum\limits_{\nu\in\mathcal{N}}\Pr(\nu)e^{\frac{\mathbf{v}_\nu(\omega)}{\lambda}}}\mu(\omega)\right)+\xi_n=\gamma$$

Further, for each $n\in\mathcal{N}$ with $\Pr(n)>0$, which there must be at least one of, $\xi_n=0$ and Theorem 2 implies:

$$\sum_{\omega\in\Omega}\Pr(n|\omega)\mu(\omega)=\Pr(n)\implies\sum_{\omega\in\Omega}\frac{\Pr(n)e^{\frac{\mathbf{v}_n(\omega)}{\lambda}}}{\sum\limits_{\nu\in\mathcal{N}}\Pr(\nu)e^{\frac{\mathbf{v}_\nu(\omega)}{\lambda}}}\mu(\omega)=\Pr(n)$$

$$\implies\sum_{\omega\in\Omega}\frac{e^{\frac{\mathbf{v}_n(\omega)}{\lambda}}}{\sum\limits_{\nu\in\mathcal{N}}\Pr(\nu)e^{\frac{\mathbf{v}_\nu(\omega)}{\lambda}}}\mu(\omega)=1$$

Which thus implies $\gamma=1$. If:

$$\sum_{\omega\in\Omega}\left(\frac{e^{\frac{\mathbf{v}_x(\omega)}{\lambda}}}{\sum\limits_{\nu\in\mathcal{N}}\Pr(\nu)e^{\frac{\mathbf{v}_\nu(\omega)}{\lambda}}}\mu(\omega)\right)>1$$

then $\Pr(x)=0$ cannot be optimal since $\xi_x\geq 0$, and thus $\Pr(x)>0$. Further, if $\Pr(x)=0$ is optimal, then since $\xi_x\geq 0$, it must be that:

$$\sum_{\omega\in\Omega}\left(\frac{e^{\frac{\mathbf{v}_x(\omega)}{\lambda}}}{\sum\limits_{\nu\in\mathcal{N}}\Pr(\nu)e^{\frac{\mathbf{v}_\nu(\omega)}{\lambda}}}\mu(\omega)\right)\leq 1.\blacksquare$$

Proof of Proposition 2. The Lagrangian for the problem described in Corollary 1 from Walker-Jones (2020) is:

$$\mathcal{L}=\sum_{\omega\in\Omega}\left(\log\left(\sum_{n=1}^{N}\Pr(n)^{\frac{\lambda_1}{\lambda_M}}\Pr(n|\mathcal{P}_{\lambda_1}(\omega))^{\frac{\lambda_2-\lambda_1}{\lambda_M}}\dots\Pr(n|\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\omega))^{\frac{\lambda_M-\lambda_{M-1}}{\lambda_M}}e^{\frac{\mathbf{v}_n(\omega)}{\lambda_M}}\right)\mu(\omega)\right)$$

$$+\sum_{A\in\times\{\mathcal{P}_{\lambda_i}\}_{i=1}^{M-1}}\sum_{n\in\mathcal{N}}\xi_n(A)\text{Pr}(n|A)-\sum_{A\in\times\{\mathcal{P}_{\lambda_i}\}_{i=1}^{M-1}}\gamma(A)\Big(\sum_{n\in\mathcal{N}}\text{Pr}(n|A)-1\Big)$$

Using Theorem 2 from Walker-Jones (2020), the first order condition with respect to $\text{Pr}(n|A)$ for some $A\in\times\{\mathcal{P}_{\lambda_i}\}_{i=1}^{M-1}$ and $\tilde{\omega}\in A$ is then:

$$\left(\sum_{\omega\in\Omega}\frac{\lambda_1\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\lambda_M\text{Pr}(n)}\text{Pr}(n|\omega)\mu(\omega)\right)+\left(\sum_{\omega\in\mathcal{P}_{\lambda_1}(\tilde{\omega})}\frac{(\lambda_2-\lambda_1)\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\lambda_M\text{Pr}(n|\mathcal{P}_{\lambda_1}(\omega))\mu(\mathcal{P}_{\lambda_1}(\tilde{\omega}))}\text{Pr}(n|\omega)\mu(\omega)\right)$$

$$+\cdots+\left(\sum_{\omega\in\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega})}\frac{(\lambda_M-\lambda_{M-1})\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\lambda_M\text{Pr}(n|\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\omega))\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}\text{Pr}(n|\omega)\mu(\omega)\right)+\xi_n(A)=\gamma(A)$$

