9-2024

Associative networks in decision making

Jiangtao LI
Singapore Management University, jtli@smu.edu.sg

Rui TANG

Mu ZHANG

Associative Networks in Decision Making*

Jiangtao Li† Rui Tang‡ Mu Zhang§

September 16, 2024

Abstract

We present a model of associative networks that captures how a decision maker expands her consideration set through mental associations between alternatives. This model serves as a tool to understand the influence of association on decision making. As a proof of concept, we characterize this model within a random attention framework and demonstrate that all the relevant parameters are uniquely identifiable. Notably, in a novel choice domain where not all observable alternatives are available, the presence of unavailable alternatives can affect the choice frequencies of other alternatives through association.

Keywords: associative network, random attention, consideration set, random choice, availability and observability

JEL: D01, D91

---

*We thank Paul Cheung, David Dillenberger, Pawel Dziewulski, Junnan He, Matthew Kovach, Jay Lu, Yusufcan Masatlioglu, Pietro Ortoleva, Erkut Ozbay, John K.-H. Quah, Satoru Takahashi, and Junjie Zhou for helpful discussions. All remaining errors are our own. Rui Tang acknowledges the support of the HKUST start-up research fund.

†School of Economics, Singapore Management University, jtli@smu.edu.sg

‡Department of Economics, Hong Kong University of Science and Technology, ruitang@ust.hk

§Department of Economics, University of Michigan, muzhang@umich.edu

1 Introduction

Memory and attention are fundamental cognitive processes essential for decision making (Simon, 1955; Payne et al., 1993; Camerer, 1997). The impact of these processes on decision making has been extensively studied in recent years (Bordalo et al., 2020, 2022). Among the various memory patterns, mental association plays a key role, linking the recall of one item to another based on an individual's prior experience or learning. In this paper, we adopt a choice-theoretical approach to explore the impact of mental association on decision making and develop a choice model to capture the effects of this cognitive process.

To fix ideas and highlight some of the motivations behind our analysis, we consider the recent launch of the Xiaomi SU7, which has sparked considerable online debate due to its striking resemblance to the Porsche Taycan. For reference, images of both vehicles are provided below: the Porsche Taycan from Porsche's official website and the Xiaomi SU7 from Xiaomi's official website.

The Xiaomi SU7 has drawn comparisons to the Porsche Taycan from internet users, with some even dubbing it the "Mi Porsche." This strategic move by Xiaomi to boost its brand recognition through the association with Porsche is evident, and it's easy to understand why this could be valuable for the company. Potential buyers who admire the exterior design of the Porsche Taycan but find it prohibitively expensive might turn to the Xiaomi SU7 as an alternative. To these buyers, while Porsche Taycan is observable, it is unaffordable and effectively unavailable.

It might not be immediately obvious why Porsche would not be concerned about the release of the Xiaomi SU7. Indeed, in an interview during the 4th China International Consumer Goods Fair, Porsche China President and CEO Michael Kirsch addressed the issue for the first time, stating:"As for the similarities between the Xiaomi SU7 and Porsche, I think it's probably that good design always has something in mind."[1] Our model of associative networks offers one possible explanation for this. For some buyers, factors such as premium branding outweigh cost in their decision-making process. While many models are affordable, these buyers might not initially consider all available options due to limited attention. During the initial launch period of the Xiaomi SU7, the press conference, news coverage, and the controversy surrounding its resemblance to the Porsche Taycan likely drew significant attention to the Xiaomi SU7. This attention could lead consumers to also consider the Porsche Taycan. Even though this association was not intentionally created by Porsche, it could increase the chances that the Porsche Taycan—but perhaps not other competitors like Maserati—ends up in consumers' final consideration set. Ultimately, this could work to Porsche's advantage.

Mental association is a cognitive process that allows the decision maker (DM) to expand her consideration set by linking relevant alternatives that may not have been initially considered. We represent this cognitive process through associative networks, a conceptual model first introduced in cognitive psychology (Anderson and Bower, 1973; Anderson, 1996; Raaijmakers and Shiffrin, 1981) and widely applied in the marketing literature (Keller, 1993; Teichert and Schöntag, 2010; Brandt et al., 2011; Cunha Jr et al., 2015). In an associative network, objects (nodes) are connected based on their semantic or conceptual relationships. When a specific node or input is activated, the network retrieves related nodes by spreading activation through the links. In our study, we employ the associative network as a descriptive model that captures how the attention to one alternative can trigger the DM to consider another alternative, abstracting away from the underlying conceptual similarities between alternatives that result in the association. More

[1] See link for a news report on this.

specifically, in our model, a link from alternative x to y indicates that the attention to x can prompt the DM to further consider y.

Notably, the process of mental association is not limited to available alternatives; it can also be triggered by *unavailable but observable* alternatives. In the case of Xiaomi SU7 versus Porsche Taycan, even before the Xiaomi SU7 was officially launched and became available for purchase, the attention it received could still prompt consideration of the Porsche Taycan. Similarly, while the Porsche Taycan may be prohibitively expensive and therefore unaffordable for some buyers—effectively making it unavailable to them—the attention drawn to it can still influence their consideration of the Xiaomi SU7. The importance of including observable but unavailable alternatives into the choice-theoretical framework has been highlighted by the recent experimental studies. For instance, Soltani et al. (2012) demonstrate that decoy alternatives can induce choice reversals, such as the attraction effect and the compromise effect, even when these alternatives are not available for selection.[2] In an experimental design with differentiated products and objective payoffs, Chadd et al. (2021) find that the presence of unavailable alternatives can lead to suboptimal decisions and longer decision times, with participants willing to pay significant amounts to avoid being exposed to these alternatives.

Throughout the paper, we study a novel choice domain where not all observable alternatives are available. Each menu is composed of two distinct sets of alternatives and is represented as a pair (A, B). While all the alternatives in A and B are observable, only the alternatives in A are available. Although the DM cannot choose any alternative from B, the presence of these unavailable alternatives can still influence the DM's attention through association, thereby affecting their choices. Choice scenarios are abundant where some alternatives are observable but unavailable—whether due to high cost, being out of stock, or restricted access for members or VIPs. Some items are costly, effectively making them unavailable. Sold-out products remain to be displayed in many online shopping platforms. Some products are exclusive to members, such as subscription-only movies on Netflix or bags reserved for loyal customers in luxury stores, making them observable

[2] Lea and Ryan (2015) report similar choice reversals in the mate choice of Túngara frogs.

but not available to non-members.[3] By incorporating observable but unavailable alternatives, our primitive—a random choice rule—is a function that maps every menu (A, B) to a distribution over A and a default option. This distribution represents the DM's choice frequencies of alternatives in A when presented with the menu (A, B).

Section 3 introduces our choice model. When presented with a menu (A, B), the DM initially considers a random subset of alternatives from $A \cup B$, which we refer to as her initial consideration set. Following Manzini and Mariotti (2014) (henceforth MM14), we assume that each alternative has a fixed probability of being initially considered by the DM, and that the DM directs her attention to each alternative independently. The DM then associates relevant alternatives in $A \cup B$ with those in her initial consideration set, a process captured by the DM's associative network, which is represented by a directed graph over the alternatives. Each link in this associative network is an ordered pair of alternatives (x, y), indicating that the consideration of x prompts the DM to further consider y. The association process continues until no further alternatives in $A \cup B$ can be linked to those already considered, resulting in a final consideration set C. The DM then selects her most preferred alternative among the available alternatives in C—she chooses her most preferred alternative in $A \cap C$ if it is not empty; otherwise, the default alternative is selected. We refer to the random choice rule induced by this choice procedure as the **A**ssociation **B**ased **C**onsideration rule (ABC).

Section 4 presents the axioms that characterize ABCs. These axioms separately address the underlying attention distribution, the association procedure, the associative network, and the revealed preference relation within our choice model. Specifically:

- Axiom 1 specifies the attention distribution: The DM directs her initial attention to each alternative independently.
- Axiom 2 states that if an observable but unavailable alternative x is revealed to prompt the DM to consider some available alternative in a given menu,

[3] There are numerous anecdotes of consumers being unable to purchase popular items like certain Hermès bags or Rolex watches even when they can afford the items and the items are in stock. For example, see link that discusses how difficult it is to buy a Birkin bag:"Buying your very first Birkin bag from a Hermès boutique is notoriously difficult. [...] you can't simply waltz into a store and buy one off the shelf."

then making x available does not affect the choice frequency of the default alternative. By this axiom, the DM's association procedure is only relevant with the observability of the alternatives and is unaffected by their availability. Consequently, so long as the attention to x leads to the consideration of some available alternative, the choice of the default alternative is blocked, no matter whether x is available or not.

- Axioms 3 and 4 characterize the associative network. Axiom 3 states that the DM can associate more alternatives with a given alternative when there are more observable alternatives. Axiom 4 states that if the DM can associate z with x when y is observable but fails to do so when y is not observable, then she must associate z with x through the intermediate alternative y.

- Axiom 5 characterizes the underlying preference relation identified from the DM's choices. The axiom states that, for two given alternatives, if the availability of one alternative affects the choice frequency of the other in some menu, then the inverse does not occur in any menu where both alternatives are observable. Essentially, the axiom posits that an inferior alternative can only affect the choice frequency of a better alternative through its observability but not its availability.

Notably, all the relevant parameters of an ABC can be uniquely identified.

We extend the ABCs to account for the possibility that the DM may stop the mental association process before considering all alternatives that can be associated. This scenario could arise due to mental fatigue or when the DM feels that she has considered enough options. Importantly, this extension captures the idea that when alternatives are mentally more distant, meaning that the DM needs more rounds of mental association to connect them, it is more challenging for the DM to initially pay attention to one alternative and eventually include the other in her final consideration set. Section 5 introduces a generalized choice rule called **A**ssociation **B**ased **C**onsideration with **T**ermination rule (ABCT), which incorporates the random termination of mental association. We show that this generalized rule can be characterized by replacing Axiom 2 with two axioms: Axiom 6 imposes a monotonicity condition on the choice frequency of the default alternative, and Axiom 7 weakens Axiom 2 by stating that moving the observable but unavailable alternatives that are directly associated with available alternatives to the available set has a consistent impact on the choice frequency of the default alternative.

Section 6 considers an application of our model in which a multi-product seller seeks to enhance the sales of a specific product by adding an additional association link to its existing network. This added link could be literal or metaphorical. For instance, platforms like Amazon might create an association from x to y by recommending y on the product page of x. Alternatively, firms could run advertisements that strengthen consumers' mental association from x to y. However, constraints such as limited space on product pages or the high costs of advertising limit the ability to add multiple links indiscriminately. We examine the optimal strategy for the seller.

In Section 7 and the Online Appendix, we consider various extensions of the ABCs that feature (i) limited data, (ii) random associative networks and preferences, and (iii) general models of initial attention. We focus on the identification of parameters in these generalized models.

1.1 Related Literature

Our paper belongs to the growing literature on choices with limited attention or limited consideration.[4] In particular, our approach is closely related to that of MM14, as both models assume that the DM allocates her initial attention randomly and independently. However, our model differs from the model of MM14 in that our DM has a follow-up procedure through which she continues to expand her consideration set via mental association. When the DM does not engage in any mental association, our model reduces to that of MM14. When every pair of alternatives are associated with each other, our model reduces to the rational choice model, where the DM always selects the best available alternative whenever she initially pays attention to some observable option.

In a concurrent paper, Yegane and Masatlioglu (2023) consider a two-stage stochastic consideration set formation process where the first stage follows MM14. For a given initial consideration set (which they refer to as the awareness set), the

[4]See, for instance, Masatlioglu et al. (2012), MM14, Brady and Rehbeck (2016), Dean et al. (2017), Lleras et al. (2017), Cattaneo et al. (2020), Dardanoni et al. (2020), Barseghyan et al. (2021a), Barseghyan et al. (2021b), and Cattaneo et al. (2023).

DM observes options sequentially and may forget previously observed ones due to limited memory (Yegane, 2022). As a result, the final consideration set is a *subset* of the initial one. By contrast, we focus on the mental association process, and the final consideration set is a *superset* of the initial one.

Our model makes three novel contributions to the literature on limited attention. First, we examine the cognitive process of mental association, which is a fundamental mechanism in forming the DM's consideration set. We provide a concrete procedure (and a random version of it) for how this process operates. Second, our model incorporates bottom-up attention (initial random attention) and top-down attention (mental association), both of which have been shown to be influential factors in decision making (Corbetta and Shulman, 2002; Geng and Behrmann, 2005; Gazzaley and Nobre, 2012).[5] Third, we investigate the impact of unavailable but observable alternatives on the DM's choices. While those alternatives, which are also called "phantom" options (Farquhar and Pratkanis, 1993), have been studied theoretically in the literature (Guney et al., 2018; Natenzon, 2019),[6] we focus on investigating how those alternatives affect the DM's attention and obtain a unique identification of our model through those alternatives.

There are a few papers studying the role of networks in individual decisions (Masatlioglu and Nakajima, 2013; Masatlioglu and Suleymanov, 2021; Ellis and Thysen, 2024; Valkanova, 2021). Among them the most related paper to ours is Masatlioglu and Suleymanov (2021). In their model, the DM is endowed with an *undirected* associative network. When faced with a menu of available products and an exogenous starting point, she forms her consideration set by including objects that are connected to the starting point through a path (with a potential cap on the

[5]Bottom-up attention involves the automatic processing of sensory stimuli in the environment, such as sudden loud noises or bright lights, that capture an individual's attention involuntarily. By contrast, top-down attention usually refers to the deliberate allocation of attention that is guided by the individual. In our model, the DM's initial attention is more likely to be bottom-up, as the DM is randomly attracted by the stimuli or salient features of the options. The second-stage mental association is a mixture of bottom-up and top-down attention, as some associated alternatives may come to mind unintentionally, and individuals may also direct their attention towards options that are relevant to what they have considered in certain dimensions.

[6]Natenzon (2019) considers a DM who is imperfectly informed about the value of available options and can derive additional information from phantom options. In the choice model of Guney et al. (2018), the best phantom option may serve as aspiration, and the DM chooses the closest available alternative to it.

length of the path). In the working paper version of Masatlioglu and Suleymanov (2021), the authors also study extensions with unobservable starting points and random network.[7] By comparison, our model studies the combination of both initial random attention and mental association through a directed associative network, and investigates the role played by unavailable but observable alternatives in this process. Masatlioglu and Nakajima (2013) study a general model of how behavioral search affects the formation of consideration sets. In their model, the connections among the alternatives that determine the search order of the DM can be represented as a network. Valkanova (2021) models the exploration of the choice set as a discrete-time Markov chain in which DMs search sequentially by making stochastic pairwise comparisons. Ellis and Thysen (2024) use a directed acyclic network to represent the DM's subjective casual model.

More broadly, our paper contributes to the literature on random choices. Various models have been proposed to rationalize random choice behavior, including the possibility that the DM has random utilities, leading to stochastic choices as a result of utility maximization (Block and Marschak, 1960; Falmagne, 1978; Gul and Pesendorfer, 2006; Gul et al., 2014),[8] and the possibility that the DM randomizes deliberately (Cerreia-Vioglio et al., 2019; Agranov and Ortoleva, 2022). While the randomness in our DM's choice behavior is driven by random attention, our work emphasizes the importance of understanding mental associations as a important channel for forming the consideration sets.

Our work also relates to the literature on how choices are influenced by factors beyond the choice menu. These factors can include frames (Salant and Rubinstein, 2008), the DM's reference points or status quo (Masatlioglu and Ok, 2005, 2014; Kovach and Suleymanov, 2023), and recommendations from external sources (Cheung and Masatlioglu, 2023, 2024), among others. While our approach shares some similarities with the work of Kovach and Suleymanov (2023) which examines how reference points can shape the DM's attention, our study focuses on understanding how unavailable but observable alternatives prompt the DM to pay

[7] In their extension with random network, the DM is assumed to consider the alternatives that are directly linked to the starting point.

[8] Among the most influential random utility models are the multinomial logit (Luce, 1959) and nested logit models (Ben-Akiva, 1973; McFadden, 1978), which are widely used in structural estimations. See also Kovach and Tserenjigmid (2022) for their behavioral foundations.

attention to available alternatives through mental association.

2 Preliminaries

There is a nonempty finite set of alternatives X, with generic elements denoted by x, y, z, etc. Denote by $\mathcal{M}$ the collection of all subsets of X, with generic elements denoted by A, B, C, etc. We allow the DM to not pick any alternative from a set of alternatives, so we also assume the existence of a default alternative (e.g., walking away from the shop, abstaining from voting).[9] When there is no confusion, we write AB for $A \cup B$, Ax for $A \cup \{x\}$, and $A \setminus x$ for $A \setminus \{x\}$.

A menu consists of two distinct sets of alternatives and is represented by a pair $(A, B) \in \mathcal{M} \times \mathcal{M}$ with $A \cap B = \emptyset$. While all the alternatives in AB are observable, only the alternatives in A are available to the DM. In other words, the DM can pay attention to alternatives in AB but can only choose from A or choose the default alternative. For ease of reference, we refer to A as the set of available alternatives, or simply the available set, and B as the set of observable but unavailable alternatives. Let $\mathcal{E}$ denote the collection of all menus.

Let $\mathcal{X} = \{(x, x) : x \in X\}$. A binary relation on X is a subset $\mathcal{R} \subseteq X \times X$. We say that $\mathcal{R}$ is reflexive if $\mathcal{X} \subseteq \mathcal{R}$. For all $x, y \in X$, we write $x\mathcal{R}y$ if $(x, y) \in \mathcal{R}$ and use these two notations interchangeably. For all $x \in X$, let $\mathcal{R}(x) = \{y \in X : x\mathcal{R}y\}$, and for all nonempty $A \subseteq X$, let $\mathcal{R}(A) = \cup_{x \in A}\mathcal{R}(x)$. Let $\mathcal{R}^0 = \mathcal{X}$. For all $k \in \mathbb{N}_+$, define $\mathcal{R}^k$ such that $x\mathcal{R}^k y$ if and only if there exists $1 \leq t \leq k$ and $\{x_1, x_2, \ldots, x_{t+1}\} \subseteq X$ such that $x_1 = x$, $x_{t+1} = y$, and $x_m\mathcal{R}x_{m+1}$ for all $m \in \{1, 2, \ldots, t\}$. We define the transitive closure of $\mathcal{R}$ as $\mathcal{R}^+ := \cup_{k=1}^{+\infty}\mathcal{R}^k$.

A random choice rule is a map $\rho : X \times \mathcal{E} \to [0, 1]$ such that for all $(A, B) \in \mathcal{E}$, (i) $A \neq \emptyset$ implies $\sum_{x \in A} \rho(x, (A, B)) \in (0, 1)$, and (ii) $\rho(x, (A, B)) > 0$ implies $x \in A$. For ease of notation, we write $\rho(x|A, B)$ rather than $\rho(x, (A, B))$. Define $\Phi_\rho(A, B) = 1 - \sum_{x \in A} \rho(x|A, B)$. The interpretation is that (1) $\rho(x|A, B)$ denotes the probability that the DM chooses the alternative x in the menu (A, B), and (2) $\Phi_\rho(A, B)$ denotes the probability that the DM chooses the default alternative in the menu (A, B).

[9] For recent work on allowing "not choosing" in a random choice setting, see MM14, Brady and Rehbeck (2016), and Dardanoni et al. (2020), among others.

In the definition of the random choice rule above, condition (i) states that the probability of choosing the default alternative is positive in any menu, which happens if the DM does not pay attention to any alternative in A. When A is the empty set, clearly we have $\Phi_\rho(A, B) = 1$.

A preference ordering $\succ$ is a strict total order defined on X. We use $\max(A; \succ)$ to denote the $\succ$-maximal alternative in A whenever A is not empty.

3 Association Based Consideration

In this section, we formally introduce our choice model in which a DM forms her consideration set through mental association.

