Attention Overload

Paul H.Y. Cheung[*] Yusufcan Masatlioglu[†]

March 6th, 2020

Abstract

Consumer has attention scarcity. When there are more products in the market, it is natural to expect that each product gets less attention due to the competition over attention. We call such phenomenon *attention overload*. However, existing models either fall short of capturing attention overload, or is inept to unbind the notions of considering and choosing. We propose a model of random attention, *Attention Overload Model*, under this innocuous behavioral assumption. We show how policy maker can learn about consumer preference purely from choice data and make welfare judgement. We demonstrate how an additional condition over attention at binaries further informs the policy maker. Characterization results are also provided for outlining the key insights to look for in the choice data for consistency.

1 Introduction

Decision-making is becoming a bustling task for consumers due to the abundance of options. For example, Amazon US sells more than 606 million products (87 million products in Home Kitchen and 62 million Books). This phenomenon is also witnessed in other domains such as healthcare plans, car insurance, or financial services. It is without doubt that consumer cannot pay attention to all products: some are going to be more appealing than others, while some are completely unnoticed. The proliferation of options forces each product must compete with each other for consumers' attention.[1] This phenomenon is known as "choice overload" in the psychology literature.

Market competitions over consumer's attention are fierce. According to eMarketer, the US has spent over \$100 billions in advertisement in 2018, where roughly half of the spending comes

[*]Department of Economics, University of Maryland.

[†]Department of Economics, University of Maryland.

[1]The limited attention has been illustrated in different markets: investment decisions (Huberman and Regev, 2001), school choice (Rosen, Curran, and Greenlee, 1998), job search (Sheridan, Richards, and Slocum, 1975), household grocery consumption (Demuynck and Seel, 2018), PC purchases (Goeree, 2008), university choice (Laroche, Rosenblatt, and Sinclair, 1984; Rosen et al., 1998), and airport choice (Başar and Bhat, 2004).

from digital media. This fierce competition makes consumers' attention problem more difficult. According to recent Ipsos eye-tracking research, the majority of TV advertising time (55%) is not paid attention to due to multitasking, switching channels, and fast-forwarding. In other words, it is possible that the same customer may attend to different products in the same environment on different occasions. The random attention idea has been utilized by new theoretical models (Manzini & Mariotti, 2014; Aguiar, 2015; Brady & Rehbeck, 2016; Horan, 2018; Cattaneo, Ma, Masatlioglu, & Suleymanov, 2019). In this stochastic environment, we define the amount of attention a product receives as the *attention frequency*.

Visschers, Hess, and Siegrist (2010) suggests that a competition for attention gets more aggressive when the number of alternatives gets larger. For example, if a product grabs the consumer's consideration in a large supermarket, then it will grab her attention in a small convenience store with fewer rivals. Hence the attention frequency for each alternative might decrease as the number of rivals increases (see Reutskaja and Hogarth (2009), Reutskaja, Nagel, Camerer, and Rangel (2011), Geng (2016)). We call this property as *attention overload*, which is the key assumption for this paper.

Choice overload is the outcome of the decision maker's cognitive limitations, it thus cannot be directly observed. From a policy-making and welfare perspective, it is important to identify whether and when larger choice sets lead to choice overload. To do this, we first reveal the preference of the individual, which are not directly observable under limited attention. Our aim is to uncover preferences solely from observed choices when attention is limited and the products compete for consumers' attention. Following the traditional insight of economics, we assume that consumers have a complete and transitive preference over the alternatives. Consumer are also assumed to pick the best alternative in their consideration sets. The key assumption is that consideration is random and unobservable for the modeler or the econometrician. With our non-parametric attention overload assumption, we can reveal preference based solely on choice behavior. Our identification informs policy makers the nature of choice overload from observed choices.

Existing models either cannot capture attention overload, or they fail to disentangle considering an alternative from choosing an alternative. Manzini and Mariotti (2014) considers an attention model with independent consideration. In their model, any alternative that is considered with positive probability entail a positive probability to be chosen. However, it is undesirable since it precludes the possibility that an inferior alternative is always considered with a better alternative (e.g. attraction effect), such that it is never chosen even when it has a positive probability to be considered. A similar feature can be also found in the rational inattention literature. Caplin, Dean, and Leahy (2019) suggests that any item that is considered are always chosen with positive probability. On the other hand, Cattaneo et al. (2019) and Brady and Rehbeck (2016) fail to satisfy the attention overload assumption. Hence, it is possible that an alternative is getting less attention even when the choice set gets smaller.

One of the behavioral consequences of choice overload is the likelihood of deferring choice. That is, when faced with too many unfamiliar choices, consumers choose the outside option (Iyengar and Lepper (2000)). To able to capture this notion, we introduce an outside option in our model. We show that our model is compatible with the fact that the outside option is chosen more often when the choice set size increases. The intuition is simple: the competition between alternatives get more fierce as the decision problems get bigger, the decision maker tends not to consider any alternative and choose the outside option. Interestingly, existing models of random attention rule, e.g. Manzini and Mariotti (2014), Brady and Rehbeck (2016) and Cattaneo et al. (2019), predict the other way.[2]

There are a number of insightful special cases of this model. Firstly, the idea of competition filter in Lleras, Masatlioglu, Nakajima, and Ozbay (2017) is a special case of the model in the deterministic environment, which says that any item which wins the consumer's attention would also prevail in the smaller set. Secondly, surprisingly, Manzini and Mariotti (2014) is a knife-edge special case of the model, where the attention frequency of an alternative is held constant throughout any set. Notions of rationalization (e.g. Cherepanov, Feddersen, and Sandroni, 2013), categorization (e.g. Aguiar (2017) and Manzini and Mariotti (2012)) and narrowing down (e.g. Lleras et al. (2017)) are nested in the model. By being special case of this property, our paper subsumes the models above. It means that any revealed preference in the using the simple non-parametric restriction would also hold in their environment. We discuss it further in Section 5.

However, due to the generality of the model, for some given data set, some preference over alternative may not be all identified. In other words, the revealed preference needs not be complete. In order to complement this property and reinforce revealed preference, we also investigate non-parametric restriction over binaries. We shows that it aids in revealed preference and we provide characterization result for the joint non-parametric restriction. On the other hand, we also show that the condition is weaker (i.e. more general) than the one considered in existing literature Cattaneo et al. (2019).

In the next section, we introduce the setup and the model. We discuss the revealed preference

[2] i.e. the outside option is chosen more often in the smaller set. For Manzini and Mariotti (2014) and Brady and Rehbeck (2016), the restriction lies in formation of attention rule. For Cattaneo et al. (2019), the restriction comes from the monotonic attention rule.

and characterization in section 3. Applications of the model are explored in Section 4. Lastly, we follow with a discussion on related literature in Section 5. Conclusion is in Section 6.

2 Choice under Attention Overload

We denote the grand alternative set as X, which are held fixed throughout the paper A typical element of X is denoted by a and its cardinality is $|X| = N$. We let $\mathcal{X}$ denote the set of all non-empty subsets of X. Each member of $\mathcal{X}$ defines a choice problem.

Definition 1 (Choice Rule)**.** A choice rule is a map $\pi : X \times \mathcal{X} \to [0, 1]$ such that for all $S \in \mathcal{X}$, $\pi(a|S) \geq 0$ for all $a \in S$, $\pi(a|S) = 0$ for all $a \notin S$, and $\sum_{a \in S} \pi(a|S) = 1$.

Therefore, $\pi(a|S)$ represents the probability that the decision maker chooses alternative a from the choice problem S. To see how the formulation allows deterministic choice rules. Take $\pi(a|S)$ as either 0 or 1, then choices are deterministic. The key ingredient of our model is probabilistic consideration sets. Given a choice problem S, each non-empty subset of S could be a consideration set with certain probability. It is natural to assume that each frequency is between 0 and 1 and that the total frequency adds up to 1. Formally,

Definition 2 (Attention Rule)**.** An attention rule is a map $\mu : \mathcal{X} \times \mathcal{X} \to [0, 1]$ such that for all $S \in \mathcal{X}$, $\mu(T|S) \geq 0$ for all $T \subset S$, $\mu(T|S) = 0$ for all $T \not\subset S$, and $\sum_{T \subseteq S} \mu(T|S) = 1$.