For $n\in\mathcal{N}$ with $\text{Pr}(n|A)>0$, which there must be at least one of, $\xi_n(A)=0$, and:

$$\sum_{\omega\in\Omega}\frac{\lambda_1\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\lambda_M\text{Pr}(n)}\text{Pr}(n|\omega)\mu(\omega)=\frac{\lambda_1\mu(A)}{\lambda_M},$$

and for for each $m\in\{1,\ldots,M-1\}$:

$$\sum_{\omega\in\cap_{i=1}^{m}\mathcal{P}_{\lambda_i}(\tilde{\omega})}\frac{(\lambda_{m+1}-\lambda_m)\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\lambda_M\text{Pr}(n|\cap_{i=1}^{m}\mathcal{P}_{\lambda_i}(\omega))\mu(\cap_{i=1}^{m}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}\text{Pr}(n|\omega)\mu(\omega)=\frac{(\lambda_{m+1}-\lambda_m)\mu(A)}{\lambda_M}.$$

This tells us:

$$\gamma(A)=\mu(A)\Big(\frac{\lambda_1}{\lambda_M}+\frac{\lambda_2-\lambda_1}{\lambda_M}+\cdots+\frac{\lambda_M-\lambda_{M-1}}{\lambda_M}\Big)=\mu(A),\quad\forall A\in\times\{\mathcal{P}_{\lambda_i}\}_{i=1}^{M-1}.$$

The rest of the proof proceeds as a proof by contradiction. Assume there is an alternative $n\in\mathcal{N}$ such that $\text{Pr}(n)>0$, and $\exists\,A\in\times\{\mathcal{P}_{\lambda_i}\}_{i=1}^{M-1}$ such that $\text{Pr}(n|A)=0$. This means for some $\tilde{\omega}\in A$ there is a $m\in\{1,\ldots,M-1\}$ such that $\text{Pr}(n|\cap_{i=1}^{m}\mathcal{P}_{\lambda_i}(\tilde{\omega}))=0$ and $\text{Pr}(n|\cap_{i=1}^{m-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))>0$. So that our objective is differentiable, we must bound $\text{Pr}(n|A)$ below by small $\epsilon>0$ (so that the multiplier $\xi_n(A)$ corresponds to the constraint $\text{Pr}(n|A)\geq\epsilon$), and for each of the k options ν such that $Pr(\nu|A)>0$, reduce $Pr(\nu|A)$ by ϵ/k (so our probabilities sum to one). Then, when we let ϵ go to zero, the first order condition with respect to $\Pr(n|A)$ is thus:

$$\left(\sum_{\omega\in\Omega}\frac{\lambda_1\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\lambda_M\Pr(n)}\Pr(n|\omega)\mu(\omega)\right)+\left(\sum_{\omega\in\mathcal{P}_{\lambda_1}(\tilde{\omega})}\frac{(\lambda_2-\lambda_1)\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\lambda_M\Pr(n|\mathcal{P}_{\lambda_1}(\omega))\mu(\mathcal{P}_{\lambda_1}(\tilde{\omega}))}\Pr(n|\omega)\mu(\omega)\right)$$

$$+\cdots+\left(\sum_{\omega\in\cap_{i=1}^{m-1}\mathcal{P}_{\lambda_i}(\tilde{\omega})}\frac{(\lambda_m-\lambda_{m-1})\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\lambda_M\Pr(n|\cap_{i=1}^{m-1}\mathcal{P}_{\lambda_i}(\omega))\mu(\cap_{i=1}^{m-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}\Pr(n|\omega)\mu(\omega)\right)$$