Initial consideration set. Following MM14, we assume that each alternative has a fixed probability of being initially considered by the DM, and that the DM attributes her attention to each alternative independently.[10] The attention probability of each alternative is given by the function $\pi : X \to (0, 1)$. For a given π, define $\mathring{\pi} : X \to (0, 1)$ such that $\mathring{\pi}(x) = 1 - \pi(x)$ for all $x \in X$. To simplify the notation, we write π_x for $\pi(x)$, $\mathring{\pi}_x$ for $\mathring{\pi}(x)$, π_A for $\prod_{x \in A} \pi(x)$, and $\mathring{\pi}_A$ for $\prod_{x \in A} \mathring{\pi}(x)$. We use the convention that $\pi_A = \mathring{\pi}_A = 1$ when A is empty.

In a given menu (A, B), the DM initially considers a subset of AB. Since the DM attributes her attention to each alternative independently, the DM initially pays attention to some $C \subseteq AB$ with probability $\pi_C \mathring{\pi}_{(AB) \setminus C}$.

Associative network and the final consideration set. The DM expands her initial consideration set through an associative network $\mathcal{N} \subseteq X \times X$, which is a reflexive binary relation on X.[11] If $(x, y) \in \mathcal{N}$, then y is **directly associated with** x, and the attention to x will prompt the DM to further consider y.

In a given menu (A, B), the DM's mental association process only depends on the restricted associative network $\mathcal{N}_{AB}$ on AB, where $\mathcal{N}_{AB} = \{(x, y) \in \mathcal{N} : x, y \in AB\}$. With $\mathcal{N}_{AB}$, the DM's mental association process works as follows. For each alternative $x \in AB$ that she initially considers, she includes every alternative y in $\mathcal{N}_{AB}(x)$ into her consideration set.[12] For each such alternative y, she then

[10] This attention rule is also studied by Manski (1977) and Barseghyan et al. (2021b).

[11] We will also use $\mathcal{W}, \mathcal{U}$ and $\mathcal{V}$ to denote generic associative networks.

[12] Note that $\mathcal{N}_{AB}$ is a binary relation and $\mathcal{N}_{AB}(x) = \{y \in X : x \mathcal{N}_{AB} y\}$.

expands her consideration set by including each alternative z in $\mathcal{N}_{AB}(y)$. The process terminates when there are no more alternatives in AB that are associated with what the DM already considers.

Formally, the association procedure described above is modeled as follows. Consider the transitive closure $\mathcal{N}^+_{AB}$ of $\mathcal{N}_{AB}$.[13] If $(x, y) \in \mathcal{N}^+_{AB}$, then there exists a path, i.e., a sequence of alternatives $x_1, x_2, \ldots, x_{n+1} \in AB$ such that $x_1 = x$, $x_{n+1} = y$, and x_{k+1} is directly associated with x_k for all $k \in \{1, 2, \ldots, n\}$. Therefore, the DM can finally consider y as long as she considers x. For a given menu (A, B), we say that y is **associated with** x in (A, B) if $(x, y) \in \mathcal{N}^+_{AB}$.[14] Note that if the DM initially considers x, the set $\mathcal{N}^+_{AB}(x)$ will be included in her final consideration set. Thus, an initial consideration set $C \subseteq AB$ leads to the final consideration set $\mathcal{N}^+_{AB}(C)$. Figure 1 is a graphic illustration of this process.

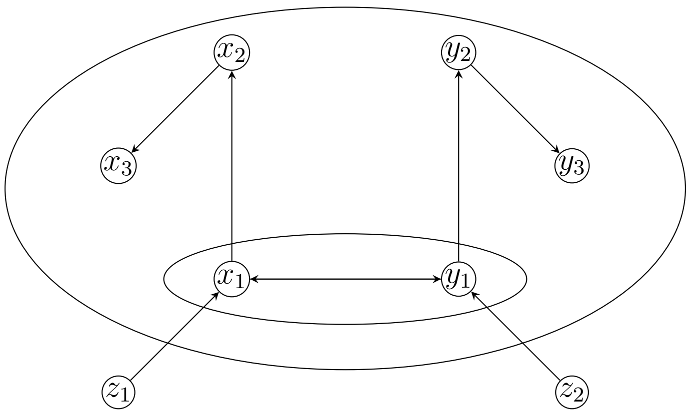

Figure 1: The menu contains 8 alternatives. The associative network is given by the arrows. The initial consideration set of the DM is $\{x_1, y_1\}$, and after association, her final consideration set is $\{x_1, x_2, x_3, y_1, y_2, y_3\}$. Alternatives z_1 and z_2 are not considered because they are not directly associated with any considered alternative.

One important feature of our model is that the DM's association process depends on unavailable yet observable alternatives.[15] In our model, such alternatives can help the DM to consider more available alternatives through association. To see this, consider two menus $(\{z\}, \{x\})$ and $(\{z\}, \{x, y\})$, and assume that y is directly associated with x and z is directly associated with y, but z is not directly associated

[13]Throughout the paper, $\mathcal{N}^+_{AB}$ denotes the transitive closure of $\mathcal{N}_{AB}$ but **not** the transitive closure of $\mathcal{N}$ restricted on AB. Similarly, $\mathcal{N}^k_{AB}$ denotes $(\mathcal{N}_{AB})^k$ but **not** $(\mathcal{N}^k)_{AB}$.

[14]When there is no confusion about menu (A, B), we simply say that y is associated with x.

[15]The importance of understanding how such alternatives affect decision-making has also been highlighted by Chadd et al. (2021).

with x. If the initial consideration set of the DM is $\{x\}$, then she cannot further consider z in menu $(\{z\},\{x\})$, but can do so through an intermediate alternative y in menu $(\{z\},\{x,y\})$.

We note that our model can also accommodate the case in which the DM is able to store the associated alternative y in her memory and use it for further association. By doing so, she can directly associate z with x even in menu $(\{z\},\{x\})$. This seems incompatible with our model because an alternative that is not observable serves as the intermediate alternative in the DM's association process. In fact, our model can accommodate such a case: If the DM can associate z with x through y without y being observable, then it is as if the DM can directly associate z with x. The two interpretations lead to the same choice behavior of the DM, and therefore, we do not distinguish them in our choice model.

Preference and choice. In the menu (A, B), if the DM's final consideration set is $C \subseteq AB$, then she chooses $\max(A \cap C; \succ)$ if $A \cap C$ is not empty. Otherwise, she chooses the default alternative.

Definition 1. *A random choice rule ρ is an **A**ssociation **B**ased **C**onsideration rule (ABC) if there exists a tuple $(\pi, \mathcal{N}, \succ)$, where π is an attention probability function, $\mathcal{N}$ is an associative network, and $\succ$ is a preference ordering, such that*

$$\rho(x|A,B) = \sum_{C\subseteq AB:\, x=\max(\mathcal{N}^{+}_{AB}(C)\cap A;\,\succ)} \pi_C \mathring{\pi}_{(AB)\setminus C} \tag{1}$$

for any $(A, B) \in \mathcal{E}$ and $x \in A$. The tuple $(\pi, \mathcal{N}, \succ)$ is said to represent ρ as an ABC.

With an ABC, the choice probability of x in the menu (A, B) is the frequency with which x is the best alternative in the final consideration set.[16]

[16]Following MM14, we assume a unique preference ordering. An interesting extension is to allow heterogeneous preferences (Barseghyan et al. 2021a, Barseghyan et al. 2021b, Kashaev and Aguiar 2022, Aguiar et al. 2023, Cattaneo et al. 2023, Cheung and Masatlioglu 2024). We discuss it in Section 7.

4 Axioms and Representation Theorem

In this section, we first introduce a reformulation of the ABC model to highlight some of its key properties. Then we present the axioms and the representation theorem, and discuss the comparative statics of our model.

4.1 Reformulation

A feature of our model is that not every available alternative in a menu is chosen with a positive probability. In particular, if the attention to some alternative x always prompts the DM to consider another better alternative, then x is never chosen. The following proposition characterizes the set of alternatives that are chosen with a positive probability in a given menu.

Proposition 1. *Consider an ABC ρ that is represented by $(\pi, \mathcal{N}, \succ)$. For all $(A, B) \in \mathcal{E}$, $\rho(x|A, B) > 0$ if and only if $x = \max(\mathcal{N}^+_{AB}(x) \cap A; \succ)$.*

In words, an alternative x is chosen with a positive probability in the menu (A, B) if and only if there is no association path from x to any available alternative y that is better than it. Otherwise, the attention to x always prompts the attention to y in this menu, which blocks the choice of x.

Next, we investigate the choice frequency of each chosen alternative. For a given menu (A, B), define $H_{\mathcal{N}}(A, B) := \{x \in AB : \mathcal{N}^+_{AB}(x) \cap A \neq \emptyset\}$ as the set of alternatives x in the menu (A, B) such that some alternative in A is associated with x. Since $\mathcal{N}$ is reflexive, we have $A \subseteq H_{\mathcal{N}}(A, B)$.

Proposition 2. *Consider an ABC ρ that is represented by $(\pi, \mathcal{N}, \succ)$. For all $(A, B) \in \mathcal{E}$,*

$$\Phi_\rho(A, B) = \mathring{\pi}_{H_{\mathcal{N}}(A,B)}. \tag{2}$$

Furthermore, if $\{x \in A : \rho(x|A, B) > 0\} = \{x_1, x_2, \dots, x_n\}$ with $x_1 \succ x_2 \succ \dots \succ x_n$, then

$$\rho(x_1|A, B) = 1 - \mathring{\pi}_{C_1} \text{ and } \rho(x_k|A, B) = \left(1 - \mathring{\pi}_{C_k}\right) \prod_{t=1}^{k-1} \mathring{\pi}_{C_t}, \ \forall k \geq 2, \tag{3}$$

where $C_k = \{y \in AB : x_k = \max(\mathcal{N}^+_{AB}(y) \cap A; \succ)\}$ for all $k \in \{1, 2, \dots, n\}$.

Equation (2) says that the probability for the default option to be chosen is equal to the probability that none of the initially considered alternatives lead to the consideration of any available alternative. Equation (3) is a reformulation of our choice rule: The probability for an alternative to be chosen is equal to the probability that it is finally considered through the association process while all better available alternatives are not.

4.2 Characterization

The first axiom captures the independent attention distribution of the DM.

Axiom 1—Default Independence: For all $x \in X$ and $A, B \in \mathcal{M}$ with $x \in A \cap B$,

$$\frac{\Phi_\rho(A, \emptyset)}{\Phi_\rho(A \setminus x, \emptyset)} = \frac{\Phi_\rho(B, \emptyset)}{\Phi_\rho(B \setminus x, \emptyset)}.$$

Axiom 1 follows from the I-Independence axiom of MM14.[17] When all observable alternatives are available, the DM's choice of the default option depends on whether her initial consideration set is empty or not. The effect of removing an available alternative on the frequency of choosing the default option is determined by the extent to which it attracts the DM's initial attention. Axiom 1 posits that the DM has a constant probability of initially considering a given alternative.

Definition 2. *For all $x \in X$, $A, B \in \mathcal{M}$ with $A \cap B = \emptyset$, some alternative in A is associated with x through B, denoted by $x \xrightarrow[\rho]{B} A$, if either $x \in A$ or $\Phi_\rho(A, B) \neq \Phi_\rho(A, B \setminus x)$.*

When there is no confusion about ρ, we write $x \xrightarrow{B} A$ for $x \xrightarrow[\rho]{B} A$. To understand Definition 2, note that if x is in A, then clearly some alternative in A (i.e., x) is associated with x. If x is not in A, then the condition $\Phi_\rho(A, B) \neq \Phi_\rho(A, B \setminus x)$ indicates $x \in B$. Furthermore, since removing x from B affects the choice frequency of the default alternative, it follows that the attention of x must lead to the choice of some alternative in A, i.e., some alternative in A is associated with x, possibly through some intermediate alternatives in B.

[17]The I-Independence axiom is stronger than Axiom 1. It additionally requires that for all $x, y \in X$ and $A, B \in \mathcal{M}$ with $x, y \in A \cap B$ and $x \neq y$, $\frac{\rho(x|A\setminus y,\emptyset)}{\rho(x|A,\emptyset)} = \frac{\rho(x|B\setminus y,\emptyset)}{\rho(x|B,\emptyset)}$.

Axiom 2—Idempotence: For all $x \in X$ and $(A, B) \in \mathcal{E}$ with $x \in B$, if $x \xrightarrow{B} A$, then $\Phi_\rho(Ax, B \setminus x) = \Phi_\rho(A, B)$.

Axiom 2 states that if some available alternative in A is associated with some observable but unavailable alternative x, then the DM's attention to x will result in the selection of an available alternative and block the choice of the default alternative. Therefore, whether x is available or unavailable has no effect on the choice frequency of the default alternative.

Axiom 3—Expansion: For all $x \in X$ and $(A, B), (C, D) \in \mathcal{E}$, if $C \subseteq A$ and $CD \subseteq AB$, then $x \xrightarrow{D} C$ implies $x \xrightarrow{B} A$.

Axiom 3 states that if the attention to x leads to the consideration of some available alternative in a given menu, then the same would be true in a less restrictive menu with more observable alternatives and more available alternatives. In other words, when there are more intermediate cues in the form of observable alternatives, more alternatives become associated with x.

Axiom 4—Path Connectedness: For all $x, y \in X$ and $(A, B) \in \mathcal{E}$ with $x \neq y$, if $x \xrightarrow{B} A$ and not $x \xrightarrow{B \setminus y} A \setminus y$, then $x \xrightarrow{B \setminus y} \{y\}$ and $y \xrightarrow{B \setminus x} A$.

Axiom 4 captures the key feature of the associative network: One alternative is associated with another through paths in the network. To see this, note that if the DM can associate some alternative z in A with x through B but cannot do so when y is not observable, then either $y = z$ or y is an intermediate alternative for this association process.

Axiom 5—Association Asymmetry: For all $x, y \in X$ and $(A, B), (C, D) \in \mathcal{E}$ with $x \neq y$ and $x, y \in A \cap C$,

$$\rho(y|A, B) \neq \rho(y|A \setminus x, Bx) \Rightarrow \rho(x|C, D) = \rho(x|C \setminus y, Dy).$$

Axiom 5 is similar to the I-Asymmetry axiom of MM14.[18] It states that if the availability of an alternative x affects the choice frequency of alternative y,

[18]The I-Asymmetry axiom states that for all distinct $x, y \in X$ and $A, B \in \mathcal{M}$, if $\rho(y|A, \emptyset) \neq \rho(y|A \setminus x, \emptyset)$, then $\rho(x|B, \emptyset) = \rho(x|B \setminus y, \emptyset)$.

then the inverse will not occur. To understand this axiom, note that if the set of observable alternatives is kept the same, so will be the distribution of the final consideration sets. Under such circumstance, the availability of x affects the choice of y only when x is better than y, and thus the inverse never occurs.

Together, Axioms 1-5 fully characterize our ABC model.

Theorem 1. *A random choice rule ρ is an ABC if and only if it satisfies Axioms 1-5. The tuple $(\pi, \mathcal{N}, \succ)$ that represents ρ as an ABC is unique.*

Identification of the parameters. The identification of the attention probability function π is the same as that in MM14. For every $x \in X$, $\pi_x = \rho(x|\{x\}, \emptyset)$.

The preference ordering $\succ$ can be identified through the DM's choice frequencies in binary menus. Consider two alternatives x and y and assume $x \succ y$. Following the interpretation of Axiom 5, the choice frequencies of x are the same in the two menus $(\{x, y\}, \emptyset)$ and $(\{x\}, \{y\})$. However, the choice frequencies of y must differ in the two menus $(\{x, y\}, \emptyset)$ and $(\{y\}, \{x\})$: In the menu $(\{x, y\}, \emptyset)$, the choice frequency of y is at most $\mathring{\pi}_x \pi_y$, while in the menu $(\{y\}, \{x\})$, the choice frequency of y is at least π_y. Therefore, $x \succ y$ if and only if $\rho(y|\{x, y\}, \emptyset) \neq \rho(y|\{y\}, \{x\})$.

For the associative network $\mathcal{N}$, note that x is associated with y if and only if $\rho(x|\{x\}, \{y\}) \neq \rho(x|\{x\}, \emptyset)$, i.e., removing the observable but unavailable alternative y affects the choice frequency of x. Given our identification strategy, the tuple $(\pi, \mathcal{N}, \succ)$ is unique.

Proof sketch of Theorem 1. We focus on the sufficiency part. With the identified parameters $(\pi, \mathcal{N}, \succ)$, we briefly demonstrate how our axioms lead to the desired representation. In Step 1, we show that for all menu (A, B) and alternative $x \in B$, $x \xrightarrow{B} A$ if and only if some alternative in A is associated with x, i.e., $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$. By this, we can focus on a subset $C \subseteq B$ such that C contains all the alternatives with which some alternative in A is associated. By the Idempotence axiom (Axiom 2) and the definition of $\xrightarrow{B}$, we have $\Phi_\rho(A, B) = \Phi_\rho(AC, B \setminus C) = \Phi_\rho(AC, \emptyset)$. Essentially, the set AC is equal to $H_\mathcal{N}(A, B)$, and we can show that $\Phi_\rho(A, B) = \mathring{\pi}_{H_\mathcal{N}(A,B)}$.

In Step 2, we show that $\succ$ is defined for each distinct pair of alternatives and satisfies asymmetry and transitivity, i.e., it is indeed a preference ordering.[19]

The final step is to show that for any given menu, each alternative is chosen with the probability specified by the model. To illustrate, we provide a simple example. Consider a menu $(\{x,y,z,w\},\{r\})$ such that

$$x \succ y \succ z \succ w \succ r \quad \text{and} \quad \mathcal{N}_{\{x,y,z,w,r\}} = \mathcal{X} \cup \{(w,y),(r,z)\}.$$

Since $r\mathcal{N}z$, by Step 1, we have

$$\begin{aligned}\sum_{\hat{x}\in\{x,y,z,w\}} \rho(\hat{x}|\{x,y,z,w\},\{r\}) &= 1 - \Phi_\rho(\{x,y,z,w\},\{r\}) \\ &= 1 - \Phi_\rho(\{x,y,z,w,r\},\emptyset) = 1 - \mathring{\pi}_{\{x,y,z,w,r\}}.\end{aligned} \tag{4}$$

Since for all $\hat{x} \in \{x,y,z\}$, $\hat{x} \succ w$, by the Association Asymmetry axiom (Axiom 5), making w unavailable does not affect the choice frequencies of x, y and z. Since $r\mathcal{N}z$ and $w\mathcal{N}y$, again by Step 1, we have

$$\begin{aligned}\sum_{\hat{x}\in\{x,y,z\}} \rho(\hat{x}|\{x,y,z,w\},\{r\}) &= \sum_{\hat{x}\in\{x,y,z\}} \rho(\hat{x}|\{x,y,z\},\{w,r\}) \\ &= 1 - \Phi_\rho(\{x,y,z\},\{w,r\}) \\ &= 1 - \Phi_\rho(\{x,y,z,w,r\},\emptyset) = 1 - \mathring{\pi}_{\{x,y,z,w,r\}}.\end{aligned} \tag{5}$$

Similarly, we can consecutively move z and y to the observable but unavailable set and obtain

$$\begin{aligned}\sum_{\hat{x}\in\{x,y\}} \rho(\hat{x}|\{x,y,z,w\},\{r\}) &= \sum_{\hat{x}\in\{x,y\}} \rho(\hat{x}|\{x,y\},\{z,w,r\}) \\ &= 1 - \Phi_\rho(\{x,y\},\{z,w,r\}) \\ &= 1 - \Phi_\rho(\{x,y,w\},\emptyset) = 1 - \mathring{\pi}_{\{x,y,w\}}.\end{aligned} \tag{6}$$

$$\begin{aligned}\rho(x|\{x,y,z,w\},\{r\}) &= \rho(x|\{x\},\{y,z,w,r\}) \\ &= 1 - \Phi_\rho(\{x\},\{y,z,w,r\}) \\ &= 1 - \Phi_\rho(\{x\},\emptyset) = 1 - \mathring{\pi}_x.\end{aligned} \tag{7}$$

[19]The binary relation $\succ$ is asymmetric if for all $x, y \in X$, $x \succ y$ implies not $y \succ x$, and is transitive if for all $x, y, z \in X$, $x \succ y$ and $y \succ z$ imply $x \succ z$.