Thus, $\mu(T|S)$ represents the probability of paying attention to the consideration set $T \subset S$ when the choice problem is S. This formulation also allows for deterministic attention rules. For example, $\mu(S|S) = 1$ represents an agent with full attention.

Another economically important variable we would like to keep track of is the amount of attention each *alternative* captures for a given μ. We can observe this information simply from μ by summing up the frequency of consideration sets containing the alternative. That is, for a given μ, the probability that a attracts attention in S is defined as:

$$\phi_\mu(a|S) := \sum_{a \in T \subseteq S} \mu(T|S)$$

When μ is clearly defined, we omit μ for convenience. Hence, $\phi(a|S)$ is the measure of attention for a in S. In deterministic attention model, the attention that one alternative receive is either zero or one. i.e. whether it is being considered or not. Yet, in stochastic environment, attention is probabilistic. This also means that the attention one alternative receives does not necessary be a zero-or-one dichotomy. We regard it as the *attention frequency* of an alternative.

When consumers are overwhelmed by the abundance of options, every product competes for consumers' attention. This implies that as the number of the alternatives increases, the competition gets more severe. That is, the frequency of attending to a product should weakly decrease when the set of available alternatives is expanded by adding more options to it. We call this property *Attention Overload*.

Assumption 1 (Attention Overload)**.** For any $a \in T \subseteq S$, $\phi(a|S) \leq \phi(a|T)$.[3]

We says that an attention rule is *overloaded* if its corresponding ϕ satisfies attention. Several models listed in Introduction satisfy this assumption. Take Manzini and Mariotti (2014) as an example, the attention frequency is held fixed for each alternative. Having this non-parametric restriction, the choice rule can be defined accordingly. A DM who follows AOM maximize his utility according to a preference ordering $\succ$ under each realized consideration set.

Definition 3. A choice rule π has a *attention overload representation* in $\succ$ if there exists a preference ordering $\succ$ over X and a overloaded attention rule μ such that

$$\pi(a|S) = \sum_{T \subseteq S} \mathbb{1}(a \text{ is } \succ\text{-best in } T) \cdot \mu(T|S)$$

for all $a \in S$ and $S \in \mathcal{X}$. In this case, we say π is represented by $(\succ, \mu)$. We also say π is a *Attention Overload Model* (AOM).

3 Revealed Preference and Characterization

Given choice data that satisfies the attention overload model, is it possible to identify consumer's underlying preference? We show how it can be done in the following. To achieve this, we exploit the fact that attention frequency satisfies attention overload in attention overload model. Note that in attention overload model, each alternative gets more attention when choice set is smaller. Hence, it is natural to *expect* that each alternative would be more likely to be picked in a smaller choice set. However, if we observe the counterfactual, i.e. an alternative has a lower probability of being selected, we can deduce that there must be something better than it in the smaller choice set. Let us first define what it means to be revealed preference in the model.

Definition 4 (Revealed Preference)**.** Let $\{(\succ_j, \mu_j)\}_{j=1,\ldots,J}$ be all attention overload representations of π. We say that b is *revealed to be preferred* to a if $b \succ_j a$ for all j.

[3] If we allow the consideration set to be empty, then we should also require that the frequency of paying attention to nothing decreases when the choice set shrinks to capture choice overload. We discuss this in further detail in Section 4.1.

This reveal preference definition checks all possible representation for choice data π and make conclusion of a is revealed to be preferred to b only if all possible representations agree. This conservative specification is also employed in Cattaneo et al. (2019) and other deterministic limited consideration models. Up next, we have the identification for revealed preference.

Our first observation is that a regularity violation at binary set reveals the decision maker's preference. More specifically, we illustrate that if $\pi(a|S) > \pi(a|\{a,b\})$, then b must be preferred to a. To reach such a conclusion, we must show that in any representation of π, say $(\succ, \mu)$, we have $b \succ a$. Let assume contrary: there exists $(\succ, \mu)$ representing π such that $a \succ b$ and μ satisfies attention overload. First note that attention is a perquisite for choice. To be able to choose an alternative, the decision maker first must pay attention to it. Hence, the attention frequency is always greater (or equal) to the choice probability for any alternative and in any choice set $\pi(a|S) \leq \phi(a|S)$. In addition, they are equal for the best alternative in any choice set: $\pi(a_S|S) = \phi(a_S|S)$ where a_S is the $\succ$-best alternative in S. Given $a \succ b$, we have

$$\phi_\mu(a|\{a,b\}) = \pi(a|\{a,b\}) < \pi(a|S) \leq \phi_\mu(a|S)$$

This contradicts with the fact that μ satisfies attention overload. The next lemma formally states this observation.

Lemma 1. Let π be a AOM. If $\pi(a|S) > \pi(a|\{a,b\})$, then b is revealed to be preferred to a.

Proof.

$$\begin{aligned}
& \phi(a|\{a,b\}) \geq \phi(a|S) && \\
\Rightarrow\quad & \sum_{a\in J\subseteq\{a,b\}} \mu(J|\{a,b\}) \geq \sum_{a\in J\subseteq S} \mu(J|S) && \text{By definition} \\
\Rightarrow\quad & \pi(a|\{a,b\}) + \sum_{\substack{a\in J\subseteq\{a,b\} \\ a \text{ is not } \succ\text{-best}}} \mu(J|\{a,b\}) \geq \pi(a|S) + \sum_{\substack{a\in J\subseteq S \\ a \text{ is not } \succ\text{-best}}} \mu(J|S) && \\
\Rightarrow\quad & \sum_{\substack{a\in J\subseteq\{a,b\} \\ a \text{ is not } \succ\text{-best}}} \mu(J|\{a,b\}) \geq \pi(a|S) - \pi(a|\{a,b\}) > 0 && \text{By the hypothesis}
\end{aligned}$$

∎

From the last line, since the consideration set that a is not $\succ$-best has positive probability, we can deduce that there *must be* something that is better than a in the set $\{a,b\}$. Hence, we can conclude that $b \succ a$. Now, the next question is whether we can generalize the set from $\{a,b\}$ to an arbitrary set $T \subseteq S$? The answer is not straightforward since the above argument does not generalize immediately. To see this, consider following two data points we have.

Example 1. Consider the following choice data. In this example, we have $\pi(c|\{a,b,c,d\}) > \pi(c|\{b,c,d\})$.

$\pi(.\|.)$	a	b	c	d
$\{a,b,c,d\}$	0	0.2	0.3	0.5
$\{b,c,d\}$		0.25	0	0.75

By the same operation, we have

$$\sum_{\substack{c\in J\subseteq\{b,c,d\} \\ c \text{ is not } \succ\text{-best}}} \mu(J|\{b,c,d\}) \geq \pi(c|\{a,b,c,d\}) - \pi(c|\{b,c,d\}) = 0.3 > 0$$

What it informs us is that there are something better c in the set $\{b,c,d\}$. However, we could not identify whether this alternative is b or c (might be both) . Hence, the identification is not clear when there are more than two alternatives in the smaller set. Therefore, we need to introduce a new perspective into viewing the choice data, such that more revealed preference can be achieved. In particular, we consider the pair $(\pi, \succ)$. i.e. we pair choice data with a possible preference ordering over alternatives. Let $U_{\succ}(a)$ be the (strict) upper counter set of a according to $\succ$.

Definition 5. $(\pi, \succ)$ satisfies Attention Compensation (AC) if $\sum\limits_{b\in T\cap U_{\succ}(a)} \pi(b|T) \geq \pi(a|S) - \pi(a|T)$ for all $a \in T \subseteq S$.