$$+\left(\sum_{\omega\in A}\frac{(\lambda_M-\lambda_m)}{\lambda_M\Pr(n|A)^{\frac{\lambda_m}{\lambda_M}}}\frac{\Pr(n)^{\frac{\lambda_1}{\lambda_M}}\Pr(n|\mathcal{P}_{\lambda_1}(\omega))^{\frac{\lambda_2-\lambda_1}{\lambda_M}}\ldots\Pr(n|\cap_{i=1}^{m-1}\mathcal{P}_{\lambda_i}(\omega))^{\frac{\lambda_m-\lambda_{m-1}}{\lambda_M}}e^{\frac{\mathbf{v}_n(\omega)}{\lambda_M}}}{\sum_{\nu\in\mathcal{N}}\Pr(\nu)^{\frac{\lambda_1}{\lambda_M}}\Pr(\nu|\mathcal{P}_{\lambda_1}(\omega))^{\frac{\lambda_2-\lambda_1}{\lambda_M}}\ldots\Pr(\nu|\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\omega))^{\frac{\lambda_M-\lambda_{M-1}}{\lambda_M}}e^{\frac{\mathbf{v}_\nu(\omega)}{\lambda_M}}}\mu(\omega)\right)$$

$$\cdot\left(\left(\frac{\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\mu(\cap_{i=1}^{m}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}\right)^{\frac{\lambda_{m+1}-\lambda_m}{\lambda_M}}\cdot\ldots\cdot\left(\frac{\mu(\cap_{i=1}^{M-1}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}{\mu(\cap_{i=1}^{M-2}\mathcal{P}_{\lambda_i}(\tilde{\omega}))}\right)^{\frac{\lambda_{M-1}-\lambda_{M-2}}{\lambda_M}}\right)+\xi_n(A)=\mu(A),$$

which cannot be satisfied for $\Pr(n|A)=\epsilon$ since $\xi_n(A)\geq 0$, and the last sum goes to infinite as ϵ goes to zero. ■

References

Caplin, A., Dean, M., & Leahy, J. (2018). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*, *86*(3), 1061–1094.

Ellis, A. (2018). Foundations for optimal inattention. *Journal of Economic Theory*, *173*, 56–94.

Gensch, D. H. (1987). A two-stage disaggregate attribute choice model. *Marketing Science*, *6*(3), 223–239.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics that make us smart* (pp. 3–34). Oxford University Press.

Hauser, J. R., Toubia, O., Evgeniou, T., Befurt, R., & Dzyabura, D. (2010). Disjunctions of conjunctions, cognitive simplicity, and consideration sets. *Journal of Marketing Research*, *47*(3), 485–496.

Hauser, J. R., & Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of consumer research*, *16*(4), 393–408.

Huettner, F., Boyacı, T., & Akçay, Y. (2019). Consumer choice under limited attention when alternatives have different information costs. *Operations Research*.

Jacobellis v. ohio. (1964). Retrieved from https://www.law.cornell.edu/supremecourt/text/378/184

Lange, K. (2013). *Optimization* (Second Edition ed.; G. Casella, I. Olkin, & S. Fienberg, Eds.). Springer.

Manzini, P., & Mariotti, M. (2012). Choice by lexicographic semiorders. *Theoretical Economics*, *7*(1), 1–23.

Manzini, P., & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, *82*(3), 1153–1176.

Masatlioglu, Y., Nakajima, D., & Ozbay, E. Y. (2012). Revealed attention. *The American Economic Review*, 2183–2205.

Matějka, F., & McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, *105*(1), 272–98.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational behavior and human performance*, *16*(2), 366–387.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, *50*(3), 665–690.

Tversky, A. (1969). Intransitivity of preferences. *Psychological review*, *76*(1), 31.

Walker-Jones, D. (2020, October). *Rational inattention and perceptual distance.* Retrieved from https://www.dwalkerjones.com