Note that equations (4)-(7) pin down the choice probabilities of x, y, z and w in menu $(\{x, y, z, w\}, \{r\})$, which are consistent with equation (3) in Proposition 2.

Comparative Statics. We end this section by discussing the comparative statics of our model. We show that we can directly compare two DMs' associative networks without imposing any restriction on the alignment of their preferences or attention probabilities. The following proposition directly follows from the construction of $\mathcal{N}$, and its proof is omitted.

Proposition 3. *For any two ABCs ρ_1 and ρ_2 that are represented by $(\pi, \mathcal{N}, \succ)$ and $(\pi', \mathcal{W}, \succ')$ respectively, the following statements are equivalent:*

(i) *For all $x \in X$ and $(A, B) \in \mathcal{E}$ with $x \notin AB$,*

$$\Phi_{\rho_1}(A, B) \neq \Phi_{\rho_1}(A, Bx) \Rightarrow \Phi_{\rho_2}(A, B) \neq \Phi_{\rho_2}(A, Bx).$$

(ii) *The associative network $\mathcal{N}$ is a subset of $\mathcal{W}$.*

5 Association Based Consideration with Termination

In this section, we examine a generalized model where the DM has a fixed probability of terminating the consideration of additional alternatives at each round of the mental association process. This generalized model becomes more relevant when there are many alternatives or when the DM experiences mental fatigue and chooses to halt mental association once a sufficient number of alternatives have been considered. The model also captures the idea that when two alternatives are mentally more distant, meaning that the DM needs more rounds of mental association to connect them, it becomes more challenging for the DM to initially pay attention to one alternative and eventually include the other in her final consideration set.

Definition 3. *A random choice rule ρ is an **A**ssociation **B**ased **C**onsideration with **T**ermination rule (ABCT) if there exists a tuple $(\pi, \mathcal{N}, \succ, \eta)$, where π is an attention probability function, $\mathcal{N}$ is an associative network, $\succ$ is a preference* *ordering, and* $\eta \in (0,1]$ *such that for all* $(A,B) \in \mathcal{E}$ *and* $x \in A$,

$$\begin{aligned}\rho(x|A,B) =& \sum_{k=0}^{+\infty}(1-\eta)\eta^k \left(\sum_{C \subseteq AB:\, x = \max(\mathcal{N}^k_{AB}(C);\succ)} \pi_C \mathring{\pi}_{(AB)\setminus C} \right) \\ &+ (1 - \sum_{k=0}^{+\infty}(1-\eta)\eta^k) \left(\sum_{C \subseteq AB:\, x = \max(\mathcal{N}^+_{AB}(C);\succ)} \pi_C \mathring{\pi}_{(AB)\setminus C} \right).\end{aligned} \tag{8}$$

The tuple $(\pi, \mathcal{N}, \succ, \eta)$ *is said to represent* ρ *as an ABCT.*

The new parameter η is the probability for the DM to continue one more round of mental association. Thus, the probability for the DM to have exactly k rounds of mental association is given by $(1-\eta)\eta^k$, in which case she expands her initial consideration set C to $\mathcal{N}^k_{AB}(C)$. The second term of the right-hand-side of equation (8) captures the case $\eta = 1$, where the DM always expands her initial consideration set C to $\mathcal{N}^+_{AB}(C)$. In this case, an ABCT reduces to an ABC.

For a given ABCT, if an observable but unavailable alternative x can direct the DM's attention to some available alternative, the attention to x may not guarantee that the DM chooses some available alternative in the end, since there is a chance that the DM does not conduct any round of mental association. Therefore, moving x to the available set may alter the choice frequency of the default option, and Axiom 2 no longer holds. We replace Axiom 2 with the next two axioms.

Axiom 6—Default Monotonicity: For all $(A,B),(C,D) \in \mathcal{E}$, if $C \subseteq A$ and $CD \subseteq AB$, then $\Phi_\rho(A,B) \leq \Phi_\rho(C,D)$.

Axiom 6 states that if there are more observable alternatives and more available alternatives, then the frequency for the DM to choose the default option becomes lower. Intuitively, more observable alternatives enable the DM to use mental association to include more alternatives into her final consideration set, and with more available alternatives, there is a higher chance that there is at least one available alternative in the DM's final consideration set.

For any set of alternatives A, define $\mathcal{D}_\rho(A) := \{x \in X \setminus A : x \xrightarrow{\{x\}} A\}$ as the set of alternatives x in $X \setminus A$ such that the attention to x leads to the consideration of some alternative in A.

Axiom 7—Association Reduction: For all $(A,B),(C,D) \in \mathcal{E}$, let $B_1 =$

$B \cap \mathcal{D}_\rho(A)$, $B_2 = B \setminus B_1$, $D_1 = D \cap \mathcal{D}_\rho(C)$ and $D_2 = D \setminus D_1$. If $B_1, D_1 \neq \emptyset$, then

$$\frac{\Phi_\rho(A, \emptyset) - \Phi_\rho(AB_1, B_2)}{\Phi_\rho(A, \emptyset) - \Phi_\rho(A, B)} = \frac{\Phi_\rho(C, \emptyset) - \Phi_\rho(CD_1, D_2)}{\Phi_\rho(C, \emptyset) - \Phi_\rho(C, D)}.$$

To understand Axiom 7, observe that the probability difference $\Phi_\rho(A, \emptyset) - \Phi_\rho(A, B)$ captures how the set of observable but unavailable alternatives B boosts the chance for the DM to end up with considering some available alternative. Now, consider the situation that we shift all alternatives in B that lead to the attention of A through one round of mental association, i.e., the set B_1, from the unavailable set to the available set. This reduces exactly one round of mental association for the DM's to pay attention to some available alternative. Hence, the ratio $\frac{\Phi_\rho(A,\emptyset)-\Phi_\rho(AB_1,B_2)}{\Phi_\rho(A,\emptyset)-\Phi_\rho(A,B)}$ is precisely the probability for the DM to have one more round of mental association, and thus is independent of the menu.

Although Axioms 6 and 7 are not direct relaxations of Axiom 2, we show in Lemmas 5 and 6 that Axioms 6 and 7 can be implied by Axioms 1-3.

To avoid triviality, we consider a non-trivial random choice rule ρ such that there exists menu (A, B) and $x \in B$ with $x \xrightarrow{B} A$.[20] The next theorem characterizes the ABCT and its uniqueness property.

Theorem 2. *A random choice rule ρ is an ABCT if and only if it satisfies Axioms 1 and 3-7. Moreover, if ρ is non-trivial, then there is a unique tuple $(\pi, \mathcal{N}, \succ, \eta)$ that represents ρ as an ABCT.*

We conclude this section with some discussion of new choice behavior that can be accommodated by ABCTs but not by ABCs. For any given ABC ρ, we argue that for all menus $(A, \emptyset)$ and $(B, \emptyset)$ with $A \subseteq B$, and all $x \in A$ and $y \in X \setminus B$, the following condition holds:

$$\rho(x|Ay, \emptyset) > \rho(x|A, \emptyset) \Rightarrow \rho(x|By, \emptyset) - \rho(x|B, \emptyset) \leq \rho(x|Ay, \emptyset) - \rho(x|A, \emptyset). \quad (9)$$

We refer to condition (9) as the *property of diminishing menu effects*: If y creates a menu effect on x (i.e., y boosts the choice frequency of x) in a smaller menu, then this effect would be weakened when more alternatives are included. To see

[20] If the random choice rule is trivial, then the DM never expands her initial consideration set through mental association. In such a case, the observable but unavailable set plays no role in affecting the DM's choices, and the ABCT model reduces to the model introduced by MM14.

why condition (9) holds, note that if y boosts the choice frequency of x, then the attention of y must prompt the consideration of x. Thus, $\rho(x|Ay, \emptyset) - \rho(x|A, \emptyset)$ is equal to the probability for y being initially considered multiplied by the probability of the event that neither x nor alternatives better than x are finally considered in $(A, \emptyset)$. Since the DM can include all associated alternatives into her consideration set, her final consideration set expands as the menu becomes larger, leading to a higher chance for it to include x or an alternative that is better than x. Hence, the menu effect should be weaker in a larger menu for an ABC.

By contrast, an ABCT may violate the property of diminishing menu effects: As the menu becomes larger, the "association distance" between y and x may decrease, meaning that it becomes easier for x to be associated with y. Consequently, y may boost the choice frequency of x to a larger extent in a larger menu. We demonstrate this observation through the following example.

Example 1. Let $A = \{x_1, ..., x_8, x\}$, $B = \{x_1, ..., x_8, x, z\}$ and $y \in X \setminus B$. Consider an ABCT ρ represented by $(\pi, \mathcal{N}, \succ, \eta)$, where π assigns each alternative an attention probability of $\frac{1}{2}$, $\mathcal{N} \setminus \mathcal{X} = \{(x_k, x_{k+1})\}_{k=1}^{7} \cup \{(x_8, x), (y, x_1), (y, z), (z, x)\}$, x is the $\succ$-best alternative in X, and $\eta = \frac{1}{2}$. It can be easily shown that $\rho(x|Ay, \emptyset) - \rho(x|A, \emptyset) = \frac{1}{2^{19}} < \frac{1}{2^{13}} < \rho(x|By, \emptyset) - \rho(x|B, \emptyset)$. The property of diminishing menu effects fails because in menu $(Ay, \emptyset)$, the DM needs 9 rounds of mental association to associate x with y, while only 2 rounds are needed in the menu $(By, \emptyset)$. $\square$

6 Application

In this section, we apply our model to the optimal design of associative networks to promote the sales of a given product. This problem is relevant for multi-product firms such as platforms like Amazon that match buyers and sellers, while also selling their own products.

Consider a multi-product firm that sells a set of products X. We assume that all products are always observable, but some products are sometimes out of stock. With this assumption, we consider a probability distribution κ over menus of the form $(A, X \setminus A)$, and let κ_A denote the probability of menu $(A, X \setminus A)$. The firm desires to maximize the sales volume of a particular alternative $x^* \in X$. We assume that the demand faced by the firm is captured by the ABC ρ represented by $(\pi, \mathcal{N}, \succ)$. That is, when the menu is (A, B), the sales volume of good x is given by $\rho(x|A, B)$.[21]

We consider the firm can add an additional association link to the existing network $\mathcal{N}$ to maximize the probability of the product being chosen. We briefly discuss the feasibility of adding such links and the rationale behind limiting it to a single additional link. This added link could be literal or metaphorical. For instance, platforms like Amazon might create an association from x to y by recommending y on the product page of x. Alternatively, firms could run advertisements that strengthen consumers' mental association from x to y. However, the firm faces constraints, such as limited space on product pages or the high costs of advertising, which prevent it from adding multiple links indiscriminately.

While our analysis below focuses on the optimal design of an associative network when only one additional link is to be added, our framework allows us to explore more general problems beyond this baseline application, such as the optimal associative network when multiple links are allowed to be added or when the firm is concerned with maximizing the total profits across all products. While all these applications are of potential interest, to formally deal with them is beyond the scope of the current paper, and we leave them for future research.

We define notations used in this section. Let $\overleftarrow{\mathcal{N}}$ be the inverse of $\mathcal{N}$ such that $x\mathcal{N}y$ if and only if $y\overleftarrow{\mathcal{N}}x$. Let $\overleftarrow{\mathcal{N}}^+$ be the transitive closure of $\overleftarrow{\mathcal{N}}$. For a given menu $(A, X \setminus A)$, let $\bar{A} = \{y \in X : y\mathcal{N}^+x^* \text{ or for some } z \in A, z \succ x^* \text{ and } y\mathcal{N}^+z\}$. In words, $\bar{A}$ contains alternatives such that the attention to any of them prompts the consideration of an available alternative that is weakly better than x^* in menu $(A, X \setminus A)$. The following proposition characterizes the optimal link to be added.

Proposition 4. *The association link (y, x^*) that satisfies the following condition is the link to be added that maximally increases the sales volume of x^*:*

$$y \in \arg\max_{z \in X} \left(\sum_{A \subseteq X :\, \rho(x^*|A, X\setminus A)>0} \kappa_A \mathring{\pi}_{\bar{A}} \left(1 - \mathring{\pi}_{\overleftarrow{\mathcal{N}}^+(z)\setminus\bar{A}}\right)\right).$$

For a given menu $(A, X \setminus A)$, if $\rho(x^*|A, X \setminus A) = 0$, then x^* is either unavailable or prompts the attention of some better alternative in A. In this case, x^* remains unchosen no matter what link we add. To understand Proposition 4, note that

[21]While we focus on ABCs, the analysis can be readily extended to the more general ABCTs.

for a given menu in which x^* is chosen, more alternatives in $\bar{A}$ indicate that it is more likely that either x^* is already considered or some available alternative that is better than x^* is considered. In either case, adding one more link does not affect the choice frequency of x^*. By contrast, if the set $\overleftarrow{\mathcal{N}}^+(z) \setminus \bar{A}$ is enlarged, the link is more likely to boost the consideration and the choice of x^*. Thus, the value of $\kappa_A \mathring{\pi}_{\bar{A}}$ can be regarded as the marginal benefit of boosting the attention of x^* in menu $(A, X \setminus A)$. The size of the menu $\overleftarrow{\mathcal{N}}^+(y) \setminus \bar{A}$ can be interpreted as the additional connectedness of x^* in menu $(A, X \setminus A)$ brought by the new link, and the value of $1 - \mathring{\pi}_{\overleftarrow{\mathcal{N}}^+(z)\setminus \bar{A}}$ captures how the attention of x^* can be boosted by the new link.

If the distribution of menus is not exogenous but can be determined by the firm, then the firm would choose the deterministic menu $(\{x^*\}, X \setminus x^*)$ to maximize the sales of product x^*. In this case, the optimal link to be added is given by the following corollary, which is an immediate implication of Proposition 4.

Corollary 1. *Suppose that the firm can choose the distribution of the menus. To maximize the sales volume of x^*, the firm can optimally choose menu $(\{x^*\}, X \setminus x^*)$ and add a link (y, x^*) such that $y \in \arg\min_{z \in X} \mathring{\pi}_{\overleftarrow{\mathcal{N}}^+(z)\setminus\overleftarrow{\mathcal{N}}^+(x^*)}$. In particular, if $\pi(\cdot) \equiv \alpha \in (0,1)$, then the link (y, x^*) satisfies $y \in \arg\max_{z \in X} |\overleftarrow{\mathcal{N}}^+(z) \setminus \overleftarrow{\mathcal{N}}^+(x^*)|$.*

When each product has the same chance to attract the attention of the consumers, the connectedness of alternative x^* is given by $\overleftarrow{\mathcal{N}}^+(x^*)$. The link we add is the simply one that increases the connectedness of x^* the most.

7 Extensions

In this section, we explore extensions of our baseline model to consider (i) limited data and (ii) random associative networks and preferences. The focus of our discussion is to investigate to what extent we can identify the parameters in these generalized models. The discussion of more general models of initial attention can be found in the Online Appendix.

7.1 Limited Data

In many applications, we are unable to observe the DM's choices in each possible menu. Instead, we may only observe the random choice rule ρ defined on a restricted collection of menus $\mathcal{E}' \subseteq \mathcal{E}$. In this section, we consider two special cases of restricted collections of menus. We first consider the case in which every observable alternative is available. Formally, $\mathcal{E}' = \mathcal{E}^F := \{(A, B) \in \mathcal{E} : B = \emptyset\}$. We then consider the case in which while some alternatives can be unavailable, all alternatives are observable. Formally, $\mathcal{E}' = \mathcal{E}^O := \{(A, B) \in \mathcal{E} : A \cup B = X\}$.

For any nonempty $\mathcal{E}' \subseteq \mathcal{E}$, we say that a random choice rule ρ is an ABC on $\mathcal{E}'$ if there is a tuple $(\pi, \mathcal{N}, \succ)$ such that for all $(A, B) \in \mathcal{E}'$ and $x \in A$, equation (1) holds for $\rho(x|A, B)$. The tuple $(\pi, \mathcal{N}, \succ)$ is said to represent ρ on $\mathcal{E}'$. Note that when $\mathcal{E}' = \mathcal{E}^F$, we are back to the standard random choice framework.

Case 1: $\mathcal{E}' = \mathcal{E}^F$. Consider an ABC ρ on $\mathcal{E}^F$ that is represented by $(\pi, \mathcal{N}, \succ)$. We show that π and $\succ$ can be uniquely identified. For all $x \in X$, $\pi(x) = \rho(x|\{x\}, \emptyset)$. For all $x, y \in X$, $x \succ y$ implies $\rho(x|\{x, y\}, \emptyset) \geq \pi_x = \rho(x|\{x\}, \emptyset)$ and $\rho(y|\{x, y\}, \emptyset) \leq \mathring{\pi}_x \pi_y < \rho(y|\{y\}, \emptyset)$. Therefore, $x \succ y$ if and only if

$$\rho(y|\{x, y\}, \emptyset) < \rho(y|\{y\}, \emptyset). \tag{10}$$

By contrast, the associative network $\mathcal{N}$ may not be uniquely identified. In what follows, we provide a partial identification of $\mathcal{N}$ by showing that we can identify the minimum associative network which is valid for representing the random choice rule on $\mathcal{E}^F$. We illustrate the idea of identification by the following examples.

Example 2. Consider two distinct alternatives x and y. If $\rho(y|\{x, y\}, \emptyset) > 0 = \rho(x|\{x, y\}, \emptyset)$, then any tuple $(\pi, \mathcal{N}, \succ)$ that represents ρ as an ABC on $\mathcal{E}^F$ must satisfy $x\mathcal{N}y$: Since x is never chosen in menu $(\{x, y\}, \emptyset)$, the attention to x must prompt the DM to consider some better alternative, which has to be y. □

Note that Example 2 also demonstrates that the associative network cannot be fully identified, since whether (y, x) is in the associative network is unclear and does not affect the DM's choice frequencies in this binary menu. The next example generalizes the identification strategy in Example 2.

Example 3. Consider alternatives x, y, z and w, and a random choice rule ρ:

(i) In menu $(\{x, y, z, w\}, \emptyset)$, only x is chosen with positive probability;

(ii) In menu $(\{x, y, w\}, \emptyset)$, only x and w are chosen with positive probability;

(iii) In menu $(\{x, y\}, \emptyset)$, only x is chosen with positive probability.

Condition (i) implies that x is better than y, z and w. By conditions (ii) and (iii), one can infer that x is directly associated with y, and neither y nor x is associated with w, since otherwise the attention to w would prompt the consideration of x and thus blocks the choice of w. Now, by adding z to menu $(\{x, y, w\}, \emptyset)$, w becomes unchosen. Hence, z must be directly associated with w. That is, for any tuple $(\pi, \mathcal{N}, \succ)$ that represents ρ as an ABC on $\mathcal{E}^F$, we must have $w\mathcal{N}z$. $\square$

The two examples above suggest the following identification of the associative network. For a given random choice rule ρ, define

$$\mathcal{N}[\rho] := \mathcal{X} \cup \Big\{(x, y) \in X^2 : x \neq y, \{y\} = \{w \in \{x, y\} : \rho(w|\{x, y\}, \emptyset) > 0\}\Big\} \cup$$
$$\Big\{(x, y) \in X^2 : \exists A \subseteq X, z \in X \setminus x \text{ such that } \{x, z\} = \{w \in Ax : \rho(w|Ax, \emptyset) > 0\}$$
$$\text{and } \{z\} = \{w \in A \cup \{x, y\} : \rho(w|A \cup \{x, y\}, \emptyset) > 0\} = \{w \in A : \rho(w|A, \emptyset) > 0\}\Big\}.$$

Note that for any set A and alternative z that satisfy the condition in the definition of $\mathcal{N}[\rho]$, it can be revealed that z is the best alternative in $A \cup \{x, y\}$, and that the attention of each alternative can prompt the consideration of z when the menu is either $(A \cup \{x, y\}, \emptyset)$ or $(A, \emptyset)$. Since x is chosen with a positive probability in menu $(Ax, \emptyset)$, x cannot prompt the consideration of z in this menu. It follows that no alternative in A is directly associated with x, and thus y must be directly associated with x so that the attention of x can lead to the consideration of z in menu $(A \cup \{x, y\}, \emptyset)$. Hence, the DM's associative network must include $\mathcal{N}[\rho]$ as a subset. Indeed, $\mathcal{N}[\rho]$ is exactly the minimal associative network such that $(\pi, \mathcal{N}[\rho], \succ)$ represents ρ as an ABC on $\mathcal{E}^F$.