Although the condition is applied on $(\pi, \succ)$, it is also safe to define the condition over the choice data only. In particular, we say that π satisfies AC if there exist an $\succ$ such that $(\pi, \succ)$ satisfies AC. There are two possible cases for the sign of $\pi(a|S) - \pi(a|T)$. Firstly, notice that if $\pi(a|S) - \pi(a|T) \leq 0$ for all $a \in T \subseteq S$, then the condition is satisfied trivially. Hence, the condition imposes restrictions only when $\pi(a|S) - \pi(a|T) > 0$. One thing to note is that if π satisfies regularity, π always satisfies AC because $\pi(a|S) - \pi(a|T) \leq 0$ for all $a \in T \subseteq S$. Specifically, $(\pi, \succ)$ satisfies AC for any $\succ$.

Given this observation, we note that several important models also satisfy AC such as RUM. Yet, any preference would be able to account for such data in AC. It is interesting to pin down the case that the preference $\succ$ is "pairwise unique" when π satisfies AC. Hence, we define a revealed preference over this property, which is in a similar spirit to the revealed preference above.

Definition 6. Let $\{\succ_j\}_{j=1,\ldots,J}$ be all the preference relation such that $(\pi, \succ_j)$ satisfies Attention Compensation. We say bPa if $b \succ_j a$ for all j.

To see how this definition aids revealed preference, let us assume that $\pi(a|S) - \pi(a|T) > 0$ and $T \subset S$. From the formula, we know that $\pi(U_{\succ}(a)|T) > 0$. Therefore, there must be something that is better than a in the set T. Also, if the smaller set contains only two elements, we can even claim revealed preference from this property.

Corollary 1. If $(\pi, \succ)$ satisfies Attention Compensation (AC) and $\pi(a|S) - \pi(a|T) > 0$, then there exists $b \in T$ such that $b \succ a$. In particular, if $T = \{a, b\}$, then bPa.

One should immediately be able to see the connection between Lemma 1 and the second part of the Corollary 1. Notice that AOM and AC are both able to claim revealed preference when they observe violation of *regularity*. Notice that Lemma 1 does not say anything when the smaller set contains more than two elements while Attention Compensation gives a condition where one can make further exploration. In particular, we show in the following that one can conclude some P for the example in Example 1.

Example 1. ***(continued)*** (Revealed Preference at attention compensating choice data) Note that in this example, we have $\pi(c|\{b, c, d\}) - \pi(c|\{a, b, c, d\}) > 0$. Hence, from Corollary 1, we know that there exists $y \in \{b, c, d\}$ such that $y \succ c$. To look further into the condition for AC, we can see that $\pi(b|\{b, c, d\})$ alone is not able to account for the decrease of the choice probability of c. i.e. $\pi(b|\{b, c, d\}) < \pi(c|\{b, c, d\}) - \pi(c|\{a, b, c, d\})$. Hence, for π to satisfy AC, it must be either $d \succ c \succ b$ or $b \succ d \succ c$ or $d \succ b \succ c$ so that we have $\pi(d|\{b, c, d\}) \geq \pi(c|\{b, c, d\}) - \pi(c|\{a, b, c, d\})$ or $\pi(b|\{b, c, d\}) + \pi(d|\{b, c, d\}) \geq \pi(c|\{b, c, d\}) - \pi(c|\{a, b, c, d\})$. We can then conclude that dPc.

We then state the key characterization theorem that guarantees us that we can use attention compensation for revealed preference purpose when choice rule is a AOM. We provide the only-if part of the proof here, which is similar in spirit to the proof for Lemma 1. The if part of the proof is a bit more involved, which requires the use of Farkas's Lemma for the existence of solution. It is provided in the Appendix.

Theorem 1 (characterization)**.** A choice rule π has a attention overload representation in $\succ$ if and only if $(\pi, \succ)$ satisfies attention compensation.

Proof. For the only-if part. We consider inequalities around $\phi(x|S)$. Note that

$$\phi(a|S) = \pi(a|S) + \sum_{\substack{a \in J \subseteq S \\ a \text{ is not } \succ\text{-best}}} \mu(J|S) \begin{cases} \geq \pi(a|S) \\ \leq \pi(a|S) + \sum_{b \in S \cap U_{\succ}(x)} \pi(b|S) \end{cases}$$

Since $\phi(a|T) \geq \phi(a|S)$, we then have $\pi(a|T) + \sum_{b \in T \cap U_{\succ}(a)} \pi(b|T) \geq \pi(a|S)$. ■

With this characterization theorem, the following corollary is a immediate result. We state it here for completeness.

Corollary 2. Let π be a AOM. Then, b is revealed preferred to a if and only if bPa.

To see the connection of overloaded attention rule to Corollary 1, let assume $\pi(a|S) - \pi(a|T) > 0$. This represents the case that the choice probability *decreases* when the choice set shrinks to T from S. Remembering that for overloaded attention rule, each alternatives get weakly more attention in smaller choice set. Hence, we know it must be case that in smaller choice set, the alternative is considered *more frequently* along with *better* alternatives such that it is chosen less frequently in the smaller choice set.

4 Extentions

4.1 Default option and choice overload

Several existing literature considers the default option, e.g. Manzini and Mariotti (2014), Brady and Rehbeck (2016) and Echenique, Saito, and Tserenjigmid (2018). To provide an accurate comparison to these models, we extend AOM to accommodate an outside option. Let a^* be the default option. In the model with the default option, we will allow an empty set consideration. Hence, now $\mu(.|S)$ is defined over all subsets of S including the empty set. The default option is always available and can be interpret as choosing nothing whenever the consideration set is empty. Let $X^* = X \cup \{a^*\}$ and $S^* = S \cup \{a^*\}$ for all $S \in \mathcal{X}$. We require that the choice rule satisfy $\sum_{a \in S^*} \pi(a|\pi) = 1$ and $\pi(a|S) \geq 0$ for all $a \in S^*$. Thus, $\pi(a^*|S) = \mu(\emptyset|S)$.

There is *a priori* no restriction on $\mu(\emptyset|S)$ from the attention overload model. In fact, for any choice data on default option, as long as the rest of the data satisfies Attention Compensation, the characterization still holds. Formally, we say that a choice rule π has a attention overload representation in $\succ$ with a default option if there exists a overloaded attention rule μ such that for each $a \in S$, $\pi(a|S) = \sum_{T \subseteq S} \mathbb{1}(a \text{ is } \succ\text{-best in } T) \cdot \mu(T|S)$ and $\pi(a^*|S) = \mu(\emptyset|S)$. It is straight-forward to see that the characterization with the property AC is necessary and sufficient.

Remark 1. A choice rule π has a AOM presentation in $\succ$ with a default option if and only if $(\pi, \succ)$ satisfies compensating attention.

The next question is whether the outside options are chosen *more* or *less* often when the decision problem S gets bigger. The Choice Overload phenomenon, e.g. Iyengar and Lepper (2000), suggests that people would tend to choose outside option *more* often if the sets get bigger. As mentioned in the Introduction, existing models of random attention rule, e.g. Manzini and Mariotti (2014), Brady and Rehbeck (2016) and Cattaneo et al. (2019), predict the other way.[4] In stark contrast, the story of our model is compatible with choice overload: while the competition between alternatives get more fierce as the decision problems get bigger, the decision maker tends not to consider any alternative and choose the outside option. Also, by the above characterization result, we know that choice overload can be explained by AOM if the rest of the data satisfies AC.

Lastly, we investigate how choice overload *affects* the explanatory power of AOM in the following scenario. Suppose an econometrican wasn't able to get the default option data but a new technology enables him to have access to such data. He also finds that the outside option data satisfies choice overload. One natural question to ask is that, if the original data satisfies AOM, does the new choice data under normalization still satisfies AOM? The answer is affirmative. The reason is that it is easier to satisfy AC under the new normalization. However, the converse of the statement is not true due exactly to the opposite reason. In other words, the existence of outside option in choice overload *enhances* the explanatory power of AOM. We put this observation in the following corollary. We denote π^* as the re-normalization of π with outside option satisfying choice overload.

Remark 2. If π is AOM, then π^* is AOM with choice overload.

4.2 Attentive at binaries

When an alternative is chosen *frequently enough* in a binary choice set, a policy maker may want to conclude that the alternative is better than the other in the binary choice set. It is up to the choice of policy maker to decide what frequency is sufficient. We first denote this threshold frequency as η.

Definition 7. $(\pi, \succ)$ satisfies η-constrained revealed preference if $a \succ b$ whenever $\pi(a|\{a,b\}) > \eta$.

As discussed before, by considering $(\pi, \succ)$, we are matching a preference ordering to a choice data. The above definition fulfills exactly our needs in revealing the preference for choice data. We can see that the η measures how *cautious* the policy maker is when making welfare judgement. If η is 1, the policy would conclude nothing from the choice data. If η is lower than 0.5, the policy maker may get a cyclic $\succ$ which does not help at all with policy making. Hence, the question is,

[4] i.e. the outside option is chosen more often in the smaller set. For Manzini and Mariotti (2014) and Brady and Rehbeck (2016), the restriction lies in formation of attention rule. For Cattaneo et al. (2019), the restriction comes from the monotonic attention rule.

given choice data, how do we know we can safely impose this definition of revealed preference? The answer is straightforward. If we put a sufficient restriction on the choice generating process, i.e. the choice rule that is generated by attention rule, we can use the above definition and make the claim of revealed preference.

Assumption 2 (η-attentive at Binaries)**.** For any $a, b \in S$, $\eta \geq \max\{\mu(\{a\}|\{a,b\}), \mu(\{b\}|\{a,b\})\}$

The above assumption fulfills our need. The condition is simple and intuitive. Consider η is 0.4. Given that the singleton consideration sets are bounded above by 0.4, if we observe a is chosen more frequently than 0.4, we know that there must be some occasions where about a and b are considered together, i.e. $\mu(\{a,b\}|\{a,b\}) > 0$, and the consumer prefers a to b. Our assumption is weaker than the assumption proposed by Cattaneo et al. (2019), meaning that any attention rule that satisfies their assumption would automatically satisfies our assumption. We then present the joint characterization results. The result guarantees us two things. Firstly, the definition 7 indeed gives us revealed preference by making the assumption. Secondly, given choice data, we know "when" these data can be represented by the AOM with the assumption of η-attentive at Binaries.

Theorem 2 (characterization)**.** π has a attention overload representation in $\succ$, where the attention satisfies assumption 1 and 2, if and only if $(\pi, \succ)$ satisfies attention compensation and η-constrained revealed preference.

5 Related Literature

5.1 Attention Rule

The *attention overload model* (AOM) is a missing piece of the puzzle in the limited consideration models. We can see this by its close connection to the *random attention model* (RAM) proposed by Cattaneo et al. (2019). While RAM generalises the *attention filter* in Masatlioglu, Nakajima, and Ozbay (2012), AOM generalizes the *competition filter* in Lleras et al. (2017). Note that AOM and RAM are independent. Attention Overload assumption is orthogonal to Monotonic Attention assumption of Cattaneo et al. (2019). We provide two attention rules to highlight their differences. The first one satisfies Monotonic Attention but not Attention Overload. According this attention rule, DM consider everything in a larger set but she only considers singleton consideration sets for smaller sets.

$\mu(T\|S)$	$\{a,b,c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\{a\}$	$\{b\}$	$\{c\}$
$\{a,b,c\}$	1	0	0	0	0	0	0
$\{a,b\}$		0			1/2	1/2	
$\{a,c\}$			0		1/2		1/2
$\{b,c\}$				0		1/2	1/2

The second one satisfies Attention Overload but not Monotonic Attention. This attention rule highlights the idea of "less is more". DM cannot deal with larger choice sets, hence she only consider singleton consideration sets when she faces larger sets but she considers everything in smaller sets.

$\mu(T\|S)$	$\{a,b,c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\{a\}$	$\{b\}$	$\{c\}$
$\{a,b,c\}$	0	0	0	0	1/3	1/3	1/3
$\{a,b\}$		1			0	0	
$\{a,c\}$			1		0		0
$\{b,c\}$				1		0	0

These two attention rule make it clear the distinction between two assumption on attention rule. However, since the attention rule is not observable, one might wonder whether these two models are behaviorally different. The example below illustrates that there are some choice data which have an AOM representation but not RAM representation, and also the other way around.

Example 2. (Explanatory Power: AOM vs. RAM)

$\pi_1(.\|.)$	a	b	c
$\{a,b,c\}$	0.4	0.3	0.3
$\{a,b\}$	0.8	0.2	
$\{a,c\}$	0.8		0.2

In RAM, revealed preference says that $b \succ c$ and $c \succ b$. Hence, it could not be explained by RAM. In AOM, revealed preference says $a \succ c$ and $a \succ b$. Interestingly, there does not exists an example of three alternatives where RAM can explain but AOM cannot.[5] We need to go for four alternatives cases to achieve this.

[5]In other words, AOM explains *more* choice data than RAM in 3-alternative cases.

$\pi_2(.\|.)$	a	b	c	d
$\{a,b,c,d\}$	1/2	1/2	0	0
$\{a,b,c\}$	0	2/3	1/3	
$\{a,b\}$	1/2	1/2		

In AOM, revealed preference says that $a \succ b$ and $b \succ a$. Hence, it could not be explained by AOM. In RAM, revealed preference says $a \succ d$ and $b \succ c$.

In deterministic environment, both Competition filter and Attention filter give revealed preference by considering taking away an element from the choice set, but the intuition behind is distinctively different. When we observe choice reversal of alternative, attention filter says that the item is *better* than another item in the *bigger* choice set, while competition filter says that the item is *worse* than another item in the *smaller* choice set. Note that these two ideas naturally extend itself into the probabilistic environments in RAM and AOM: When one observes a violation of regularity of an alternative, RAM says the alternatives is *better* than another alternative in the bigger choice set, and AOM says the alternative is *worse* than another alternative in the smaller choice set. This is illustrated in Example above.

In introduction, we mentioned Manzini and Mariotti (2014) as a special case of the model. Manzini and Mariotti (2014) assumes that each alternative a has a fixed probability, $\gamma(a)$, to be considered. It is equivalent to say that the attention frequency, $\phi_{MM}(a|S)$, in their world is held fixed across different decision problem S. i.e.

$$\phi_{MM}(a|S) = \gamma(a) \text{ for all } S$$

whenever $a \in S$. Since attention overload requires only weak inequality, the model falls into AOM. Secondly, Aguiar (2017) is also a special case of the model. Aguiar (2017) assumes that each category D has a fixed probability $m(D)$. If the category is *available*, the DM picks the best alternative out of it. If not, the DM chooses the default option. Let the set of all category be $\mathcal{D}$. Therefore, the attention frequency is given by, whenever $a \in S$,

$$\phi_{Aguiar}(a|S) = \sum_{a \in D \in \mathcal{D}} m(D)$$

Notice that the attention frequency is also held fixed in the model. Hence, it immediately satisfied attention overload.

Lastly, we put our attention on random competition filter (RCF), which is a major special case of the model. Let $\Gamma_i(.)$ be consideration set mapping which satisfies competition filter and $\sum_{i=1}^{n} \alpha_i = 1$. A random competition filter model is specified by the following attention rule with the respective attention frequency,

$$\mu_{RCF}(T|S) = \sum_{i=1}^{n} \alpha_i \mathbb{1}(\Gamma_i(S) = T)$$
$$\phi_{RCF}(a|S) = \sum_{i:a\in\Gamma_i(S)} \alpha_i$$

One can immediately see that random competition filter satisfies attention overload, since $a \in \Gamma_i(T)$ for all $T \subseteq S$ if $a \in \Gamma_i(S)$. Random attention filter nests two others model in the introduction, which are Bounded Rationalization and Imprecise Narrowing Down. Bounded Rationalization is a straightforward and meaningful generalization of Cherepanov et al. (2013). It states that the DM does not always stick to the same set of rationale given the same choice set. Hence, it is *as if* the DM assigns a probability distribution over the power set on the set of rationale. Since Cherepanov et al. (2013) is a special case of Lleras et al. (2017), it immediately follows that Bounded Rationalization model is a special case of random competition filter. Imprecise Narrowing Down shares similar idea. Given the same choice set, the DM does not necessarily follow the same procedure on setting up criteria. Thus, it is *as if* the DM assigns a probability distribution over the set of all possible procedure. It makes Imprecise Narrow Down again a special case of random competition filter.