Proposition 5. *For any random choice rule ρ that can be represented by $(\pi, \mathcal{N}, \succ)$ as an ABC on $\mathcal{E}^F$, π and $\succ$ can be uniquely identified. Furthermore, $\mathcal{N}[\rho] \subseteq \mathcal{N}$, and $(\pi, \mathcal{N}[\rho], \succ)$ also represents ρ as an ABC on $\mathcal{E}^F$.*

Since $\mathcal{E}^F$ is the standard choice domain considered by the literature, for completeness, we axiomatize ABCs on $\mathcal{E}^F$ in the Online Appendix.

Case 2: $\mathcal{E}' = \mathcal{E}^O$. Since for each menu $(A, B) \in \mathcal{E}^O$, the set of observable alternatives is fixed to be $X = A \cup B$, the DM's mental association is governed by the transitive closure of her associative network. That is, if a random choice rule ρ is represented by $(\pi, \mathcal{N}, \succ)$ as an ABC on $\mathcal{E}^O$, then ρ is also represented by $(\pi, \mathcal{N}^+, \succ)$ as an ABC on $\mathcal{E}^O$. Therefore, one immediate observation is that we cannot distinguish whether the DM's true associative network is $\mathcal{N}$ or $\mathcal{N}^+$. Nevertheless, the transitive closure $\mathcal{N}^+$ can be uniquely identified. To see this, consider two distinct alternatives x and y. Note that $x\mathcal{N}^+y$ is equivalent to $H_{\mathcal{N}}(\{y\}, X \setminus y) = H_{\mathcal{N}}(\{x,y\}, X \setminus \{x,y\})$. Thus, $x\mathcal{N}^+y$ if and only if

$$\Phi_\rho(\{y\}, X \setminus y) = \Phi_\rho(\{x,y\}, X \setminus \{x,y\}). \tag{11}$$

For the preference relation, we can show that $x \succ y$ if and only if $\rho(x|\{x\}, X \setminus x) = \rho(x|\{x,y\}, X \setminus \{x,y\})$.[22] Thus, $\succ$ can also be uniquely identified.

Proposition 6. *For any random choice rule ρ that can be represented by $(\pi, \mathcal{N}, \succ)$ as an ABC on $\mathcal{E}^O$, both $\mathcal{N}^+$ and $\succ$ can be uniquely identified. Furthermore, for all $A \subseteq X$ with $\mathcal{N}^+(X \setminus A) \cap A = \emptyset$, $\mathring{\pi}_A$ can be uniquely identified, and no further information regarding π can be obtained.*

By Proposition 6, the attention probability function π can be partially identified. Notably, whether π can be uniquely identified depends on $\mathcal{N}^+$. To see this, let $X = \{x_j\}_{j=1}^m$. Define $\mathcal{A}^\rho := \{A \subseteq X : A \neq \emptyset, \mathcal{N}^+(X \setminus A) \cap A = \emptyset\}$, and let $\mathcal{A}^\rho = \{A_i\}_{i=1}^n$. Consider an $n \times m$ matrix $\mathbb{M}^\rho$ such that for all $i \in \{1, ..., n\}$ and $j \in \{1, ..., m\}$, $\mathbb{M}^\rho_{i,j} = 1$ if $x_j \in A_i$, and $\mathbb{M}^\rho_{i,j} = 0$ if $x_j \notin A_i$. Consider an $m \times 1$ vector Π such that for all $j \in \{1, ..., m\}$, $\Pi_j = \ln(\mathring{\pi}_{x_j})$, and an $n \times 1$ vector $\mathbb{L}^\rho$ such that for all $i \in \{1, ..., n\}$, $\mathbb{L}^\rho_i = \ln(\mathring{\pi}_{A_i})$. The information for π we obtain from the random choice rule can be summarized by the following equation:

$$\mathbb{M}^\rho \cdot \Pi = \mathbb{L}^\rho. \tag{12}$$

The unique identification of π is equivalent to the unique solution of Π in equation (12). Therefore, we have the following proposition the proof of which is omitted.

Proposition 7. *For any random choice rule ρ that can be represented by $(\pi, \mathcal{N}, \succ)$ as an ABC on $\mathcal{E}^O$, π can be uniquely identified if and only if the rank of the matrix $\mathbb{M}^\rho$ equals $|X|$.*

[22]The necessity follows from the proof of the necessity of Axiom 5 for ABCs (Lemma 3). For sufficiency, note that if $y \succ x$, then $\rho(x|\{x\}, X \setminus x) < \rho(x|\{x,y\}, X \setminus \{x,y\})$.

Below, we provide one demonstration example.

Example 4. Let $X = \{x_1, x_2\}$. A random choice rule ρ satisfies $\rho(x_1|\{x_1, x_2\}, \emptyset) = \rho(x_1|\{x_1\}, \{x_2\}) = \frac{3}{4}$, $\rho(x_2|\{x_1, x_2\}, \emptyset) = 0$ and $\rho(x_2|\{x_2\}, \{x_1\}) = \frac{1}{2}$. Following our identification strategy, we have $x_1 \succ x_2$, $\mathcal{N}^+ = \mathcal{X} \cup \{(x_2, x_1)\}$ and $\mathcal{A}^\rho = \{\{x_2\}, \{x_1, x_2\}\}$. Let $A_1 = \{x_2\}$ and $A_2 = \{x_1, x_2\}$, and we have

$$\mathbb{M}^\rho = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$

which has rank 2. Therefore, π can be uniquely identified. Indeed, note that $\rho(x_2|\{x_2\}, \{x_1\}) = \frac{1}{2}$ implies that $\pi_{x_2} = \frac{1}{2}$. Combining it with $\rho(x_1|\{x_1\}, \{x_2\}) = \frac{3}{4}$ and $x_2 \mathcal{N} x_1$, we have $\pi_{x_1} = \frac{1}{2}$. $\square$

7.2 Random Association Based Consideration

In this section, we extend our analysis by introducing randomness to the associative network and the preference ordering.

Let $\mathscr{P}$ be the set of all possible preference orderings over X. A random preference is a probability distribution τ over $\mathscr{P}$. Let $\mathscr{N}$ be the set of all possible associative networks over X. A random associative network is a probability distribution μ over $\mathscr{N}$. A random associative network μ is said to be link-independent if there is a link-formation probability function (LPF) $\theta : X^2 \to [0, 1]$ such that (i) for all $z \in X$, $\theta(z, z) = 1$, and (ii) for all $\mathcal{N} \in \mathscr{N}$,

$$\mu(\mathcal{N}) = \left(\prod_{(x,y) \in \mathcal{N}} \theta(x, y) \right) \left(\prod_{(x,y) \notin \mathcal{N}} (1 - \theta(x, y)) \right)$$

The value of $\theta(x, y)$ is the probability that the DM can associate y with x. Condition (ii) indicates that the DM forms each association link independently.

We say that a random choice rule ρ is a **R**andom **A**ssociation **B**ased **C**onsideration rule (RABC) if there exists a tuple (π, μ, τ), where π is attention probability function, μ is a random associative network, and τ is a random preference, such that for all $(A, B) \in \mathcal{E}$ and $x \in A$,

$$\rho(x|A,B) = \sum_{\succ \in \mathscr{P}} \sum_{\mathcal{N} \in \mathscr{N}} \tau(\succ)\mu(\mathcal{N}) \left(\sum_{C \subseteq AB:\, x = \max(\mathcal{N}^{+}_{AB}(C) \cap A;\, \succ)} \pi_C \mathring{\pi}_{(AB)\setminus C} \right). \quad (13)$$

If in addition, μ is link-independent, then ρ is said to be a **R**andom **L**ink-independent **A**ssociation **B**ased **C**onsideration rule (RLABC) rule.

We regard RLABCs, which are special cases of RABCs, as a suitable starting point to investigate the joint effect of random attention, random association, and random preferences, since the model assumes that the formation of each directed association link, the realization of the preference, and the attention towards each alternative are all independent. Below, we show that parameters of RLABCs have nice identification properties.

For a given RLABC, the unique identification of the attention probability function π is similar to that in our baseline model. The LPF θ can also be uniquely identified as follows. For any two distinct alternatives x and y, we have

$$\Phi_\rho(\{y\}, \{x\}) = (1 - \pi_y)(1 - \pi_x \theta(x, y)),$$

i.e., the probability for y being unselected in menu $(\{y\}, \{x\})$ is equal to the probability that y is not initially paid attention to and not considered through the initial consideration of x. Thus, the probability for y being associated with x is

$$\theta(x, y) = \frac{1}{\pi_x} - \frac{\Phi_\rho(\{y\}, \{x\})}{\pi_x - \pi_x \pi_y}.$$

While the distribution of preferences τ may not be uniquely identified, for each nonempty $A \subseteq \mathcal{M}$ and each $x \in A$, we can pin down the probability that x is the best alternative in A under τ, i.e., we can identify

$$Z(x, A) := \sum_{\succ \in \mathscr{P}:\, x = \max(A;\succ)} \tau(\succ).$$

This is because with θ and π being identified, we can obtain the distribution of the final consideration sets of the DM in each menu. In particular, if for all $B \subsetneq A$,

$Z(\cdot, B)$ is known, then we can inductively derive $Z(\cdot, A)$ by the following formula

$$\forall x \in A,\ Z(x, A) = \frac{\rho(x|A, \emptyset) - \sum_{B \subsetneq A} \mathcal{F}(B, A) Z(x, B)}{1 - \sum_{B \subsetneq A} \mathcal{F}(B, A)},$$

where $\mathcal{F}(B, A)$ denotes the probability that B is the final consideration set in menu $(A, \emptyset)$. Note that the set of preference distributions that generate $\{Z(\cdot, A)\}_{A \subseteq X}$ can be partially identified, although not always uniquely, through the Block-Marschak polynomial introduced by Block and Marschak (1960).[23] Thus, introducing the channels of random attention and association does not make the identification of random preferences more complicated under our independence assumption.[24]

However, if we relax the assumption of link-independence for the random associative network by considering RABCs, then the uniqueness no longer holds. The following example illustrates that two distinct random associative networks can lead to the same random choice rule, even if the preference is deterministic.[25]

Example 5. Let $X = \{x, y\}$. Consider the following random choice rule ρ.

$$\rho(y|\{x, y\}, \emptyset) = \frac{1}{8},\ \rho(x|\{x, y\}, \emptyset) = \rho(x|\{x\}, \{y\}) = \rho(y|\{y\}, \{x\}) = \frac{5}{8},$$
$$\rho(x|\{x\}, \emptyset) = \rho(y|\{y\}, \emptyset) = \frac{1}{2}.$$

It can be revealed that DM's random preference τ satisfies $\tau(\succ) = 1$, where $\succ$ satisfies $x \succ y$, and her attention probability function is given by $\pi_x = \pi_y = \frac{1}{2}$. However, the random choice rule can be represented by more than one random associative networks: Any random associative network μ can represent the random choice rule if the probability of y being associated with x and that of x being

[23] Falmagne (1978) shows that with $\{Z(\cdot, A)\}_{A \subseteq X}$, for all $x \in X$ and partition $\{D, D'\}$ of $X \setminus x$, we can uniquely identify the probability for x being better than all alternatives in D and worse than all alternatives in D'. Any distribution over preference orderings that induces the above probabilities is consistent with $\{Z(\cdot, A)\}_{A \subseteq X}$.

[24] If τ is deterministic, i.e., $\tau(\succ) = 1$ for some $\succ \in \mathscr{P}$, then τ can be uniquely identified. The unique identification property also holds for single-crossing random preferences (Apesteguia et al., 2017; Barseghyan et al., 2021b). Another special class of random preferences is the Luce model, where there is a weight function $\omega : X \to \mathbb{R}_+$ such that for all $x \in A$, $Z(x, A) = \frac{\omega(x)}{\sum_{y \in A} \omega(y)}$. Clearly, since $Z(\cdot, \cdot)$ can be fully identified, ω can also be uniquely identified up to rescaling. A similar observation is made in Cheung and Masatlioglu (2024).

[25] Although it is an intriguing open question to explore when two different random associative networks lead to the same random choice rule, this question is beyond the scope of the current paper and is left for future research.

associated with y are both equal to $\frac{1}{2}$ under μ. Thus, we can consider four (deterministic) associative networks $\mathcal{N} = \mathcal{X}$, $\mathcal{W} = \mathcal{X} \cup \{(x, y)\}$, $\mathcal{V} = \mathcal{X} \cup \{(y, x)\}$, and $\mathcal{U} = \mathcal{X} \cup \{(x, y), (y, x)\}$, and the two random associative networks μ_1 and μ_2, with $\mu_1(\mathcal{N}) = \mu_1(\mathcal{U}) = \frac{1}{2}$ and $\mu_2(\mathcal{W}) = \mu_2(\mathcal{V}) = \frac{1}{2}$, can both represent ρ. $\square$

A Appendix

Proof of Proposition 1. If $\rho(x|A, B) > 0$, then there exists $C \subseteq AB$ such that $x = \max(\mathcal{N}^+_{AB}(C) \cap A; \succ)$. Since $x \in \mathcal{N}^+_{AB}(C)$, we have $\mathcal{N}^+_{AB}(x) \subseteq \mathcal{N}^+_{AB}(C)$, and thus $x = \max(\mathcal{N}^+_{AB}(x) \cap A; \succ)$. Inversely, if $x = \max(\mathcal{N}^+_{AB}(x) \cap A; \succ)$, then x is chosen when the initial consideration set is $\{x\}$. Thus, $\rho(x|A, B) \geq \pi_x \mathring{\pi}_{(AB)\setminus x} > 0$. $\square$

Proof of Proposition 2. For equation (2), note that for any initial consideration set $C \subseteq AB$, we have $\mathcal{N}^+_{AB}(C) \cap A \neq \emptyset$ if and only if $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$ for some $x \in C$. Thus, the default option is chosen if and only if any alternative x that satisfies $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$ is not initially considered. This leads to equation (2).

For equation (3), we just need to show that it holds for each $k \in \{2, ..., n\}$. When the initial consideration set is D, x_k is chosen if and only if $\mathcal{N}^+_{AB}(D) \cap \{x_1, ..., x_k\} = \{x_k\}$. That is, $D \cap \{y \in AB : x_k = \max(\mathcal{N}^+_{AB}(y) \cap A; \succ)\} \neq \emptyset$ and for all $m \leq k-1$, $D \cap \{y \in AB : x_m = \max(\mathcal{N}^+_{AB}(y) \cap A; \succ)\} = \emptyset$. This leads to equation (3). $\square$

Proof of Theorems 1 and 2. **(Reformulation)** Prior to the proof, we provide a reformulation of the choice frequency of the default option. Consider an ABCT ρ that is represented by $(\pi, \mathcal{N}, \succ, \eta)$. For any given menu (A, B), define $H^0_{\mathcal{N}}(A, B) = A$. Then, we can define inductively $H^k_{\mathcal{N}}(A, B)$ for every $k \in \mathbb{N}_+$ such that

$$H^k_{\mathcal{N}}(A, B) = \{x \in AB : x\mathcal{N}y \text{ for some } y \in H^{k-1}_{\mathcal{N}}(A, B)\}.$$

To interpret, the set $H^k_{\mathcal{N}}(A, B)$ contains all alternatives in AB which can direct the DM's attention to some available alternative in A through no more than k rounds of mental association. We reformulate $\Phi_\rho(A, B)$ as:

$$\Phi_\rho(A, B) = \sum_{k=0}^{+\infty} (1-\eta)\eta^k \mathring{\pi}_{H^k_{\mathcal{N}}(A,B)} + \left(1 - \left(\sum_{k=0}^{+\infty} (1-\eta)\eta^k\right)\right) \mathring{\pi}_{H_{\mathcal{N}}(A,B)}.$$

Note that $\{H^k_{\mathcal{N}}(A,B)\}_{k=0}^{+\infty}$ is an increasing sequence of sets, and there exists $m \in \mathbb{N}_+$ such that for all $k \geq m$, $H^k_{\mathcal{N}}(A,B) = H_{\mathcal{N}}(A,B)$. For any such m, we also have

$$\Phi_\rho(A,B) = \sum_{k=0}^{m-1} (1-\eta)\eta^k \mathring{\pi}_{H^k_{\mathcal{N}}(A,B)} + \eta^m \mathring{\pi}_{H^m_{\mathcal{N}}(A,B)}.$$

(Necessity) Since an ABC is a special ABCT, we first consider an ABCT ρ and show that it satisfies Axioms 1 and 3-7. We then show that if ρ is an ABC, then it satisfies Axiom 2.

Lemma 1. *If a random choice rule ρ is an ABCT, then it satisfies Axiom 1.*

Proof of Lemma 1. Let ρ be represented by $(\pi, \mathcal{N}, \succ, \eta)$. For all $A \in \mathcal{M}$, $\Phi_\rho(A, \emptyset) = \mathring{\pi}_A$. Thus, for all $x \in A$, $\frac{\Phi_\rho(A,\emptyset)}{\Phi_\rho(A\setminus x,\emptyset)} = \mathring{\pi}_x$. □

Lemma 2. *If a random choice rule ρ is represented by $(\pi, \mathcal{N}, \succ, \eta)$ as an ABCT, then for all $(A,B) \in \mathcal{E}$ and $x \in AB$, $x \xrightarrow{B} A$ if and only if $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$.*

Proof of Lemma 2. The case for $x \in A$ is trivial. Consider the case where $x \in B$. For necessity, suppose that $\mathcal{N}^+_{AB}(x) \cap A = \emptyset$. By independent attention, we have $\Phi_\rho(A,B) = \mathring{\pi}_x \Phi_\rho(A, B\setminus x) + \pi_x \Phi_\rho(A, B|x)$, where $\Phi_\rho(A,B|x)$ denotes the probability that the default option is chosen in menu (A,B) conditioning on that x is initially paid attention to. However, note that for any initial consideration set $C \subseteq AB$ such that $x \notin C$. We have $\mathcal{N}^+_{AB}(C) \cap A \neq \emptyset$ if and only if $\mathcal{N}^+_{AB}(Cx) \cap A \neq \emptyset$, and for all $k \in \mathbb{N}$, $\mathcal{N}^k_{AB}(C) \cap A \neq \emptyset$ if and only if $\mathcal{N}^k_{AB}(Cx) \cap A \neq \emptyset$. Therefore, $\Phi_\rho(A,B|x) = \Phi_\rho(A, B \setminus x)$, and we have $\Phi_\rho(A,B) = \Phi_\rho(A, B\setminus x)$.

For sufficiency, suppose that $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$. Consider an arbitrary set $C \subseteq (AB) \setminus x$. We have $\pi_C \mathring{\pi}_{(AB)\setminus(Cx)} = \pi_C \mathring{\pi}_{(AB)\setminus C} + \pi_{Cx} \mathring{\pi}_{(AB)\setminus(Cx)}$. Note that $\mathcal{N}^+_{(AB)\setminus x}(C) \cap A \neq \emptyset$ implies $\mathcal{N}^+_{AB}(C) \cap A \neq \emptyset$ and $\mathcal{N}^+_{AB}(Cx) \cap A \neq \emptyset$, and for all $k \in \mathbb{N}$, $\mathcal{N}^k_{(AB)\setminus x}(C) \cap A \neq \emptyset$ implies $\mathcal{N}^k_{AB}(C) \cap A \neq \emptyset$ and $\mathcal{N}^k_{AB}(Cx) \cap A \neq \emptyset$. Thus, we have $\Phi_\rho(A,B) \leq \Phi_\rho(A, B\setminus x)$, and to show $\Phi_\rho(A,B) \neq \Phi_\rho(A, B\setminus x)$, we need to show $\Phi_\rho(A,B) < \Phi_\rho(A, B\setminus x)$. Since every $C \subseteq (AB)\setminus x$ has a positive probability to be initially considered, it suffices to show that there exists $C \subseteq (AB) \setminus x$ such that $\mathcal{N}^+_{(AB)\setminus x}(C) \cap A = \emptyset$ and $\mathcal{N}^+_{AB}(Cx) \cap A \neq \emptyset$. By taking $C = \emptyset$, we are done. □

Lemma 3. *An ABCT ρ satisfies Axioms 3-7.*

Proof of Lemma 3. Let ρ be represented by $(\pi, \mathcal{N}, \succ, \eta)$. For Axiom 3, consider two menus (A, B) and (C, D) with $C \subseteq A$ and $CD \subseteq AB$. Since $x \xrightarrow{D} C$, either (1) $x \in C$ or (2) $\Phi_\rho(C, D) \neq \Phi_\rho(C, D \setminus x)$. Case (1) directly implies $x \in A$ since $C \subseteq A$, and thus we have $x \xrightarrow{B} A$. Case (2) implies $\mathcal{N}^+_{CD}(x) \cap C \neq \emptyset$ by Lemma 2, and thus $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$. It follows that either $x \in A$ or $x \in B$. The former case directly implies $x \xrightarrow{B} A$, and the latter case implies $x \xrightarrow{B} A$ by Lemma 2.

For Axiom 4, consider $x, y \in X$ and $(A, B) \in \mathcal{E}$ such that $x \neq y$, $x \xrightarrow{B} A$, and not $x \xrightarrow{B \setminus y} A \setminus y$. It follows that $x \in B$, $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$ and $\mathcal{N}^+_{(AB)\setminus y}(x) \cap (A \setminus y) = \emptyset$. Since $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$, there is a sequence $(x_k)_{k=1}^n$ in AB such that $x_1 = x$, $\{x_1, ..., x_n\} \cap A = x_n$, and for all $k \in \{1, ..., n-1\}$, $x_k \mathcal{N} x_{k+1}$. However, since $\mathcal{N}^+_{(AB)\setminus y}(x) \cap (A \setminus y) = \emptyset$, for all such sequence, there exists $k \in \{2, ..., n\}$ such that $x_k = y$. Therefore, we have $x \xrightarrow{B \setminus y} \{y\}$ and $y \xrightarrow{B \setminus x} A$.

For Axiom 5, it suffices to show that if $x \succ y$, then for all $(A, B) \in \mathcal{E}$ with $x, y \in A$, we have $\rho(x|A, B) = \rho(x|A \setminus y, By)$. Note that since $AB = (A \setminus y) \cup (By)$, we have for all $C \subseteq AB$ and $k \in \mathbb{N}$, $x \succ y$ implies that $x = \max(\mathcal{N}^k_{AB}(C) \cap A; \succ)$ if and only if $x = \max(\mathcal{N}^k_{(A\setminus y)\cup(By)}(C) \cap (A \setminus y); \succ)$. It then follows from the definition of ABCT (equation (8)) that $\rho(x|A, B) = \rho(x|A \setminus y, By)$.

For Axiom 6, note that in the proof of Lemma 2, we have shown that for all menu (A, B) and $x \in B$, if $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$, then $\Phi_\rho(A, B) < \Phi_\rho(A, B \setminus x)$. If $\mathcal{N}^+_{AB}(x) \cap A = \emptyset$, then by Lemma 2, not $x \xrightarrow{B} A$, and thus $\Phi_\rho(A, B) = \Phi_\rho(A, B \setminus x)$. In both cases, we have $\Phi_\rho(A, B) \leq \Phi_\rho(A, B \setminus x)$. It remains to show that $\Phi_\rho(Ax, B \setminus x) \leq \Phi_\rho(A, B)$. Note that for all $C \subseteq AB = (Ax) \cup (B \setminus x)$, $\mathcal{N}^+_{AB}(C) \cap A \neq \emptyset$ implies $\mathcal{N}^+_{(Ax)\cup(B\setminus x)}(C) \cap (Ax) = \mathcal{N}^+_{AB}(C) \cap (Ax) \neq \emptyset$, and for all $k \in \mathbb{N}$, $\mathcal{N}^k_{AB}(C) \cap A \neq \emptyset$ implies $\mathcal{N}^k_{(Ax)\cup(B\setminus x)}(C) \cap (Ax) = \mathcal{N}^k_{AB}(C) \cap (Ax) \neq \emptyset$. Thus, Axiom 6 holds.

For Axiom 7, note that since $B_1 = \{x \in B : x \xrightarrow{\{x\}} A\}$, by Lemma 2, we have $x \in B_1$ if and only if $\mathcal{N}^+_{Ax}(x) \cap A \neq \emptyset$, which further implies that $B_1 = \{x \in B : x \mathcal{N} y \text{ for some } y \in A\}$. It then follows that for all $k \in \mathbb{N}$, $H^k_{\mathcal{N}}(AB_1, B \setminus B_1) = H^{k+1}_{\mathcal{N}}(A, B)$. Pick $m \in \mathbb{N}_+$ with $m \geq 3$ such that for all $k \geq m-1$,

$H^k_{\mathcal{N}}(A,B) = H_{\mathcal{N}}(A,B)$ and $H^k_{\mathcal{N}}(AB_1, B\setminus B_1) = H_{\mathcal{N}}(AB_1, B\setminus B_1)$. We have

$$\begin{aligned}
\Phi_\rho(A,\emptyset) - \Phi_\rho(A,B) &= \mathring{\pi}_A - \sum_{k=0}^{m-1}(1-\eta)\eta^k \mathring{\pi}_{H^k_{\mathcal{N}}(A,B)} - \eta^m \mathring{\pi}_{H^m_{\mathcal{N}}(A,B)} \\
&= \eta\mathring{\pi}_A - \sum_{k=1}^{m-1}(1-\eta)\eta^k \mathring{\pi}_{H^k_{\mathcal{N}}(A,B)} - \eta^m \mathring{\pi}_{H^m_{\mathcal{N}}(A,B)} \\
&= \eta\left(\mathring{\pi}_A - \sum_{k=0}^{m-2}(1-\eta)\eta^k \mathring{\pi}_{H^k_{\mathcal{N}}(AB_1,B\setminus B_1)} - \eta^{m-1} \mathring{\pi}_{H^{m-1}_{\mathcal{N}}(AB_1,B\setminus B_1)}\right) \\
&= \eta\Phi_\rho(A,\emptyset) - \eta\Phi_\rho(AB_1, B\setminus B_1).
\end{aligned}$$

Since $B_1 \neq \emptyset$, we have $\Phi_\rho(A,\emptyset) > \Phi_\rho(AB_1, B\setminus B_1)$. Therefore, we have

$$\frac{\Phi_\rho(A,\emptyset) - \Phi_\rho(AB_1, B\setminus B_1)}{\Phi_\rho(A,\emptyset) - \Phi_\rho(A,B)} = 1/\eta.$$

Hence, Axiom 7 holds for an ABCT. □

Lemma 4. *An ABC ρ satisfies Axiom 2.*

Proof of Lemma 4. Let ρ be represented by $(\pi, \mathcal{N}, \succ)$. For a given menu (A,B) and $x \in B$, we have $H_{\mathcal{N}}(A,B) \subseteq H_{\mathcal{N}}(Ax, B\setminus x)$. To show Axiom 2, it suffices to show that if $x \xrightarrow{B} A$, then we have $H_{\mathcal{N}}(Ax, B\setminus x) \subseteq H_{\mathcal{N}}(A,B)$. To see this, consider $y \in H_{\mathcal{N}}(Ax, B\setminus x)$. It follows that $\mathcal{N}^+_{AB}(y) \cap (Ax) \neq \emptyset$. If $\mathcal{N}^+_{AB}(y) \cap A \neq \emptyset$, then $y \in H_{\mathcal{N}}(A,B)$. If $x \in \mathcal{N}^+_{AB}(y)$, then $\mathcal{N}^+_{AB}(x) \subseteq \mathcal{N}^+_{AB}(y)$. Since $x \xrightarrow{B} A$, by Lemma 2, we have $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$, and thus $\mathcal{N}^+_{AB}(y) \cap A \neq \emptyset$. Hence, $y \in H_{\mathcal{N}}(A,B)$. □

(Sufficiency) Consider a random choice rule ρ. Throughout the proof of sufficiency, we define π, $\mathcal{N}$ and $\succ$ as follows. Let the attention probability π be such that for all $x \in X$, $\pi_x = \rho(x|\{x\},\emptyset) \in (0,1)$. Define the associative network $\mathcal{N}$ such that

$$\mathcal{N} = \mathcal{X} \cup \{(x,y) \in X^2 : x \neq y \text{ and } \rho(y|\{y\},\emptyset) \neq \rho(y|\{y\},\{x\})\}.$$

Define the binary relation $\succ$ over X such that for any two distinct alternatives x and y, $x \succ y$ if and only if $\rho(y|\{x,y\},\emptyset) \neq \rho(y|\{y\},\{x\})$. Note that if Axiom 5 holds, then $x \succ y$ implies that for all menu (A,B) with $x,y \in A$, $\rho(x|A,B) = \rho(x|A\setminus y, By)$.

Lemma 5. *For any random choice rule ρ, Axioms 1-3 imply Axiom 6.*

Proof of Lemma 5. Consider a random choice rule ρ that satisfies Axioms 1-3. Let (A, B) and (C, D) be two arbitrary menus such that $C \subseteq A$ and $CD \subseteq AB$. For any $x \in B$ such that $\Phi_\rho(A, B) = \Phi_\rho(A, B \setminus x)$, i.e., not $x \xrightarrow{B} A$, by Axiom 3, we have not $x \xrightarrow{D} C$, i.e., $\Phi_\rho(C, D) = \Phi_\rho(C, D \setminus x)$. Therefore, we can consecutively delete those alternatives in B to obtain $\hat{B} \subseteq B$ such that $\Phi_\rho(A, B) = \Phi_\rho(A, \hat{B})$ and $\Phi_\rho(C, D) = \Phi_\rho(C, \hat{D})$ where $\hat{D} = D \setminus (B \setminus \hat{B})$. Now, for menu $(A, \hat{B})$, we have for all $x \in \hat{B}$, $x \xrightarrow{\hat{B}} A$. By Axioms 2 and 3, we can consecutively shifting alternatives from $\hat{B}$ to A without affect the choice frequency of the default option, that is, $\Phi_\rho(A, B) = \Phi_\rho(A, \hat{B}) = \Phi_\rho(A\hat{B}, \emptyset)$. By a similar argument, we can show that there exists some subset $\bar{D} \subseteq \hat{D}$ such that $\Phi_\rho(C, D) = \Phi_\rho(C, \hat{D}) = \Phi_\rho(C\bar{D}, \emptyset)$. By Axiom 1, we have $\Phi_\rho(A, B) = \mathring{\pi}_{A\hat{B}}$ and $\Phi_\rho(C, D) = \mathring{\pi}_{C\bar{D}}$. Since $C\bar{D} \subseteq C\hat{D} \subseteq A\hat{B}$, we have $\Phi_\rho(A, B) \leq \Phi_\rho(C, D)$, i.e., Axiom 6 holds. □

Lemma 6. *For any random choice rule ρ, Axioms 1-3 imply Axiom 7.*

Proof of Lemma 6. Consider menu (A, B) and nonempty $B_1 = \{x \in B : x \xrightarrow{\{x\}} A\}$. By Axioms 2 and 3, we have $\Phi_\rho(A, B) = \Phi_\rho(AB_1, B_2)$ where $B_2 = B \setminus B_1$. Following a similar argument as the proof of Lemma 5, we have $\Phi_\rho(AB_1, B_2) = \Phi_\rho(AB_1G, \emptyset)$ for some $G \subseteq B_2$. Thus, by Axiom 1, we have $\Phi_\rho(A, \emptyset) - \Phi_\rho(AB_1, B_2) \neq \emptyset$. Therefore, $\frac{\Phi_\rho(A,\emptyset)-\Phi_\rho(A,B)}{\Phi_\rho(A,\emptyset)-\Phi_\rho(AB_1,B_2)}$ is constantly equal to 1, and Axiom 7 holds. □

Lemma 7. *If a random choice rule ρ satisfies Axiom 4, then for all $(A, B) \in \mathcal{E}$ and $x \in B$, $x \xrightarrow{B} A$ implies $\mathcal{N}^+_{AB}(x) \cap A \neq \emptyset$.*

Proof of Lemma 7. Consider $(A, B) \in \mathcal{E}$ and $x \in B$ such that $x \xrightarrow{B} A$. It follows that $A \neq \emptyset$, since otherwise we have $\Phi_\rho(A, B) = \Phi_\rho(A, B \setminus x) = 1$. Since not $x \xrightarrow{B} \emptyset$, by Axiom 4 and a simple induction, there exists $y \in A$ such that $x \xrightarrow{B} \{y\}$. It then suffices to show $y \in \mathcal{N}^+_{By}(x)$, and we show this by induction.

First, if $|B| = 1$, then $B = \{x\}$. In this case, $x \xrightarrow{B} \{y\}$ is equivalent to $x \xrightarrow{\{x\}} \{y\}$, which further implies $\Phi_\rho(\{y\}, \{x\}) \neq \Phi_\rho(\{y\}, \emptyset)$. Since only y is available in menus $(\{y\}, \{x\})$ and $(\{y\}, \emptyset)$, we conclude that $\rho(y|\{y\}, \{x\}) \neq \rho(y|\{y\}, \emptyset)$. By the definition of $\mathcal{N}$, we have $x\mathcal{N}y$. Thus, we have $y \in \mathcal{N}^+_{\{x,y\}}(x) = \mathcal{N}^+_{By}(x)$.

Next, assume by induction that when $|B| \leq n$, $x \xrightarrow{B} \{y\}$ implies $y \in \mathcal{N}^+_{By}(x)$. We want to show that when $|B| = n + 1$, $x \xrightarrow{B} \{y\}$ also implies $y \in \mathcal{N}^+_{By}(x)$. To see this, note that if there exists $z \in B \setminus x$ such that $x \xrightarrow{B \setminus z} \{y\}$, then by the induction hypothesis, $y \in \mathcal{N}^+_{(By)\setminus z}(x) \subseteq \mathcal{N}^+_{By}(x)$. Otherwise, for all $z \in B \setminus x$, $x \xrightarrow{B\setminus z} \{y\}$ does not hold. It then follows from Axiom 4 that for all $z \in B \setminus x$, $x \xrightarrow{B\setminus z} \{z\}$ and $z \xrightarrow{B\setminus x} \{y\}$. By our induction hypothesis, we have for all $z \in B \setminus x$, $z \in \mathcal{N}^+_B(x) \subseteq \mathcal{N}^+_{By}(x)$ and $y \in \mathcal{N}^+_{(By)\setminus x}(z) \subseteq \mathcal{N}^+_{By}(z)$. By the transitivity of $\mathcal{N}^+_{By}$, we have $y \in \mathcal{N}^+_{By}(x)$. □

Lemma 8. *If a random choice rule ρ satisfies Axiom 3, then for all $x \in X$ and $(A, B) \in \mathcal{E}$ with $x \in B$, $\mathcal{N}(x) \cap A \neq \emptyset$ implies $x \xrightarrow{B} A$.*

Proof of Lemma 8. Consider $x \in B$ with $\mathcal{N}(x) \cap A \neq \emptyset$. It then follows that there exists $y \in A$ such that $x\mathcal{N}y$, i.e., $x \xrightarrow{\{x\}} \{y\}$. By Axiom 3, we have $x \xrightarrow{B} A$. □

Lemma 9. *If a random choice rule ρ satisfies Axioms 3 and 4, then for all $x \in X$ and $A \in \mathcal{M}$ with $x \notin A$, $x \in \mathcal{D}_\rho(A)$ if and only if $x\mathcal{N}y$ for some $y \in A$.*

Proof of Lemma 9. Suppose that $x \in \mathcal{D}_\rho(A)$. By Lemma 7, we have $\mathcal{N}^+_{Ax}(x) \cap A \neq \emptyset$. It follows that there is $y \in A$ such that $x\mathcal{N}y$. Inversely, suppose that $x\mathcal{N}y$ for some $y \in A$, i.e., $\mathcal{N}(x) \cap A \neq \emptyset$. By Lemma 8, we have $x \in \mathcal{D}_\rho(A)$. □

Lemma 10. *If a random choice rule ρ satisfies Axioms 1, 3, 4, 6 and 7, then there exists $\eta \in (0, 1]$ such that for all $(A, B) \in \mathcal{E}$,*

$$\Phi_\rho(A, B) = \sum_{k=0}^{+\infty} (1-\eta)\eta^k \mathring{\pi}_{H^k_{\mathcal{N}}(A,B)} + \left(1 - \left(\sum_{k=0}^{+\infty}(1-\eta)\eta^k\right)\right) \mathring{\pi}_{H_{\mathcal{N}}(A,B)}; \tag{14}$$

if ρ additionally satisfies Axiom 2, then $\eta = 1$ in equation (14).

Proof of Lemma 10. If for all $(A, B) \in \mathcal{E}$ and all $x \in B$, not $x \xrightarrow{B} A$, then we have $\mathcal{N} = \mathcal{X}$ and $\Phi_\rho(A, B) = \Phi_\rho(A, \emptyset) = \mathring{\pi}_A$. Equation (16) holds for any $\eta \in (0, 1]$ since $H^k_{\mathcal{N}}(A, B) = H_{\mathcal{N}}(A, B) = A$ for all $k \in \mathbb{N}_+$.

Consider the non-trivial situation in which there exists a menu (A^*, B^*) and $x \in B^*$ such that $x \xrightarrow{B^*} A^*$. By Lemma 7, $\mathcal{N}^+_{A^*B^*}(x) \cap A^* \neq \emptyset$. It follows that the set $B^*_1 = B^* \cap \mathcal{D}_\rho(A^*)$ is not empty. Let $B^*_2 = B^* \setminus B^*_1$. By Axiom 6, we have $\Phi_\rho(A^*B^*_1, B^*_2) \leq \Phi_\rho(A^*, B^*) \leq \Phi_\rho(A^*, B^* \setminus x) \leq \Phi_\rho(A^*, \emptyset)$. Since $\Phi_\rho(A^*, B^*) \neq \Phi_\rho(A^*, B^* \setminus x)$, we have $\Phi_\rho(A^*B^*_1, B^*_2) \leq \Phi_\rho(A^*, B^*) < \Phi_\rho(A^*, B^* \setminus x) \leq \Phi_\rho(A^*, \emptyset)$. Thus, we define

$$\eta := \frac{\Phi_\rho(A^*, \emptyset) - \Phi_\rho(A^*, B^*)}{\Phi_\rho(A^*, \emptyset) - \Phi_\rho(A^*B^*_1, B^*_2)} \in (0, 1]. \tag{15}$$

Note that if Axiom 2 additionally holds, then $\Phi_\rho(A^*, B^*) = \Phi_\rho(A^*B_1^*, B_2^*)$. To see this, note that for all $\hat{A} \subseteq A^*B^*$, if $A^* \subseteq \hat{A}$ and $x \in (A^*B^*) \setminus \hat{A}$, by Axiom 3, $x \xrightarrow{B^*} A^*$ implies $x \xrightarrow{(A^*B^*)\setminus\hat{A}} \hat{A}$. Thus, by Axiom 2, we can consecutively move alternatives from B_1^* to the available set without affecting the choice frequency of the default option $\Phi_\rho(\cdot,\cdot)$. Therefore, we have $\Phi_\rho(A^*, B^*) = \Phi_\rho(A^*B_1^*, B_2^*)$. That is, when Axiom 2 additionally holds, $\eta = 1$.