5.2 Regularity

In the following, we consider models which do not have a reference to attention rule. Notice that a number of models in this aspect respect regularity. The seminal work of the Random Utility Model (RUM) is one of those. By previous discussion, the condition AC is automatically satisfied when models satisfy regularity. Hence, any RUM is AOM. Note that there are a number of models are included in RUM. For example, Gul, Natenzon, and Pesendorfer (2014) considers attribute rule in which the DM first draw an attribute and then pick an alternative which contains such attribute. They show that every attribute rule is a RUM. Hence, every attribute rule is a AOM. On the other hand, Fudenberg, Iijima, and Strzalecki (2015) introduces the additive perturbed utility model where the decision intentionally randomized as deterministic choices can be costly. Since the choices in their model always satisfies regularity, any choice rule in APU has a representation of AOM.

5.3 Others

There are several other stochastic choice model which are compatible of producing violation of regularity. Intriguingly, we can show that some of them are AOM by directly checking the condition AC. Echenique et al. (2018) considers priorities in alternatives *before* the DM applies the Luce rule, which is called the Perception-adjusted Luce model (PALM). In the model, DM impose a weak order $\succeq$ over alternatives as priority and attach Luce weight $u(x)$ to each of them. To explain their model, We take their primitive $\succeq$ as our preference $\succ$ and consider an arbitrary tie-breaking rule. We put the observation in the following.

Proposition 1. *Any PALM satisfies AC.*

Echenique and Saito (2018) Proposes a model, General Luce Model (GLM), where a deterministic consideration set mapping is applied before the DM use Luce rule over alternatives. Notice GLM reduces to the standard Luce model when every alternative is chosen with positive probability. We can construct example where every alternative has positive probability but does not satisfy Luce rule. Hence, GLM does not includes AOM. On the other hand, AOM does not include GLM because the restriction-free consideration set mapping in GLM allows for cyclic P in our model. However, the Threshold GLM, a special case of GLM, is included in AOM. In threshold GLM, alternatives with too low a Luce weight, u(x), would not be considered. We take their primitive $u(x)$ and construct $a \succ b$ if $u(a) < u(b)$, and take an arbitrary tie-breaking rule if $u(a) = u(b)$.

Proposition 2. *Any threshold GLM satisfies AC.*

6 Conclusions

As there are more products while time is limited, it is likely that consumer's attention span on each product decreases due to competition, which we call "attention overload". In this paper, we develop the notion of attention frequency, and propose a model, *Attention Overload Model*, to capture attention overload. We show that the condition, attention compensation, is key to checking whether the data is consistent with the model. We also show that several existing models fall under this model, but at the same time, the richness of the model allows us to explain more different phenomena such as *Choice Overload*. We show how policy marker can draw inference over revealed preference by purely observing choice data. A more stringent requirement, attention at binaries, is proposed to provide more information to the policy marker for welfare judgement. The research on *AOM* opens up a path intelligent enquiry into consumer's attention under competition. For example, one may ask, what would be the stylistic parametric model of attention overload given existing parametric model of attention fall short of capturing attention overload? A useful parametric model can definitely further benefit the level of grip over consumer's behavior from the policy-making point of view. On the other hand, along the line of non-parametric restriction, one may be interested to find out what is the intersection between the RAM from Cattaneo et al. (2019) and AOM. Further researches in this agenda are encouraged.

References

Aguiar, V. H. (2015). Stochastic choice and attention capacities: Inferring preferences from psychological biases. *SSRN Working Paper 2607602.*

Aguiar, V. H. (2017). Random categorization and bounded rationality. *Economics Letters*, *159*, 46–52.

Başar, G. & Bhat, C. (2004). A parameterized consideration set model for airport choice: An application to the San Francisco Bay Area. *Transportation Research Part B: Methodological*, *38*(10), 889–904.

Brady, R. L. & Rehbeck, J. (2016). Menu-dependent stochastic feasibility. *Econometrica*, *84*(3), 1203–1223.

Caplin, A., Dean, M., & Leahy, J. (2019). Rational Inattention, Optimal Consideration Sets, and Stochastic Choice. *The Review of Economic Studies*, *86*(3), 1061–1094.

Cattaneo, M., Ma, X., Masatlioglu, Y., & Suleymanov, E. (2019). A Random Attention Model. *Journal of Political Economy.*

Cherepanov, V., Feddersen, T., & Sandroni, A. (2013). Rationalization. *Theoretical Economics*, *8*(3), 775–800.

Demuynck, T. & Seel, C. (2018). Revealed preference with limited consideration. *American Economic Journal: Microeconomics*, *10*(1), 102–131.

Echenique, F. & Saito, K. (2018). General luce model. *Economic Theory*, 1–16.

Echenique, F., Saito, K., & Tserenjigmid, G. (2018). The perception-adjusted luce model. *Mathematical Social Sciences*, *93*, 67–76.

Fudenberg, D., Iijima, R., & Strzalecki, T. (2015). Stochastic choice and revealed perturbed utility. *Econometrica*, *83*(6), 2371–2409.

Geng, S. (2016). Decision time, consideration time, and status quo bias. *Economic Inquiry*, *54*(1), 433–449.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica*, *76*(5), 1017–1074.

Gul, F., Natenzon, P., & Pesendorfer, W. (2014). Random choice as behavioral optimization. *Econometrica*, *82*(5), 1873–1912.

Horan, S. (2018). Random consideration and choice: A case study of "default" options. *Working Paper, Université de Montréal and CIREQ.*

Huberman, G. & Regev, T. (2001). Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *Journal of Finance*, *56*(1), 387–396.

Iyengar, S. S. & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*(6), 995–1006.

Laroche, M., Rosenblatt, J., & Sinclair, I. (1984). Brand Categorization Strategies in an Extensive Problem Solving Situation: a Study of University Choice. *ACR North American Advances*, *NA-11*.

Lleras, J. S., Masatlioglu, Y., Nakajima, D., & Ozbay, E. Y. (2017). When more is less: Limited consideration. *Journal of Economic Theory*, *170*, 70–85.

Manzini, P. & Mariotti, M. (2012). Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association*, *10*(5), 1141–1165.

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. *American Economic Review*, *102*(5), 2183–2205.

Reutskaja, E. & Hogarth, R. M. (2009). Satisfaction in choice as a function of the number of alternatives: When "goods satiate". *Psychology and Marketing*, *26*(3), 197–203.

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, *101*(2), 900–926.

Rosen, D. E., Curran, J. M., & Greenlee, T. B. (1998). College choice in a brand elimination framework: The high school students perspective. *Journal of Marketing for Higher Education*, *8*(3), 73–92.

Sheridan, J. E., Richards, M. D., & Slocum, J. W. (1975). Comparative analysis of expectancy and heuristic models of decision behavior. *Journal of Applied Psychology*, *60*(3), 361–368.

Visschers, V. H., Hess, R., & Siegrist, M. (2010). Health motivation and product design determine consumers visual attention to nutrition information on food products. *Public Health Nutrition*, *13*(7), 1099–1106.

Appendix

Proof for the if-part in Theorem 1

Proof. The idea of the proof: The proof is mainly divided into two parts idea-wise. The first part sets up the system of linear equation which pins down the μ that satisfies the desired property. Some algebraic operations are devoted into lining up the system in a way to prepare for the second part. The second part shows how we can utilize the Farkas's Lemma for proving the existence of a solution to the system for any parameter value which satisfies the property AC. Claim 4 concludes the proof.