To proceed, consider an arbitrary menu (A, B). By Lemma 9, we have for all $k \in \mathbb{N}$,

$$\mathcal{D}_\rho(H_\mathcal{N}^k(A, B)) \cap (AB) = H_\mathcal{N}^{k+1}(A, B) \setminus H_\mathcal{N}^k(A, B).$$

Note that there exists $m \in \mathbb{N}$ such that for all $k \geq m$, $H_\mathcal{N}^k(A, B) = H_\mathcal{N}(A, B)$, and for all $k < m$, $H_\mathcal{N}^k(A, B) \subsetneq H_\mathcal{N}^{k+1}(A, B)$. By Axiom 7, for all $k < m$, we have

$$\frac{\Phi_\rho(H_\mathcal{N}^k(A, B), \emptyset) - \Phi_\rho(H_\mathcal{N}^k(A, B), B \setminus H_\mathcal{N}^k(A, B))}{\Phi_\rho(H_\mathcal{N}^k(A, B), \emptyset) - \Phi_\rho(H_\mathcal{N}^{k+1}(A, B), B \setminus H_\mathcal{N}^{k+1}(A, B))} = \eta. \tag{16}$$

Since for all $x \in B \setminus H_\mathcal{N}(A, B)$, $\mathcal{N}(x) \cap H_\mathcal{N}(A, B) = \emptyset$, by Lemma 9, we have for all $x \in B \setminus H_\mathcal{N}(A, B)$, $x \notin \mathcal{D}_\rho(H_\mathcal{N}(A, B))$. Thus, for all $x \in B \setminus H_\mathcal{N}(A, B)$, $\mathcal{N}_{AB}^+(x) \cap H_\mathcal{N}(A, B) = \emptyset$. It then follows that for all $x \in B \setminus H_\mathcal{N}(A, B)$, not $x \xrightarrow{B\setminus H_\mathcal{N}(A,B)} H_\mathcal{N}(A, B)$. By Axiom 3, for all $C \subseteq B \setminus H_\mathcal{N}(A, B)$, for all $x \in C$, not $x \xrightarrow{C} H_\mathcal{N}(A, B)$. Therefore, we can consecutively delete alternatives in $B \setminus H_\mathcal{N}(A, B)$ and have $\Phi_\rho(H_\mathcal{N}(A, B), B \setminus H_\mathcal{N}(A, B)) = \Phi_\rho(H_\mathcal{N}(A, B), \emptyset)$.

Note that by Axiom 1, for all $D \in \mathcal{M}$, we have $\Phi_\rho(D, \emptyset) = \mathring{\pi}_D$. Combining it with the fact that $\Phi_\rho(H_\mathcal{N}^m(A, B), B \setminus H_\mathcal{N}(A, B)) = \Phi_\rho(H_\mathcal{N}(A, B), B \setminus H_\mathcal{N}(A, B)) = \Phi_\rho(H_\mathcal{N}(A, B), \emptyset)$, we can use equation (16) to inductively derive the following:

$$\Phi_\rho(A, B) = \sum_{k=0}^{m-1} (1 - \eta)\eta^k \mathring{\pi}_{H_\mathcal{N}^k(A,B)} + \eta^m \mathring{\pi}_{H_\mathcal{N}^m(A,B)}.$$

One can easily verify that the equation above is the same as equation (14) since for all $k \geq m$, $H_\mathcal{N}^k(A, B) = H_\mathcal{N}(A, B)$. □

In the remaining part of the proof, we fix η to be the one in Lemma 10.

Lemma 11. *If the random choice rule ρ satisfies Axioms 1 and 3-7, then the revealed preference relation $\succ$ is asymmetric and satisfies that for all distinct $x, y \in X$, either $x \succ y$ or $y \succ x$.*

Proof of Lemma 11. By the definition of $\succ$ and Axiom 5, if $x \succ y$, then not $y \succ x$. It remains to show that $\succ$ is well-defined for all distinct $x, y \in X$. Suppose to the contrary that not $x \succ y$ and not $y \succ x$. Then we have $\rho(x|\{x\},\{y\}) = \rho(x|\{x,y\},\emptyset)$ and $\rho(y|\{y\},\{x\}) = \rho(y|\{x,y\},\emptyset)$. Since $\rho(x|\{x\},\{y\}) = 1 - \Phi_\rho(\{x\},\{y\})$, by Lemma 10, $\rho(x|\{x\},\{y\})$ is either equal to $(1-\eta)\pi_x + \eta(1-\mathring{\pi}_{\{x,y\}})$ or π_x, depending on whether $y\mathcal{N}x$ holds or not. Similarly, $\rho(y|\{y\},\{x\})$ is either equal to $(1-\eta)\pi_y + \eta(1-\mathring{\pi}_{\{x,y\}})$ or π_y. It follows that $\rho(x|\{x\},\{y\}) + \rho(y|\{y\},\{x\}) \geq \pi_x + \pi_y > 1 - \mathring{\pi}_{\{x,y\}}$. However, $\rho(x|\{x\},\{y\}) + \rho(y|\{y\},\{x\}) = 1 - \Phi_\rho(\{x,y\},\emptyset) = 1 - \mathring{\pi}_{\{x,y\}}$, which is a contradiction. This indicates that either $x \succ y$ or $y \succ x$. $\square$

Lemma 12. *If the random choice rule ρ satisfies Axioms 1 and 3-7, then the revealed preference relation $\succ$ is transitive.*

Proof of Lemma 12. It suffices to show that for all pairwise distinct $x, y, z \in X$, $x \succ y$ and $y \succ z$ imply $x \succ z$. Suppose to the contrary that we have $x \succ y$, $y \succ z$, and $z \succ x$. By Axiom 5, we have $\rho(x|\{x,y,z\},\emptyset) = \rho(x|\{x,z\},\{y\}) = 1 - \Phi_\rho(\{x,z\},\{y\}) - \rho(z|\{x,z\},\{y\}) = 1 - \Phi_\rho(\{x,z\},\{y\}) - \rho(z|\{z\},\{x,y\}) = 1 - \Phi_\rho(\{x,z\},\{y\}) - (1 - \Phi_\rho(\{z\},\{x,y\})) = \Phi_\rho(\{z\},\{x,y\}) - \Phi_\rho(\{x,z\},\{y\})$. Similarly, we have

$$\rho(y|\{x,y,z\},\emptyset) = \Phi_\rho(\{x\},\{y,z\}) - \Phi_\rho(\{x,y\},\{z\}),$$
$$\rho(z|\{x,y,z\},\emptyset) = \Phi_\rho(\{y\},\{x,z\}) - \Phi_\rho(\{y,z\},\{x\}).$$

It then follows that

$$\begin{aligned} 1 - \Phi_\rho(\{x,y,z\},\emptyset) = &\ \Phi_\rho(\{x\},\{y,z\}) + \Phi_\rho(\{y\},\{x,z\}) + \Phi_\rho(\{z\},\{x,y\}) \\ &- \Phi_\rho(\{y,z\},\{x\}) - \Phi_\rho(\{x,z\},\{y\}) - \Phi_\rho(\{x,y\},\{z\}). \end{aligned} \tag{17}$$

To prove the lemma, it suffices to show that equation (17) never holds. Note that the left-hand-side (LHS) of equation (17) has a fixed value while the value of the right-hand-side (RHS) of equation (17) depends on the associative network $\mathcal{N}$. We will show that the RHS is always strictly smaller than the LHS under any associative network $\mathcal{N}$. To do so, first note that for any menu (A, B) with $|B| \leq 2$, we can decompose $\Phi_\rho(A, B)$ as the summation of the three terms: $(1-\eta)\Phi^0_\rho(A, B)$, $(1-\eta)\eta\Phi^1_\rho(A, B)$ and $\eta^2\Phi^2_\rho(A, B)$, where $\Phi^k_\rho(A, B)$ denotes the probability for the

DM to choose the default option conditional on that the DM exercises k rounds of mental association. Since for any menu (A, B), $\Phi^2_\rho(A, B) \leq \Phi^1_\rho(A, B)$, and $|B| = 1$ implies $\Phi^2_\rho(A, B) = \Phi^1_\rho(A, B)$, to show that the value of the RHS of equation (17) is strictly less than that of the LHS, it suffices to show the following two inequalities:

$$\begin{aligned} 1 - \Phi_\rho(\{x, y, z\}, \emptyset) > \Phi^0_\rho(\{x\}, \{y, z\}) + \Phi^0_\rho(\{y\}, \{x, z\}) + \Phi^0_\rho(\{z\}, \{x, y\}) \\ - \Phi^0_\rho(\{y, z\}, \{x\}) - \Phi^0_\rho(\{x, z\}, \{y\}) - \Phi^0_\rho(\{x, y\}, \{z\}), \end{aligned} \tag{18}$$

$$\begin{aligned} 1 - \Phi_\rho(\{x, y, z\}, \emptyset) > \Phi^1_\rho(\{x\}, \{y, z\}) + \Phi^1_\rho(\{y\}, \{x, z\}) + \Phi^1_\rho(\{z\}, \{x, y\}) \\ - \Phi^1_\rho(\{y, z\}, \{x\}) - \Phi^1_\rho(\{x, z\}, \{y\}) - \Phi^1_\rho(\{x, y\}, \{z\}). \end{aligned} \tag{19}$$

Note that the RHS of (18) minus the LHS of (18) is equal to

$$\begin{aligned} &\sum_{w \in \{x,y,z\}} (\Phi_\rho(\{w\}, \emptyset) - \Phi_\rho(\{x, y, z\} \setminus w, \emptyset)) + \Phi_\rho(\{x, y, z\}, \emptyset) - 1 \\ &= \mathring{\pi}_x + \mathring{\pi}_y + \mathring{\pi}_z - \mathring{\pi}_{\{x,y\}} - \mathring{\pi}_{\{y,z\}} - \mathring{\pi}_{\{x,z\}} + \mathring{\pi}_{\{x,y,z\}} - 1 = -\pi_{\{x,y,z\}} < 0. \end{aligned}$$

Therefore, inequality (18) holds. It remains to show that inequality (19) holds. Consider the following possible associative networks.

Case 1: $\mathcal{N}_{\{x,y,z\}} = \{(x, x), (y, y), (z, z), (x, z), (y, x), (z, y)\}$. In this case, the RHS of (19) minus the LHS of (19) is equal to

$$\begin{aligned} &\Phi_\rho(\{x, y\}, \emptyset) + \Phi_\rho(\{y, z\}, \emptyset) + \Phi_\rho(\{x, z\}, \emptyset) - 2\Phi_\rho(\{x, y, z\}, \emptyset) - 1 \\ &= \mathring{\pi}_{\{x,y\}} + \mathring{\pi}_{\{y,z\}} + \mathring{\pi}_{\{x,z\}} - 2\mathring{\pi}_{\{x,y,z\}} - 1 = 2\pi_{\{x,y,z\}} - \pi_{\{x,y\}} - \pi_{\{x,z\}} - \pi_{\{y,z\}} < 0. \end{aligned}$$

Case 2: $\mathcal{N}_{\{x,y,z\}} = \{(x, x), (y, y), (z, z), (x, z), (y, z), (z, x)\}$. In this case, the RHS of (19) minus the LHS of (19) is equal to

$$\begin{aligned} &\Phi_\rho(\{x, z\}, \emptyset) + \Phi_\rho(\{y\}, \emptyset) - \Phi_\rho(\{x, y, z\}, \emptyset) - 1 \\ &= \mathring{\pi}_{\{x,z\}} + \mathring{\pi}_y - \mathring{\pi}_{\{x,y,z\}} - 1 = \pi_{\{x,y,z\}} - \pi_{\{x,y\}} - \pi_{\{y,z\}} < 0. \end{aligned}$$

Case 3: $\mathcal{N}_{\{x,y,z\}} = \{(x, x), (y, y), (z, z), (x, z), (y, x)\}$. In this case, the RHS of (19) minus the LHS of (19) is equal to

$$\Phi_\rho(\{x,y\},\emptyset)+\Phi_\rho(\{y\},\emptyset)+\Phi_\rho(\{x,z\},\emptyset)-\Phi_\rho(\{x,y,z\},\emptyset)-\Phi_\rho(\{x,y\},\emptyset)-1$$
$$=\mathring{\pi}_y+\mathring{\pi}_{\{x,z\}}-\mathring{\pi}_{\{x,y,z\}}-1=\pi_{\{x,y,z\}}-\pi_{\{x,y\}}-\pi_{\{y,z\}}<0.$$

Case 4: $\mathcal{N}_{\{x,y,z\}}=\{(x,x),(y,y),(z,z),(x,z),(y,z)\}$. In this case, the RHS of (19) minus the LHS of (19) is equal to

$$\Phi_\rho(\{x\},\emptyset)+\Phi_\rho(\{y\},\emptyset)-\Phi_\rho(\{x,y\},\emptyset)-1$$
$$=\mathring{\pi}_x+\mathring{\pi}_y-\mathring{\pi}_{\{x,y\}}-1=-\pi_{\{x,y\}}<0.$$

Case 5: $\mathcal{N}_{\{x,y,z\}}=\{(x,x),(y,y),(z,z),(x,z)\}$. In this case, the RHS of (19) minus the LHS of (19) is equal to

$$\Phi_\rho(\{x\},\emptyset)+\Phi_\rho(\{y\},\emptyset)-\Phi_\rho(\{x,y\},\emptyset)-1$$
$$=\mathring{\pi}_x+\mathring{\pi}_y-\mathring{\pi}_{\{x,y\}}-1=-\pi_{\{x,y\}}<0.$$

Case 6: $\mathcal{N}_{\{x,y,z\}}=\{(x,x),(y,y),(z,z)\}$. In this case, the RHS of (19) minus the LHS of (19) is equal to

$$\sum_{w\in\{x,y,z\}}(\Phi_\rho(\{w\},\emptyset)-\Phi_\rho(\{x,y,z\}\setminus w,\emptyset))+\Phi_\rho(\{x,y,z\},\emptyset)-1$$
$$=\mathring{\pi}_x+\mathring{\pi}_y+\mathring{\pi}_z-\mathring{\pi}_{\{x,y\}}-\mathring{\pi}_{\{y,z\}}-\mathring{\pi}_{\{x,z\}}+\mathring{\pi}_{\{x,y,z\}}-1=-\pi_{\{x,y,z\}}<0.$$

It can be shown that for any other associative network $\hat{\mathcal{N}}$ that is not covered by the above cases, there is one associative network $\mathcal{N}$ from the above cases such that either $\hat{\mathcal{N}}$ is symmetric to $\mathcal{N}$, or $\hat{\mathcal{N}}$ leads to the same value of $\Phi^1_\rho(\{y,z\},\{x\})+\Phi^1_\rho(\{x,z\},\{y\})+\Phi^1_\rho(\{x,y\},\{z\})$ as $\mathcal{N}$ does but a lower value of $\Phi^1_\rho(\{x\},\{y,z\})+\Phi^1_\rho(\{y\},\{x,z\})+\Phi^1_\rho(\{z\},\{x,y\})$ than $\mathcal{N}$ does. Therefore, inequality (19) holds for all associative networks. □

Lemma 13. *A random choice rule ρ is represented by $(\pi,\mathcal{N},\succ,\eta)$ as an ABCT if and only if (i) for all menu (A,B), equation (14) holds, and (ii) for all $x,y\in X$, if $x\succ y$, then for all menu (C,D) with $x,y\in C$, $\rho(x|C,D)=\rho(x|C\setminus y,Dy)$.*

Proof of Lemma 13. As we have shown in the necessity part, if ρ is represented by $(\pi, \mathcal{N}, \succ, \eta)$ as an ABCT, then conditions (i) and (ii) hold. Inversely, suppose that conditions (i) and (ii) hold. For any given menu (A, B), we can enumerate alternatives in A such that $A = \{x_i\}_{i=1}^n$ with $x_1 \succ ... \succ x_n$. By condition (ii), we have for all $k \in \{1, ..., n\}$,

$$\begin{aligned}\sum_{t=1}^{k} \rho(x_t|A, B) &= \sum_{t=1}^{k} \rho(x_t|A \setminus \{x_{k+1}, ..., x_n\}, B \cup \{x_{k+1}, ..., x_n\}) \\ &= 1 - \Phi_\rho(A \setminus \{x_{k+1}, ..., x_n\}, B \cup \{x_{k+1}, ..., x_n\}).\end{aligned}$$

With condition (i), we can uniquely identify $\rho(x_i|A, B)$ for all $i \in \{1, ..., n\}$. Therefore, there is a unique choice rule that satisfies conditions (i) and (ii). Since the random choice rule that is represented by $(\pi, \mathcal{N}, \succ, \eta)$ as an ABCT also satisfies conditions (i) and (ii), we conclude that any choice rule that satisfies conditions (i) and (ii) is the one that is represented by $(\pi, \mathcal{N}, \succ, \eta)$ as an ABCT. $\square$

The sufficiency for Axioms 1 and 3-7 to guarantee ρ to be an ABCT can be implied by Lemmas 10-13. If Axiom 2 is additionally satisfied, then $\eta = 1$, as we have shown in the proof of Lemma 10. It then follows that the random choice rule is an ABC. The uniqueness of π, $\mathcal{N}$ and $\succ$ is evident according to their definitions. The uniqueness of η (defined by equation (15)) is guaranteed by the non-triviality assumption on the random choice rule ρ. $\square$

Proof of Propsition 4. Let $\hat{\rho}(x^*|A, X \setminus A)$ be the choice frequency of x^* in menu $(A, X \setminus A)$ after the link (y, x^*) is added. It remains to show that when $\rho(x^*|A, X \setminus A) > 0$, we have $\hat{\rho}(x^*|A, X \setminus A) - \rho(x^*|A, X \setminus A) = \mathring{\pi}_{\bar{A}} \left(1 - \mathring{\pi}_{\overline{\mathcal{N}^+(z)} \setminus \bar{A}}\right)$. To see this, consider a partition $\{\bar{A}_1, \bar{A}_2\}$ of $\bar{A}$ with $\bar{A}_1 = \{y \in X : \text{for some } z \in A, z \succ x^* \text{ and } y\mathcal{N}^+ z\}$ and $\bar{A}_1 = \bar{A} \setminus \bar{A}_1$. Let events 1 and 2 denote that no alternative in $\bar{A}_1$ and $\bar{A}_2$ is initially considered, respectively. If event 1 does not occur, then some alternative better than x^* in A will be considered and always blocks the choice x^*. If event 2 does not occur, then x^* already appears in the final consideration set, and adding one more link cannot further boost the choice of x^*. Hence, the new link only affects the choice of x^* when both events occurs, of which the probability is $\mathring{\pi}_{\bar{A}}$. Conditional on the two events, the extra choice probability of x^* by the new link equals the chance that some alternative prompts the attention to x^* through the new link, i.e., $1 - \mathring{\pi}_{\overline{\mathcal{N}^+(z)} \setminus \bar{A}}$. $\square$

Proof of Proposition 5. Consider a random choice rule ρ that is represented by $(\pi, \mathcal{N}, \succ)$ as an ABC on $\mathcal{E}^F$. It remains to show that ρ is also represented by $(\pi, \mathcal{N}[\rho], \succ)$ on $\mathcal{E}^F$. To simplify the notation, let $\mathcal{W} = \mathcal{N}[\rho]$. Since $\mathcal{W} \subseteq \mathcal{N}$, it suffices to show that for all menu (A, B), $x \in A$ and $C \subseteq AB$, if $x = \max(\mathcal{N}^+_{AB}(C) \cap A; \succ)$, then $x \in \mathcal{W}^+_{AB}(C)$. The case $x \in C$ is trivial. Suppose that $x \notin C$. Since $x = \max(\mathcal{N}^+_{AB}(C) \cap A; \succ)$, there exists a sequence of mutually distinct alternatives $(x_k)_{k=1}^{n+1}$, where $n \in \mathbb{N}_+$, such that $x_1 \in C$, $x_{n+1} = x = \max(\{x_t\}_{t=1}^{n+1}; \succ)$, and for all $k \leq n$, $\mathcal{N}(x_k) \cap \{x_t\}_{t=k}^{n+1} = \{x_k, x_{k+1}\}$. It then suffices to show that for all $k \leq n$, $x_k \mathcal{W} x_{k+1}$. We prove this by induction. If $n = 1$, then $x_2 \succ x_1$ and $x_1 \mathcal{N} x_2$ implies $\rho(x_1|\{x_1, x_2\}, \emptyset) = 0$, which further implies $x_1 \mathcal{W} x_2$ by the definition of $\mathcal{W}$. Suppose that our hypothesis is true for all $n \leq m$. Consider the case where $n = m+1$. If the second best alternative in $\{x_t\}_{t=1}^{n+1}$ is not x_1, then apply our induction hypothesis twice, and we are done. If the second best alternative in $\{x_t\}_{t=1}^{n+1}$ is x_1, then by our assumption on $\mathcal{N}$, x_{n+1} is the only alternative chosen with positive probabilities in menus $(\{x_t\}_{t=1}^{n+1}, \emptyset)$ and $(\{x_t\}_{t=3}^{n+1}, \emptyset)$, and x_1 and x_{n+1} are the only two alternatives chosen with positive probabilities in menu $(\{x_1\} \cup \{x_t\}_{t=3}^{n+1}, \emptyset)$. Therefore, by the definition of $\mathcal{W}$, we have $x_1 \mathcal{W} x_2$. By applying the induction hypothesis to $(x_k)_{k=2}^{n+1}$, we are done. $\square$

Proof of Proposition 6. Let $\mathcal{A} = \{A \subseteq X : A \neq \emptyset, \mathcal{N}^+(X \setminus A) \cap A = \emptyset\}$. We have for each $A \in \mathcal{A}$, $\mathring{\pi}_A = \Phi_\rho(A, X \setminus A)$. Thus, it remains to show that only $\{\mathring{\pi}_A\}_{A \in \mathcal{A}}$ can be identified. Note that for every menu $(A, X \setminus A)$, $H_\mathcal{N}(A, X \setminus A) \in \mathcal{A}$, and we have $\Phi_\rho(A, X \setminus A) = \mathring{\pi}_{H_\mathcal{N}(A, X \setminus A)}$. For every $x \in A$, let $C = \{y \in A : y \succ x\}$. We have

$$\rho(x|A, X \setminus A) = 1 - \sum_{y \in C} \rho(y|Cx, X \setminus (Cx)) - \Phi_\rho(Cx, X \setminus (Cx))$$

$$= 1 - \sum_{y \in C} \rho(y|C, X \setminus C) - \Phi(Cx, X \setminus (Cx)) = \Phi_\rho(C, X \setminus C) - \Phi_\rho(Cx, X \setminus (Cx)).$$

Thus, ρ does not reveal additional information on π given $\{\mathring{\pi}_A\}_{A \in \mathcal{A}}$. $\square$

References

Agranov, M. and P. Ortoleva (2022): "Revealed Preferences for Randomization: An Overview," in *AEA Papers and Proceedings*, vol. 112, 426–430.

Aguiar, V. H., M. J. Boccardi, N. Kashaev, and J. Kim (2023): "Random Utility and Limited Consideration," *Quantitative Economics*, 14, 71–116.

Anderson, J. R. (1996): *The Architecture of Cognition*, vol. 5, Psychology Press.

Anderson, J. R. and G. H. Bower (1973): *Human Associative Memory*, Winston & Sons.

Apesteguia, J., M. A. Ballester, and J. Lu (2017): "Single-Crossing Random Utility Models," *Econometrica*, 85, 661–674.

Barseghyan, L., M. Coughlin, F. Molinari, and J. C. Teitelbaum (2021a): "Heterogeneous Choice Sets and Preferences," *Econometrica*, 89, 2015–2048.

Barseghyan, L., F. Molinari, and M. Thirkettle (2021b): "Discrete Choice under Risk with Limited Consideration," *American Economic Review*, 111, 1972–2006.

Ben-Akiva, M. E. (1973): "Structure of Passenger Travel Demand Models." Ph.D. thesis, Massachusetts Institute of Technology.

Block, H. D. and J. Marschak (1960): "Random Orderings and Stochastic Theories of Responses," *Contributions to Probability and Statistics*, 97–132.

Bordalo, P., N. Gennaioli, and A. Shleifer (2020): "Memory, Attention, and Choice," *The Quarterly Journal of Economics*, 135, 1399–1442.

——— (2022): "Salience," *Annual Review of Economics*, 14, 521–544.

Brady, R. L. and J. Rehbeck (2016): "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84, 1203–1223.

Brandt, C., C. P. De Mortanges, C. Bluemelhuber, and A. C. Van Riel (2011): "Associative Networks: A New Approach to Market Segmentation," *International Journal of Market Research*, 53, 187–207.

Camerer, C. F. (1997): "Progress in Behavioral Game Theory," *Journal of Economic Perspectives*, 11, 167–188.

Cattaneo, M. D., P. Cheung, X. Ma, and Y. Masatlioglu (2023): "Attention Overload," Working paper.

Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov (2020): "A Random Attention Model," *Journal of Political Economy*, 128, 2796–2836.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019): "Deliberately Stochastic," *American Economic Review*, 109, 2425–2445.

Chadd, I., E. Filiz-Ozbay, and E. Y. Ozbay (2021): "The Relevance of Irrelevant Information," *Experimental Economics*, 24, 985–1018.

Cheung, P. and Y. Masatlioglu (2023): "Decision Making with Recommendation," Working paper.

Cheung, P. H. and Y. Masatlioglu (2024): “Disentangling Attention and Utility Channels in Recommendations,” Working paper.

Corbetta, M. and G. L. Shulman (2002): “Control of Goal-directed and Stimulus-driven Attention in the Brain,” *Nature Reviews Neuroscience*, 3, 201–215.

Cunha Jr, M., M. R. Forehand, and J. W. Angle (2015): “Riding Coattails: When Co-Branding Helps versus Hurts Less-Known Brands,” *Journal of Consumer Research*, 41, 1284–1300.

Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2020): “Inferring Cognitive Heterogeneity from Aggregate Choices,” *Econometrica*, 88, 1269–1296.

Dean, M., Ö. Kibris, and Y. Masatlioglu (2017): “Limited Attention and Status Quo Bias,” *Journal of Economic Theory*, 169, 93–127.

Ellis, A. and H. C. Thysen (2024): “Subjective Causality in Choice,” Working paper.

Falmagne, J.-C. (1978): “A Representation Theorem for Finite Random Scale Systems,” *Journal of Mathematical Psychology*, 18, 52–72.

Farquhar, P. H. and A. R. Pratkanis (1993): “Decision Structuring with Phantom Alternatives,” *Management Science*, 39, 1214–1226.

Gazzaley, A. and A. C. Nobre (2012): “Top-down Modulation: Bridging Selective Attention and Working Memory,” *Trends in Cognitive Sciences*, 16, 129–135.

Geng, J. J. and M. Behrmann (2005): “Spatial Probability as an Attentional Cue in Visual Search,” *Perception & Psychophysics*, 67, 1252–1268.

Gul, F., P. Natenzon, and W. Pesendorfer (2014): “Random Choice as Behavioral Optimization,” *Econometrica*, 82, 1873–1912.

Gul, F. and W. Pesendorfer (2006): “Random Expected Utility,” *Econometrica*, 74, 121–146.

Guney, B., M. Richter, and M. Tsur (2018): “Aspiration-Based Choice,” *Journal of Economic Theory*, 176, 935–956.

Kashaev, N. and V. H. Aguiar (2022): “A Random Attention and Utility Model,” *Journal of Economic Theory*, 204, 105487.

Keller, K. L. (1993): “Conceptualizing, Measuring, and Managing Customer-Based Brand Equity,” *Journal of Marketing*, 57, 1–22.

Kovach, M. and E. Suleymanov (2023): “Reference Dependence and Random Attention,” *Journal of Economic Behavior & Organization*, 215, 421–441.

Kovach, M. and G. Tserenjigmid (2022): “Behavioral Foundations of Nested Stochastic Choice and Nested Logit,” *Journal of Political Economy*, 130, 2411–2461.

Lea, A. M. and M. J. Ryan (2015): "Irrationality in Mate Choice Revealed by Túngara Frogs," *Science*, 349, 964–966.

Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2017): "When More is Less: Limited Consideration," *Journal of Economic Theory*, 170, 70–85.

Luce, R. D. (1959): *Individual Choice Behavior: A Theoretical Analysis*, New York: Wiley.

Manski, C. F. (1977): "The structure of random utility models," *Theory and decision*, 8, 229.

Manzini, P. and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82, 1153–1176.

Masatlioglu, Y. and D. Nakajima (2013): "Choice by Iterative Search," *Theoretical Economics*, 8, 701–728.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012): "Revealed Attention," *American Economic Review*, 102, 2183–2205.

Masatlioglu, Y. and E. A. Ok (2005): "Rational Choice with Status Quo Bias," *Journal of economic theory*, 121, 1–29.

——— (2014): "A Canonical Model of Choice with Initial Endowments," *The Review of Economic Studies*, 81, 851–883.

Masatlioglu, Y. and E. Suleymanov (2021): "Decision Making within a Product Network," *Economic Theory*, 71, 185–209.

McFadden, D. (1978): "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlqvist, J. W. Weibull, L. Lundqvist, and F. Snickars, Amsterdam: North-Holland, 75–96.

Natenzon, P. (2019): "Random Choice and Learning," *Journal of Political Economy*, 127, 419–457.

Payne, J. W., J. R. Bettman, and E. J. Johnson (1993): *The Adaptive Decision Maker*, Cambridge University Press.

Raaijmakers, J. G. and R. M. Shiffrin (1981): "Search of Associative Memory," *Psychological Review*, 88, 93.

Salant, Y. and A. Rubinstein (2008): "(A,f): Choice with Frames," *The Review of Economic Studies*, 75, 1287–1296.

Simon, H. A. (1955): "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics*, 69, 99–118.

Soltani, A., B. De Martino, and C. Camerer (2012): "A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence," *PLoS Computational Biology*, 8, e1002607.

Teichert, T. A. and K. Schöntag (2010): “Exploring Consumer Knowledge Structures Using Associative Network Analysis,” *Psychology & Marketing*, 27, 369–398.

Valkanova, K. (2021): “Markov Stochastic Choice,” Working paper.

Yegane, E. (2022): “Stochastic Choice with Limited Memory,” *Journal of Economic Theory*, 205, 105548.

Yegane, E. and Y. Masatlioglu (2023): “Revealed Attention and Memory,” Working paper.

Online Appendix (for online publication only)

This online appendix to "*Associative Networks in Decision Making*" is organized as follows. In Section OA-1, we provide a characterization of the association based consideration rule (ABC) restricted on the domain $\mathcal{E}^F = \{(A, B) \in \mathcal{E} : B = \emptyset\}$ where every observable alternative is available. In Section OA-2, we discuss more general models of initial attention distributions.

OA-1. Restricted Domain

In this section, we axiomatize ABCs on $\mathcal{E}^F$. The first axiom is the Default Independence axiom that we introduced in the main text.

Axiom A1—Default Independence: For all $x \in X$ and $A, B \in \mathcal{M}$ with $x \in A \cap B$:

$$\frac{\Phi_\rho(A, \emptyset)}{\Phi_\rho(A \setminus x, \emptyset)} = \frac{\Phi_\rho(B, \emptyset)}{\Phi_\rho(B \setminus x, \emptyset)}.$$

For a given choice rule ρ and menu $(A, \emptyset)$, let $c_\rho(A, \emptyset) := \{x \in A : \rho(x|A, \emptyset) > 0\}$ be the set of chosen alternatives in A, i.e., those in A that are chosen with positive probabilities. We impose the next two axioms on the set of chosen alternatives.

Axiom A2—Sen's α: For all $A, B \in \mathcal{M}$, $B \subseteq A$ implies $c_\rho(A, \emptyset) \cap B \subseteq c_\rho(B, \emptyset)$.

Axiom A3—Reducibility: For all $A \in \mathcal{M}$, if for every $x \in A$, $c_\rho(A, \emptyset) \neq c_\rho(A \setminus x, \emptyset)$, then $c_\rho(A, \emptyset) = A$.

Axiom A2 states that if an alternative is selected from a larger menu, it must also be selected from any smaller menu that contains it. To interpret, if a particular alternative x is not chosen in a smaller menu, then given the presence of more competitive alternatives in a larger menu, it should also remain unselected. In our context, if an alternative is not chosen, its consideration must lead to the consideration of a better alternative. Consequently, in a larger menu, the superior alternative remains to be associated with x and thus blocks the choice of x.

The contrapositive of Axiom A3 states that if not all alternatives are selected, then there exists an unselected alternative whose removal does not alter the set of chosen alternatives. To illustrate this axiom, consider a menu $(\{x, y, z\}, \emptyset)$ where only x is chosen. As both y and z are unselected, their consideration must lead to the consideration of a superior alternative in this menu, which has to be x. If the removal of y results in a change in the set of chosen alternatives such that z becomes chosen, then x must be associated with z through y, and x must be directly associated with y. In this scenario, the deletion of z does not alter the association relation between x and y, and thus does not affect the set of chosen alternatives. In summary, Axiom A3 establishes the existence of an unselected alternative (if not all alternatives are chosen) whose removal does not impact the association relation among the remaining alternatives, thereby preserving the set of chosen alternatives.

For any $x \in X$ and $A \in \mathcal{M}$, we say that x is associatively independent of A, denoted by $x \vdash A$, if $x \notin A$ and for all $y \in A$, $\rho(y|A, \emptyset) = \rho(y|Ax, \emptyset)$. Note that according to this definition, for all $x \in X$, we have $x \vdash \emptyset$.

Axiom A4—Weak I-Independence: For all $x \in X$ and $A, B \in \mathcal{M}$, if $x \vdash A$ and $x \vdash B$, then $x \vdash A \cup B$.

Axiom A4 posits that if x is associatively independent of both A and B, then it is also associatively independent of their union. Notably, this axiom can be implied by the I-Independence axiom of MM14. According to the I-Independence axiom, if x does not affect the frequency of selecting alternative y in a particular menu, then it should not impact the frequency of choosing y in *every* menu.

For any two alternatives x and y, we say that x *weakly dominates* y, denoted by $x \trianglerighteq y$, if there is a menu A such that $y \in A$ and $c_\rho(A, \emptyset) = \{x\}$.

Axiom A5—Dominance Asymmetry: For all $x, y, z \in X$ and $A, B \in \mathcal{M}$ such that $x \neq z$, $x \trianglerighteq y$, $y \in c_\rho(A, \emptyset)$ and $z \in c_\rho(B, \emptyset)$, we have

$$\rho(z|A, \emptyset) \neq \rho(z|A \setminus y, \emptyset) \Rightarrow \rho(x|B, \emptyset) = \rho(x|B \setminus z, \emptyset).$$

To understand Axiom A5, observe that deleting y from menu $(A, \emptyset)$ changes the choice frequency of z. It then follows that either y is worse than z, in which case the consideration of y prompts the consideration of z and boosts the choice of z, or y is better than z, in which case the presence of y hinders the choice of z. Since y is chosen in $(A, \emptyset)$, the former case cannot be true. Thus, y must be better than z, and the alternative x that weakly dominates y is even better. Therefore, deleting z from any menu in which z is chosen will not affect the choice frequency of x since the consideration of z cannot prompt the consideration of x in that menu.

Theorem A1. *Axioms A1-A5 are sufficient and necessary for a random choice rule ρ to be an ABC on $\mathcal{E}^F$.*

Proof of Theorem A1. **(Necessity)** Consider an ABC ρ on $\mathcal{E}^F$ that is represented by $(\pi, \mathcal{N}, \succ)$. Axiom A1 holds trivially. For Axiom A2, consider $A, B \in \mathcal{M}$ with $B \subseteq A$. If for some $x \in B$, $x \notin c_\rho(B, \emptyset)$, then we have $x \neq \max(\mathcal{N}_B^+(x); \succ)$. Since $\mathcal{N}_B^+(x) \subseteq \mathcal{N}_A^+(x)$, we have $x \neq \max(\mathcal{N}_A^+(x); \succ)$. Thus, $x \notin c_\rho(A, \emptyset)$.

For Axiom A3, consider $A \in \mathcal{M}$ such that $c_\rho(A, \emptyset) \neq A$. Let $c_\rho(A, \emptyset) = \{x_1, ..., x_n\}$ such that for all $k \in \{1, ..., n-1\}$, $x_k \succ x_{k+1}$. Consider a partition $\{B_k\}_{k=1}^n$ of A such that for every k, $B_k = \{y \in A : \mathcal{N}_A^+(y) \cap \{x_1, ..., x_k\} = \{x_k\}\}$. Note that each B_k contains x_k, and for all $y \in B_k \setminus x_k$, we have $y \notin c_\rho(A, \emptyset)$, $x_k \succ y$ and $x_k \in \mathcal{N}_{B_k}^+(y)$. Consider some k such that $B_k \setminus x_k \neq \emptyset$. It is easy to show that there is an alternative $y \in B_k \setminus x_k$ such that for all $z \in B_k \setminus y$, $x_k \in \mathcal{N}_{B_k \setminus y}^+(z)$. Deleting y from menu $(A, \emptyset)$ will not affect the choices.

For Axiom A4, note that $x \vdash A$ if and only if for all $y \in c_\rho(A, \emptyset)$, $y \succ x$, and for all $z \in A$, $(x, z) \notin \mathcal{N}$. It follows that $x \vdash A$ and $x \vdash B$ imply $x \vdash AB$.

For Axiom A5, it suffices to show that for two distinct alternatives x and y, if $x \in c_\rho(A, \emptyset)$ and $\rho(y|A, \emptyset) \neq \rho(y|A \setminus x, \emptyset)$, then $x \succ y$. Let $c_\rho(A, \emptyset) = \{x_1, ..., x_n\}$ such that for all $k \in \{1, ..., n-1\}$, $x_k \succ x_{k+1}$. Consider the partition $\{B_k\}_{k=1}^n$ constructed in the proof for Axiom A3. We have $x = x_k$ for some k. If $y \notin c_\rho(A, \emptyset)$, then $\rho(y|A, \emptyset) \neq \rho(y|A \setminus x, \emptyset)$ implies $y \in B_k$, and thus $x \succ y$. If $y \in c_\rho(A, \emptyset)$, then $\rho(y|A, \emptyset) \neq \rho(y|A \setminus x, \emptyset)$ implies $y = x_t$ for some $t > k$, and thus $x \succ y$.

(Sufficiency) Through out the proof of sufficiency, we assume that Axioms A1-A5 hold. For the attention probability function π, let $\pi_x = \rho(x|\{x\}, \emptyset)$ for every $x \in X$. For the preference ordering $\succ$, let $x \succ y$ if $x \neq y$, $x \in c_\rho(\{x, y\}, \emptyset)$ and $\rho(y|\{x, y\}, \emptyset) \neq \rho(y|\{y\}, \emptyset)$. For the associative network $\mathcal{N}$, let $(x, y) \in \mathcal{N}$ if and only if either (i) $x = y$, or (ii) $x \neq y$ and there exists $A \in \mathcal{M}$ such that $x \vdash A$ and $x \notin c_\rho(A \cup \{x, y\}, \emptyset)$. We proceed with a sequence of lemmas.