Assume $(\pi(.), \succ)$ satisfies property AC. For every S and $x \in S$, we set $\sum_{\substack{x \in J \subseteq S \\ x \text{ is } \succ\text{-best}}} \mu(J|S) = \pi(x|S)$ and $\phi(x|S) = \max_{R \supseteq S} \pi(x|R)$. It is immediate to see that the attention rule satisfies the desired non-parametric properties. What remains is to show that there exists a solution to the system of linear equation. Let $x_1 \succ x_2 \succ ...x_n$. Then, we have for $i = 1, ...n$

$$\sum_{\substack{x_i \in J \subseteq S \\ x_i \text{ is } \succ\text{-best}}} \mu(J|S) = \pi(x_i|S) \qquad \text{.....(denoted by } \mathcal{P}_i\text{)}$$

$$\phi(x_i|S) = \max_{R \supseteq S} \pi(x_i|R) \qquad \text{.....(denoted by } \mathcal{M}_i\text{)}$$

Note that for x_1, $\max_{R \supseteq S} \pi(x_i|R) = \pi(x_i|S)$ (or it violates property AC). Also, $\sum_{\substack{x_i \in J \subseteq S \\ x_i \text{ is } \succ\text{-best}}} \mu(J|S) = \phi(x_i|S)$. Hence, $(\mathcal{P}_1) = (\mathcal{M}_1)$. On the other hand, $\mathcal{P}_n$ is $\pi(x_n|S) = \mu(\{x_n\}|S)$, which immediate gives the solution to the "unknown" $\mu(\{x_n\}|S)$. Hence, we are left with $(\mathcal{P}_i)$, $i = 1, ..n-1$ and $(\mathcal{M}_i)$, $i = 2, ..n$. Then, we create $\mathcal{M}'_i \equiv \sum_{j \leq i} (\mathcal{P}_j) - (\mathcal{M}_i)$ for every $i = 2, ...n$, i.e.

$$\sum_{j<i} \sum_{\substack{x_i \notin J \subseteq S \\ x_j \text{ is } \succ\text{-best}}} \mu(J|S) = \sum_{j \leq i} \pi(x_j|S) - \max_{R \supseteq S} \pi(x_i|R) \qquad \text{.....(denoted by } \mathcal{M}'_i\text{)}$$

where $\sum_{j \leq i} \pi(x_j|S) - \max_{R \supseteq S} \pi(x_i|R) \geq 0$ for $i = 2, ...n$ by Property K. Lastly, we create $(\mathcal{P}'_1) \equiv (\mathcal{P}_1) - \sum_{j>1} (\mathcal{M}_j)$. Hence, we are left with $(\mathcal{P}'_1), (\mathcal{P}_i), i = 2, ..n-1$ and $(\mathcal{M}'_1), i = 2, ...n$. We utilize the Farkas's Lemma to prove the existence of solution to the above system of linear equations. Note that the system is straightforward when $n < 3$. Hence, we focus only the case that $n \geq 3$.

Farkas's Lemma : Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then exactly one of the following is true:

1. There exists an $x \in \mathbb{R}^n$ such that $Ax = b$ and $x \geq 0$.
2. There exists a $y \in \mathbb{R}^m$ such that $yA \geq 0$ and $yb < 0$.

We let A be the matrix and b be the vector which represents the above system of linear equations by $A\mu = b$. i.e. $A := \left(r_1, r_2, ..., r_{2n-2}\right)^T$, and $b := (b_1, b_2, ...b_{2n-2})^T$, where r_j's are row vector. In particular, we let r_1 and b_1 correspond to the LHS and RHS of $\mathcal{P}'_1$ respectively; r_j and b_j correspond to the LHS and RHS of $\mathcal{M}'_{n+2-j}$ respectively for $j = 2, ..., n$; r_j and b_j correspond to the LHS and RHS of $\mathcal{P}_{-n+1+j}$ respectively for $j = n+1, ..., 2n-2$.

For shorthand, we write, for all i, whenever the RHS is defined, $m_i \equiv \max_{R \supseteq S} \pi(x_i|R)$, $\pi_i \equiv \pi(x_i|S)$ and $k_i \equiv \sum_{j \leq i} \pi(x_j|S) - \max_{R \supseteq S} \pi(x_i|R) = \sum_{j \leq i} \pi_j - m_i$. We let the set B as the set of b which is generated by a data set that satisfies AC . i.e.

$$\begin{aligned} B = \Big\{ b \in \mathbb{R}^{2n-2} : & b_1 = \pi_1 - k_n - k_{n-1} - .... - k_2, \\ & b_i = k_{n-i+2}, \text{ for } i = 2, .., n, \\ & b_i = \pi_{i+1-n} \text{ for } i = n+1, ..., 2n-2, \\ & \text{where } \pi(.|S) \text{ satisfies AC.} \Big\} \end{aligned}$$

We would show that there does not exist $y = (y_1, y_2, y_3, ...y_{2n-2}) \in \mathbb{R}^{2n-2}$ such that $yA \geq 0$ and $yb < 0$ for all $b \in B$. We define the set $Y(A)$ as the set of y which satisfies $yA \geq 0$. Hence, it suffices to show that for all $b \in B$, $\min_{y \in Y(A)} yb \geq 0$. Note that except b_1, all b_j are positive for all possible $\pi(.|S)$ that satisfies AC. Hence, the key insight in the following proof is to show that how we can guarantee $yb \geq 0$ despite the possibility of b_1 being negative.

Claim 1. *For all $y \in Y(A)$, $y \geq 0$.*

Proof. It is due to the fact that A admits a reduced row-echelon form by construction. To see this, first note that $yA \geq 0$. Since A admits a reduced row-echelon form, the leading entry is 1 and the leading entry in each row is the only non-zero entry in its column. It gives us $y_j \geq 0$ for all j. ■

With claim 1, we can see that if $b_1 \geq 0$, the proof is trivially done. We then state claim 2, which is a special case of claim 3.

Claim 2. *For all $y \in Y(A)$ and for $i = 2, ...n-1$, we have $y_i + y_j \geq y_1$ for $j = i+1, ..., 2n-i$.*

Let $\mathbb{P}_n$ be the power set of the set $\{2, 3, ...n-1\}$

Claim 3. *(A generalization of claim 2) For all $y \in Y(A)$ and for all $P \in \mathbb{P}_n$, $\sum_{i\in P} y_i + y_j \geq |P| * y_1$, for $j = \max_{i\in P} i + 1, \max_{i\in P} i + 2, ..., 2n - \max_{i\in P} i$.*

Proof. For any set P, we get

$$\sum_{i\in P} y_i + y_j \geq |P| * y_1 \text{ from the column of } \mu(S - \cup_{i\in P\cup\{j\}}\{x_{i+n-2}\}|S)$$

for any $j \in \{\max_{i\in P} i + 1, i + 2, ...n\}$. For the LHS: it is because for any $i \in P$, the row vector r_{n-i+2} has the coefficient of 1 in the column of $\mu(S - \cup_{i\in P\cup\{j\}}\{x_{i+n-2}\}|S)$ by construction. For the RHS: it is because the row vector r_1 has the coefficient of $|P|$ in the same column by construction. Also, we get

$$\sum_{i\in P} y_i + y_j \geq |P| * y_1 \text{ from the column of } \mu(S - \cup_{i\in P}\{x_{i+n-2}\} - \cup_{i<j-n}\{x_i\}|S)$$

for any $j \in \{n + 1, n + 2, ...2n - \max_{i\in P} i\}$. For the LHS: it is because for any $i \in P$, the row vector r_{n-i+2} has the coefficient of 1 in the column of $\mu(S - \cup_{i\in P}\{x_{i+n-2}\} - \cup_{i<j-n}\{x_i\}|S)$ by construction. For the RHS: it is because the row vector r_1 has the coefficient of $|P|$ in the same column by construction. Hence, we have covered any j in $\{\max_{i\in P} i + 1, \max_{i\in P} i + 2, ..., 2n - \max_{i\in P}\}$. The proof is complete. ■

We need to show an auxiliary minimization problem to complete the proof. Let $\mathbf{c}_n$ and $\mathbf{z}_n$ be two vectors. To be consistent with the above in notation, both vectors start with subscript 2 and end with $2n - 2$. i.e. $\mathbf{c}_n = (c_2, c_3, ...c_{2n-2})$