Lemma A1. *For all $x \in X$ and $A \in \mathcal{M}$, if $x \vdash A$, then $x \in c_\rho(Ax, \emptyset)$.*

Proof of Lemma A1. Since $x \vdash A$, we have $\sum_{y \in A} \rho(y|A, \emptyset) = \sum_{y \in A} \rho(y|Ax, \emptyset)$. Since $\Phi_\rho(Ax, \emptyset) < \Phi_\rho(A, \emptyset)$, we have $\rho(x|Ax, \emptyset) \neq 0$, i.e., $x \in c_\rho(Ax, \emptyset)$. $\square$

Lemma A2. *The binary relation $\succ$ is a preference ordering and satisfies that for all $x, y \in X$ and $A \in \mathcal{M}$, if $x \succ y$ and $y \in c_\rho(A, \emptyset)$, then $\rho(x|A, \emptyset) = \rho(x|A \setminus y, \emptyset)$.*

Proof of Lemma A2. The claim that $x \succ y$ and $y \in c_\rho(A, \emptyset)$ imply $\rho(x|A, \emptyset) = \rho(x|A \setminus y, \emptyset)$ follows from the definition of $\succ$ and Axiom A5. Showing that $\succ$ is well-defined for each distinct pair of alternatives and asymmetric is trivial. To see that $\succ$ is transitive, suppose to the contrary that there are three mutually distinct alternatives x, y and z such that $x \succ y$, $y \succ z$, and $z \succ x$. By symmetry, we can focus on three representative cases, where in case 1, $c_\rho(\{x, y, z\}, \emptyset) = \{x, y, z\}$, in case 2, $c_\rho(\{x, y, z\}, \emptyset) = \{x, y\}$, and in case 3, $c_\rho(\{x, y, z\}, \emptyset) = \{x\}$. We want to show that all the three cases lead to contradiction.

For case 1, we have $1 - \mathring{\pi}_{\{x,y,z\}} = 1 - \Phi(\{x, y, z\}, \emptyset) = \sum_{w \in \{x,y,z\}} \rho(w|\{x, y, z\}, \emptyset)$ $= \rho(x|\{x, z\}, \emptyset) + \rho(y|\{x, y\}, \emptyset) + \rho(z|\{y, z\}, \emptyset) = 1 - \Phi(\{x, z\}, \emptyset) - \rho(z|\{x, z\}, \emptyset) + 1 - \Phi(\{x, y\}, \emptyset) - \rho(x|\{x, y\}, \emptyset) + 1 - \Phi(\{y, z\}, \emptyset) - \rho(y|\{y, z\}, \emptyset) = 3 - \Phi(\{x, z\}, \emptyset) - \rho(z|\{z\}, \emptyset) - \Phi(\{x, y\}, \emptyset) - \rho(x|\{x\}, \emptyset) - \Phi(\{y, z\}, \emptyset) - \rho(y|\{y\}, \emptyset) = \pi_x + \pi_y + \pi_z - \pi_{\{x,y\}} - \pi_{\{y,z\}} - \pi_{\{x,z\}} < 1 - \mathring{\pi}_{\{x,y,z\}}$, which is a contradiction.

For case 2, since $c_\rho(\{x, y, z\}, \emptyset) = \{x, y\}$ and $z \succ x$, we have $c_\rho(\{y, z\}, \emptyset) = \{y\}$. It follows that $y \trianglerighteq z$, and thus by Axiom A5 and $x \succ y$, we have for all $A \in \mathcal{M}$ with $x \in c_\rho(A, \emptyset)$, $\rho(y|A, \emptyset) = \rho(y|A \setminus x, \emptyset)$. By Axiom A2, $x \in c_\rho(\{x, y\}, \emptyset)$, and thus $\rho(y|\{x, y\}, \emptyset) = \rho(y|\{y\}, \emptyset)$, which contradicts to the fact that $x \succ y$.

For case 3, since $c_\rho(\{x, y, z\}, \emptyset) = \{x\}$, we have $x \trianglerighteq z$. By $z \succ x$, we have $z \in c_\rho(\{x, z\}, \emptyset)$ and $\rho(x|\{x, z\}, \emptyset) \neq \rho(x|\{x\}, \emptyset)$. By Axiom A5, this contradicts to $x \trianglerighteq z$. $\square$

Lemma A3. *For all $x \in X$ and $A \in \mathcal{M}$ with $x \in A$, if $x \notin c_\rho(A, \emptyset)$, then there exists $y \in \mathcal{N}_A^+(x) \cap c_\rho(A, \emptyset)$ such that $y \succ x$.*

Proof of Lemma A3. We prove by induction on $|A|$. First, if $|A| = 2$, then $A = \{x, y\}$ and $x \notin c_\rho(\{x, y\}, \emptyset)$. Then by the construction of $\mathcal{N}$ and $\succ$, we have $(x, y) \in \mathcal{N}$ and $y \succ x$. Therefore, the lemma holds when $|A| = 2$.

Assume that the lemma holds whenever $|A| \leq n$, where $n \geq 2$. Consider the case where $|A| = n+1$. Since $x \notin c_\rho(A, \emptyset)$, by Axiom A3, there exists $y \in A$ such that $c_\rho(A, \emptyset) = c_\rho(A \setminus y, \emptyset)$. If $y \neq x$, then $x \notin c_\rho(A \setminus y, \emptyset)$, and we are done by our induction hypothesis. Hence, consider the case where x is the only alternative in A such that $c_\rho(A, \emptyset) = c_\rho(A \setminus x, \emptyset)$. By a similar argument, we can additionally assume that for all $z \in c_\rho(A, \emptyset)$, we have $x \in c_\rho(A \setminus z, \emptyset)$. Thus, by Lemma A2, we have for all $z \in c_\rho(A, \emptyset)$, $z \succ x$.

To proceed, consider $(A \setminus x, \emptyset)$. We first show that if $c_\rho(A \setminus x, \emptyset) = A \setminus x$, then $|A \setminus x| = 1$, and we are done. To see this, suppose to the contrary that $|A \setminus x| \geq 2$, and let y and z be two distinct alternatives in $A \setminus x$. By Axiom A2 and the assumptions we impose on the case we consider, we have $x \in c_\rho(A \setminus y, \emptyset) = A \setminus y$ and $x \in c_\rho(A \setminus z, \emptyset) = A \setminus z$. Since for all $\hat{x} \in A \setminus x$, we have $\hat{x} \succ x$, by Lemma A2, we have $x \vdash A \setminus \{x, y\}$ and $x \vdash A \setminus \{x, z\}$. It follows from Axiom A4 that $x \vdash A \setminus x$, which by Lemma A1 is a contradiction since $x \notin c_\rho(A, \emptyset)$.

By the above argument, we can additionally assume that $c_\rho(A \setminus x, \emptyset) \neq A \setminus x$. By Axiom A3, there exists $z \in A \setminus x$ such that $c_\rho(A \setminus \{x, z\}, \emptyset) = c_\rho(A \setminus x, \emptyset) = c_\rho(A, \emptyset)$. By Axiom A2, we have $c_\rho(A \setminus z) = \{x\} \cup c_\rho(A)$. Since $x \in c_\rho(A \setminus z, \emptyset) = \{x\} \cup c_\rho(A)$ and for all $w \in c_\rho(A)$, $w \succ x$, by Lemma A2, we have $x \vdash A \setminus \{x, z\}$. Since $x \notin c_\rho(A, \emptyset)$, we have $(x, z) \in \mathcal{N}$. Since $z \notin c_\rho(A \setminus x, \emptyset)$, by the induction hypothesis, there exists $y \in c_\rho(A \setminus x, \emptyset)$ such that $y \succ z$ and $y \in \mathcal{N}^+_{A \setminus x}(z)$. Thus, we have $y \in \mathcal{N}^+_A(x)$. Since $y \in c_\rho(A \setminus x, \emptyset)$, we have $y \in c_\rho(A, \emptyset)$, and thus $y \succ x$. $\square$

Lemma A4. *For all $x \in X$ and $A \in \mathcal{M}$, if $x \in c_\rho(A, \emptyset)$, then $x = \max(\mathcal{N}^+_A(x); \succ)$.*

Proof of Lemma A4. We show that in menu $(A, \emptyset)$, if there is an alternative that is associated with x and $\succ$-better than x, then x is not chosen. By Axiom A2, it suffices to show that for any sequence of alternatives $(x_k)_{k=1}^n$, where $n \geq 2$, if for all $k \in \{1, ..., n-1\}$, $x_n \succ x_k$ and $(x_k, x_{k+1}) \in \mathcal{N}$, then $x_1 \notin c_\rho(\{x_1, ..., x_n\}, \emptyset)$. We show this by induction on n. First, let $n = 2$. We have $x_1 \mathcal{N} x_2$ and $x_2 \succ x_1$. Suppose to the contrary that $x_1 \in c_\rho(\{x_1, x_2\}, \emptyset)$, then by Lemma A2 and the construction of $\succ$, we have $c_\rho(\{x_1, x_2\}, \emptyset) = \{x_1, x_2\}$ and $\rho(x_2 | \{x_1, x_2\}, \emptyset) = \rho(x_2 | \{x_2\}, \emptyset)$, i.e., $x_1 \vdash \{x_2\}$. However, since $x_1 \mathcal{N} x_2$, by the construction of $\mathcal{N}$, we can find $A \in \mathcal{M}$ such that $x_1 \vdash A$ and $x_1 \notin c_\rho(A \cup \{x_1, x_2\}, \emptyset)$. By Axiom A4, we have $x_1 \vdash Ax_2$, and by Lemma A1, we have $x_1 \in c_\rho(A \cup \{x_1, x_2\}, \emptyset)$, which is a contradiction. Thus, we must have $x_1 \notin c_\rho(\{x_1, x_2\}, \emptyset)$.

Next, suppose that the induction hypothesis holds for all $n \leq m$ ($m \geq 2$). Consider the case where $n = m + 1$. Since for all $k \in \{1, ..., n-1\}$, $x_n \succ x_k$, we have $c_\rho(\{x_2, ..., x_n\}, \emptyset) = \{x_n\}$ by our induction hypothesis. Suppose to the contrary that $x_1 \in c_\rho(\{x_1, ..., x_n\}, \emptyset)$, by Axiom A2, we have $c_\rho(\{x_1, ..., x_n\}, \emptyset) = \{x_1, x_n\}$. By Lemma A2, we have $\rho(x_n|\{x_1, ..., x_n\}, \emptyset) = \rho(x_n|\{x_2, ..., x_n\}, \emptyset)$. Thus $x_1 \vdash \{x_2, ..., x_n\}$. Since $x_1 \mathcal{N} x_2$, we can find $A \in \mathcal{M}$ such that $x_1 \vdash A$ and $x_1 \notin c_\rho(A \cup \{x_1, x_2\}, \emptyset)$. By Axiom A4, we have $x_1 \vdash A \cup \{x_2, ..., x_n\}$, and by Lemma A1, we have $x_1 \in c_\rho(A \cup \{x_1, ..., x_n\}, \emptyset)$. It follows from Axiom A2 that $x_1 \in c_\rho(A \cup \{x_1, x_2\}, \emptyset)$, which is a contradiction. Therefore, we have $x_1 \notin c_\rho(\{x_1, ..., x_n\}, \emptyset)$. $\square$

With Lemmas A3 and A4, we have for all $A \in \mathcal{M}$ and $x \in A$, $x \in c_\rho(A, \emptyset)$ if and only if $x = \max(\mathcal{N}_A^+(x); \succ)$. Let $c_\rho(A, \emptyset) = \{x_1, ..., x_n\}$ such that for all $k \in \{1, ..., n-1\}$, $x_k \succ x_{k+1}$. We can have a partition $\{B_k\}_{k=1}^n$ of A such that for every k, $B_k = \{y \in A : \mathcal{N}_A^+(y) \cap \{x_1, ..., x_k\} = \{x_k\}\}$. Note that for each k, $x_k \in B_k$, and for all $y \in B_k \setminus x_k$, $y \notin c_\rho(A, \emptyset)$, $x_k \succ y$ and $x_k \in \mathcal{N}_{B_k}^+(y)$. To show that ρ can be represented by $(\pi, \mathcal{N}, \succ)$ as an ABC on $\mathcal{E}^F$, it suffices to show that for all $k \in \{1, ..., n\}$,

$$\sum_{t=1}^{k} \rho(x_t|A, \emptyset) = 1 - \mathring{\pi}_{C_k}, \tag{20}$$

where $C_k = \cup_{t=1}^k B_k$. Note that equation (20) holds when $k = n$. Consider some $k < n$. Let $D_k = A \setminus C_k$. It follows that $\mathcal{N}_A^+(D_k) \cap C_k = \emptyset$. Thus, for all $D \subseteq D_k$, $c_\rho(C_k \cup D, \emptyset) \cap D \neq \emptyset$. Therefore, we can enumerate $D_k = \{y_1, ..., y_m\}$ such that for all $t \in \{1, ..., m\}$, $y_t \in c_\rho(C_k \cup \{y_1, ..., y_t\}, \emptyset)$. Note that for all $D \subseteq D_k$, $\{x_1, ..., x_k\} = c_\rho(C_k \cup D, \emptyset) \cap C_k$. It follows from Lemma A2 that for all $t \in \{1, ..., k\}$ and $s \in \{1, ..., m\}$, $\rho(x_t|C_k, \emptyset) = \rho(x_t|C_k \cup \{y_1, ..., y_s\}, \emptyset) = \rho(x_t|A, \emptyset)$. Therefore,

$$\sum_{t=1}^{k} \rho(x_t|A, \emptyset) = \sum_{t=1}^{k} \rho(x_t|C_k, \emptyset) = 1 - \Phi(C_k, \emptyset) = 1 - \mathring{\pi}_{C_k}.$$

The sufficiency is thus shown. $\square$

OA-2. General Models of Initial Attention

In this section, we study two relaxations of our assumption regarding how the DM's initial attention set is formed. First, we consider more general initial attention distributions by relaxing the assumption of independent attention. Second, we maintain the assumption of independent attention but allow the DM's attention probability for an alternative to depend on its availability.

General attention distributions. Consider a general attention distribution function $\sigma : \mathcal{M} \times \mathcal{M} \to [0,1]$ such that for all $A \in \mathcal{M}$, $\sum_{B \subseteq A} \sigma(B, A) = 1$, and $\sigma(B, A) > 0$ if and only if $B \subseteq A$. To interpret, $\sigma(B, A)$ is the probability that the DM's initial consideration set is B when A is the set of all observable alternatives. With σ, if the DM's associative network and preference ordering are given by $\mathcal{N}$ and $\succ$ respectively, then for all menu (A, B) and $x \in A$, we have

$$\rho(x|A, B) = \sum_{C \subseteq AB:\, x = \max(\mathcal{N}^{+}_{AB}(C) \cap A;\, \succ)} \sigma(C, AB).$$

In what follows, we demonstrate that both $\mathcal{N}$ and $\succ$ can be uniquely identified. To see this, consider two distinct alternatives x and y. The identification of $\mathcal{N}$ is exactly the same as that in our baseline model. If $x\mathcal{N}y$, then

$$\Phi_\rho(\{y\}, \{x\}) = \sigma(\emptyset, \{x, y\}) = \Phi_\rho(\{x, y\}, \emptyset),$$

and if not $x\mathcal{N}y$, then

$$\Phi_\rho(\{y\}, \{x\}) = \sigma(\emptyset, \{x, y\}) + \sigma(\{x\}, \{x, y\}) > \Phi_\rho(\{x, y\}, \emptyset).$$

Therefore, $x\mathcal{N}y$ if and only if $\Phi_\rho(\{y\}, \{x\}) = \Phi_\rho(\{x, y\}, \emptyset)$. For the identification of the preference ordering, note that if $x \succ y$, then we have

$$\begin{aligned}
&\rho(x|\{x, y\}, \emptyset) = \rho(x|\{x\}, \{y\}), \text{and} \\
&\rho(y|\{x, y\}, \emptyset) \leq \sigma(\{y\}, \{x, y\}) \\
&< \sigma(\{y\}, \{x, y\}) + \sigma(\{x, y\}, \{x, y\}) \leq \rho(y|\{y\}, \{x\}).
\end{aligned}$$

Therefore, $x \succ y$ if and only if $\rho(x|\{x, y\}, \emptyset) = \rho(x|\{x\}, \{y\})$.

While the general attention distribution may be hard to identify, certain parametric assumptions on σ can lead to a unique identification. For instance, consider the attention distribution σ introduced by Brady and Rehbeck (2016):

There is a function $\zeta : \mathcal{M} \to (0,1)$ such that for all $B \subseteq A$, $\sigma(B,A) = \frac{\zeta(B)}{\sum_{C \subseteq A} \zeta(C)}$. Brady and Rehbeck (2016) show that such an attention distribution generalizes the independent attention distribution in MM14. In fact, for all $A \subseteq X$, $\frac{\zeta(A)}{\zeta(\emptyset)}$ can be pinned down inductively via $\Phi(A,\emptyset)$.[26] Therefore, ζ is unique up to rescaling.

Availability-dependent attention. Our baseline model assumes that the attention probability of an observable alternative is the same regardless of its availability. However, this assumption may not hold in certain contexts. For example, in some online shopping platforms, products that are sold out are explicitly labeled as "out of stock" on the display page. It is plausible that such labeling may result in excessive or less attention from the consumer. Therefore, a natural extension of our baseline model is to incorporate the availability of alternatives as a factor that influences the attention probability assigned to them.

Formally, let $\pi^f : X \to (0,1)$ be the attention probability function for available alternatives, and $\pi^n : X \to (0,1)$ be the attention probability function for unavailable but observable alternatives.[27] Let $\mathcal{N}$ be the associative network and $\succ$ be the DM's preference ordering. For each menu (A,B) and $x \in A$, we have

$$\rho(x|A,B) = \sum_{C \subseteq AB:\, x = \max(\mathcal{N}^+_{AB}(C) \cap A; \succ)} \pi^f_{C \cap A} \pi^n_{C \cap B} \mathring{\pi}^f_{A \setminus C} \mathring{\pi}^n_{B \setminus C}.$$

We argue that all relevant parameters of the model above can be uniquely identified. First, the preference ordering $\succ$ can be identified similarly as Case 1 in Section 7.1, and the associative network $\mathcal{N}$ can be identified similarly as in our baseline model. Second, the attention probability for each alternative x when it is available is given by $\pi^f_x = \rho(x|\{x\},\emptyset)$. Finally, the attention probabilities for unavailable but observable alternatives can be identified through the extent to which they boost the choice frequencies of other alternatives. To see this, consider alternative x and assume that there exists a distinct alternative y such that $x\mathcal{N}y$. We have $\rho(y|\{y\},\{x\}) = 1 - (1-\pi^n_x)(1-\pi^f_y)$, which implies

$$\pi^n_x = 1 - \frac{1 - \rho(y|\{y\},\{x\})}{1 - \pi^f_y}.$$

[26] Specifically, suppose that for all $C \subsetneq A$, the ratio $\frac{\zeta(C)}{\zeta(\emptyset)}$ is already pinned down. Then we have $\frac{\zeta(A)}{\zeta(\emptyset)} = \frac{\sum_{B \subseteq A: B \neq \emptyset} \zeta(B)}{\zeta(\emptyset)} - \frac{\sum_{C \subsetneq A: C \neq \emptyset} \zeta(C)}{\zeta(\emptyset)} = \frac{1 - \Phi(A,\emptyset)}{\Phi(A,\emptyset)} - \frac{\sum_{C \subsetneq A: C \neq \emptyset} \zeta(C)}{\zeta(\emptyset)}$.

[27] The definitions of π^f_A, π^n_A, $\mathring{\pi}^f_A$ and $\mathring{\pi}^n_A$ are similar to those of π_A and $\mathring{\pi}_A$ in Section 3.

We note that π^n cannot be fully identified: For a given alternative x, if every other alternative is not associated with it, then we are unable to identify π^n_x. Nevertheless, in this case, since the consideration of x does not prompt the consideration of any other alternative, the value of π^n_x is irrelevant.

References for Online Appendix

Brady, R. L. and J. Rehbeck (2016): "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84, 1203–1223.