Claim 4. *For all $n \geq 3$,*

$$\min_{\mathbf{c}_n\in\mathbf{C}_n,\mathbf{z}_n\in\mathbf{Z}_n} \mathbf{c}_n \cdot \mathbf{z}_n \geq 1$$

where

$$\mathbf{C}_n = \{\mathbf{c}_n \in \mathbb{R}_+^{2n-3}| \sum_{i=2}^{n+1-j} c_i + \sum_{i=n}^{n-3+j} (c_{i+1} + c_{i+3-j}) \geq 1, j = 2, 3, ...n\}$$

$$\mathbf{C}_n = \{\mathbf{c}_n| \sum_{i=2}^{n+1-j} c_i + \sum_{i=n+3-j}^{n} c_i + \sum_{i=n+1}^{n-2+j} c_i \geq 1, j = 2, 3, ...n\}$$

$$\mathbf{Z}_n = \{\mathbf{z}_n \in \mathbb{R}_+^{2n-3}|\sum_{i\in P} z_i + z_j \geq |P|, \forall P \in \mathbb{P}_n, j = \max_{i\in P} i + 1, \max_{i\in P} i + 2, ..., 2n - \max_{i\in P} i\}$$

Proof. We prove by induction. Consider n=3. We have

$$\mathbf{C}_3 = \{\mathbf{c}_3 | c_2 \geq 1, c_4 + c_3 \geq 1\}$$

$$\mathbf{Z}_3 = \{\mathbf{z}_3 | z_2 + z_3 \geq 1, z_2 + z_4 \geq 1\}$$

It is straight-forward to see that minimum must be attained by binding constraint $c_4 + c_3 = 1$.

$$\begin{aligned}
\mathbf{c}_3 \cdot \mathbf{z}_3 &= z_2 c_2 + z_3 c_3 + z_4 c_4 \\
&\geq z_2 + c_3(1 - z_2) + c_4(1 - z_2) \\
&= z_2 + (c_3 + c_4) * (1 - z_2) = 1
\end{aligned}$$

Suppose $n = k - 1$ is true. Consider $n = k$. We set up the Lagrangian minimization problem and assigns Lagrangian multiplier $\lambda_i's$ to the constraints in $\mathbf{C}_n$. For notational convenience, we adopt a slightly different notational convention for the λ. We give each multiplier the subscript as a set of all the subscript involved in the constraint. Take $n = 3$ as an example, we would have multiplier λ_2 for $c_2 \geq 1$ and $\lambda_{3,4}$ for $c_4 + c_3 \geq 1$. It is simple to check that each constraint has it own unique respect set of subscript. One advantage of using this subscript convention is that it is more informative than giving natural number to the constraints. We collect all possible subscript of λ and name it Λ_k. The Lagrangian multiplier for the constraints in $\mathbf{Z}_n$ is not impactful in the proof.

First order condition of the Lagrangian equation gives:

$$\frac{\partial L}{\partial c_i} = z_i - \sum_{i \in S \in \Lambda_k} \lambda_S \geq 0; \qquad (z_i - \sum_{i \in S \in \Lambda_k} \lambda_S) c_i = 0, i = 2, 3, ...2k - 2$$

By plugging in first order condition, we can get the following,

$$\begin{aligned}
\mathbf{c}_n \cdot \mathbf{z}_n &\geq \sum_{i=2}^{2k-2} c_i (\sum_{i \in S \in \Lambda_k} \lambda_S) \\
&= \sum_{S \in \Lambda_k} \lambda_S \sum_{i \in S} c_i \\
&\geq \sum_{S \in \Lambda_k} \lambda_S
\end{aligned}$$

Note that if all $c_i \neq 0$ for all i, it is straight-forward to see that $\sum_{S \in \Lambda_k} \lambda \geq 1$. For example, if $c_{2k-2} \neq 0$ and $c_2 \neq 0$, we can get from $\mathbb{Z}_k$,

$$z_2 + z_{2n-2} \geq 1$$

$$\sum_{2\in S\in\Lambda_k}\lambda_S+\sum_{2n-2\in S\in\Lambda_k}\lambda_S\geq 1$$
$$\sum_{S\in\Lambda_k}\lambda_S\geq 1$$

In fact, it is straight-forward to check that as long as

$$(c_2\neq 0 \text{ and } c_{2k-2}\neq 0) \text{ or}$$
$$(c_2\neq 0, c_3\neq 0 \text{ and } c_{2k-3}\neq 0) \text{ or}$$
$$....$$
$$(c_2\neq 0, c_3\neq 0, ...c_{k-1}\neq 0 \text{ and } c_{k+1}\neq 0)$$
$$(c_2\neq 0, c_3\neq 0, ...c_{k-1}\neq 0 \text{ and } c_k\neq 0)$$

then $\sum_{S\in\Lambda}\lambda_S\geq 1$. For cases outside the above, we check sequentially and apply mathematical induction in each cases:

Case 1: $c_2=0$. By re-numbering some of the variables, in particular, write $z'_i=z_{i+1}$ and $c'_i=c_{i+1}$ for $i=2,3,..2(k-1)-2$. We name this set of constraint as $\mathbb{C}_k$ where both c'_i and c_j for some i,j co-exist. We perform the same procedure on and $\mathbb{Z}_k$. Then, by restricting attention only at c'_i and z'_i, it is straightforward to see that $\mathbb{C}_k\subset\mathbf{C'}_{k-1}$ and $\mathbb{Z}_k\subset\mathbf{Z}'_{k-1}$, where $\mathbf{C}'_{k-1}$ is the same set as $\mathbf{C}_{k-1}$ by just renaming c to c'. Hence, in this case, by induction hypothesis,

$$\min_{\mathbf{c}_k\in\mathbb{C}_k,\mathbf{z}_k\in\mathbb{Z}_k}\mathbf{c}_k\cdot\mathbf{z}_k\geq\min_{\mathbf{c}_{k-1}\in\mathbf{C}'_{k-1},\mathbf{z}_{k-1}\in\mathbf{Z}'_{k-1}}\mathbf{c}_{k-1}\cdot\mathbf{z}_{k-1}\geq 1$$

Case 2: $c_3=0$ and $c_{2k-2}=0$. We re-number the variable, in particular, write $c'_i=c_i$ for $i=2$, write $c'_i=c_{i+1}$ for $i=3,...,2(k-1)-3$ and write $c'_i=c_{i+2}$ for $i=2(k-1)-2$. Analogously, we do the same for z. By a similar argument. We show $\min_{\mathbf{c}_k\in\mathbb{C}_k,\mathbf{z}_k\in\mathbb{Z}_k}\mathbf{c}_k\cdot\mathbf{z}_k\geq 1$.

....

Case $k-3$: $c_{k-1}=0$ and $c_{k+2}=...=c_{2k-3}=c_{2k-2}=0$. Write $c'_i=c_i$ for $i=2,..k-3$, write $c'_i=c_{i+1}$ for $i=k-3,...,k$, write $c'_i=i+2$ for $i=k+1,..,2(k-1)-2$.

Last Case: $c_k=c_{k+1}=....=c_{2k-2}=0$. This case needs its special attention. We need to prove a auxiliary claim to finish this proof.

Claim 5. *For all $n\geq 4$*

$$\min_{\mathbf{d}_n\in\mathbf{D}_n,\mathbf{w}_n\in\mathbf{W}_n}\mathbf{d}_n\cdot\mathbf{w}_n\geq 1$$

where

$$\mathbf{D}_n = \{\mathbf{d}_n \in \mathbb{R}_+^{n-2} | \sum_{i=2}^{n+1-j} d_i + \sum_{i=n}^{n-4+j} d_{i+3-j} \geq 1, j = 3, ...n\}$$

$$\mathbf{W}_n = \{\mathbf{w}_n \in \mathbb{R}_+^{n-2} | \sum_{i \in P} w_i + w_j \geq |P|, \forall P \in \mathbb{P}_n, j = \max_{i \in P} i + 1, ..., n\}$$

Proof. For $n = 4$. We have

$$\mathbf{D}_4 = \{\mathbf{d}_4 | d_2 \geq 1, d_3 \geq 1\}$$

$$\mathbf{W}_4 = \{\mathbf{w}_4 | w_2 + w_3 \geq 1\}$$

Hence, we have

$$\begin{aligned}\mathbf{d_4} \cdot \mathbf{w_4} &= d_2 w_2 + d_3 w_3 \\ &\geq w_2 + w_3 \\ &\geq 1\end{aligned}$$

Suppose $n = k - 1$ is true. Consider $n = k$. We apply the same technique for naming the Lagrangian multiplier. We assign Lagrangian multiplier $\omega_i' s$ to the constraint in $\mathbf{D}_k$, and collect the subscript of those multiplier in the set Ω_k. The first order condition of the Lagrangian equation gives:

$$\frac{\partial L}{\partial d_i} = w_i - \sum_{i \in S \in \Omega_k} \omega_S \geq 0; \qquad (w_i - \sum_{i \in S \in \Omega_k} \omega_S) d_i = 0, i = 2, 3, ..., k-2$$

By plugging in the first order condition, we can get

$$\begin{aligned}\mathbf{d}_n \cdot \mathbf{w}_n &\geq \sum_{i=2}^{k-2} (d_i \sum_{i \in S \in \Omega_k} \omega_k) \\ &= \sum_{S \in \Omega_k} (\omega_S \sum_{i \in S} d_i) \\ &\geq \sum_{S \in \Omega_k} \omega_S\end{aligned}$$

Note that if all $d_i \neq 0$ for all i, it is straight-forward to see that $\sum_{S \in \Omega_k} \omega_S \geq 1$. Since, by the constraint in $\mathbf{W}_k$, we have

$$\sum_{i=2}^{k-1} z_i \geq k - 3$$

$$(k-3)\sum_{S\in\Omega_k}\omega_S \geq k-3$$
$$\sum_{S\in\Omega_k}\omega_S \geq 1$$

If any of the $d_i = 0$, the problem reduces to the minimization problem for $k-1$. Hence, by mathematical induction, we proved claim 5. ■

Hence, in the last case, we can apply claim 5. Hence, we finish the proof of claim 4. ■

Recall the previous problem that if $b_1 \geq 0$, the proof is trivially done. If not, i.e. $b_1 < 0$ then we can apply claim 4 by setting

$$c_i = -\frac{b_i}{b_1} \text{ for } i = 2, ..., 2n-2$$
$$z_i = \frac{y_i}{y_1} \text{ for } i = 2, ..., 2n-2$$

Hence, the statement that all $b \in B$, $\min_{y\in Y(A)} yb \geq 0$ is equivalent to the statement that $\min_{\mathbf{c}_n\in\mathbf{C}_n,\mathbf{z}_n\in\mathbf{Z}_n} \mathbf{c}_n \cdot \mathbf{z}_n \geq 1$. It completes the proof. ■

Proof for Theorem 2

Proof. The only-if part is immediate. For the if-part, we need to be concerned about constructing the μ. For non-binaries choice set, we follow the technique in the proof for Theorem 1. For binaries choice set, WLOG, we first let $x \succ y$ throughout the proof. We assume

$$\mu(\{x,y\}|\{x,y\}) = \pi(x|\{x,y\})$$
$$\mu(\{y\}|\{x,y\}) = \pi(y|\{x,y\})$$

To check that it fulfills assumption 2. Suppose not. i.e. there exists y s.t. $\eta < \mu(y|\{x,y\}$. Hence, by definition 7, we know that $y \succ x$, which violates the fact that $x \succ y$.

To check that it fulfills assumption 1. Note that $\phi(x|\{x,y\}) = \pi(x|\{x,y\}) \geq \pi(x|S)$ for all S (or there would be a contradiction that $y \succ x$). Hence, $\phi(x|\{x,y\}) \geq \phi(x|S)$ for all S since $\phi(x|S) = \max_{R\supseteq S}\pi(x|R)$. On the other hand, $\phi(y|\{x,y\}) = 1$, which automatically satisfies assumption 1.

■

Proof for Proposition 1

It suffices to show the following claim.

Claim 6. *For any $n \in \mathbb{N}$ and $M, u_1, u_2, ..., u_n \in \mathbb{R}_+$, we have, for $k = 2.., n$,*

$$\frac{u_1}{\sum_{i=1}^n u_i} + \sum_{s=1}^{k-1} \prod_{j=1}^{s} (1 - \frac{u_j}{\sum_{i=1}^n u_i}) \frac{u_{s+1}}{\sum_{i=1}^n u_i} \geq \prod_{j=1}^{k-1} (1 - \frac{u_j}{\sum_{i=1}^n u_i + M}) \frac{u_k}{\sum_{i=1}^n u_i + M}$$

Proof. Fix n, M, we prove by induction. Let $k = 2$, we have

$$\begin{aligned} LHS - RHS &= \frac{u_1}{\sum_{i=1}^n u_i} + (1 - \frac{u_1}{\sum_{i=1}^n u_i}) \frac{u_2}{\sum_{i=1}^n u_i} - (1 - \frac{u_1}{\sum_{i=1}^n u_i + M}) \frac{u_2}{\sum_{i=1}^n u_i + M} \\ &\geq \frac{u_1}{\sum_{i=1}^n u_i} + (1 - \frac{u_1}{\sum_{i=1}^n u_i}) \frac{u_2}{\sum_{i=1}^n u_i} - (1 - \frac{u_1}{\sum_{i=1}^n u_i + M}) \frac{u_2}{\sum_{i=1}^n u_i} \\ &\geq \frac{u_1}{\sum_{i=1}^n u_i} - \frac{u_1}{\sum_{i=1}^n u_i} \frac{u_2}{\sum_{i=1}^n u_i} \frac{M}{\sum_{i=1}^n u_i + M} \geq 0 \end{aligned}$$

Let the statement be true for $k - 1$.

$$\begin{aligned} LHS - RHS = &\frac{u_1}{\sum_{i=1}^n u_i} + \sum_{s=1}^{k-1} \prod_{j=1}^{s} (1 - \frac{u_j}{\sum_{i=1}^n u_i}) \frac{u_{s+1}}{\sum_{i=1}^n u_i} - \prod_{j=1}^{k-1} (1 - \frac{u_j}{\sum_{i=1}^n u_i + M}) \frac{u_k}{\sum_{i=1}^n u_i + M} \\ \geq &\frac{u_1}{\sum_{i=1}^n u_i} + (1 - \frac{u_1}{\sum_{i=1}^n u_i}) \prod_{j=2}^{k-1} (1 - \frac{u_j}{\sum_{i=1}^n u_i + M}) \frac{u_k}{\sum_{i=1}^n u_i + M} - \prod_{j=1}^{k-1} (1 - \frac{u_j}{\sum_{i=1}^n u_i + M}) \frac{u_k}{\sum_{i=1}^n u_i + M} \\ \geq &\frac{u_1}{\sum_{i=1}^n u_i} - \frac{u_1}{\sum_{i=1}^n u_i} \frac{M}{\sum_{i=1}^n u_i + M} \prod_{j=2}^{k-1} (1 - \frac{u_j}{\sum_{i=1}^n u_i + M}) \frac{u_k}{\sum_{i=1}^n u_i + M} \geq 0 \end{aligned}$$

where the second step is by induction hypothesis. ■

Proof for Proposition 2

It suffices to show the following claim.

Claim 7. *For any $n \in \mathbb{N}$ and $M, u_1, u_2, ..., u_n \in \mathbb{R}_+$, where $u_i < u_{i+1}$ we have, for $k = 1, 2, ..., n$ and $l = 1, 2..., k$ with $M > u_l$*

$$\frac{u_1}{\sum_{i=1}^n u_i} + \frac{u_2}{\sum_{i=1}^n u_i} + ... + \frac{u_k}{\sum_{i=1}^n u_i} \geq \frac{u_k}{\sum_{i=l}^n u_i + M}$$

Proof. It suffices to show that for any $s, A, B > 0$, the fraction $\frac{s+A}{s+A+B}$ is increasing in A, which is straightforward. ■