Attention, Memory, and Preferences*

Yusufcan Masatlioglu† Ece Yegane‡

November 7, 2024

Abstract

We study how the allocation of attention to alternatives and the accessibility of alternatives from memory affect decision making. To distinguish between attention and memory, we propose a two-stage stochastic consideration set formation process. An alternative enters the decision maker's consideration set if it is investigated in the initial attention stage and is remembered in the subsequent recall stage. In the initial attention stage, the decision maker investigates each available alternative with some alternative-specific probability. In the recall stage, the decision maker remembers each alternative that she investigated in the attention stage with some probability. The probability of remembering an alternative depends on the memorability of the alternative and its position in the order of investigation in the attention stage. In particular, investigating an alternative more recently enhances the probability of recalling it. The decision maker chooses the alternative that maximizes her preference relation over her consideration set. Under the assumption that the investigation of alternatives is observable, we provide testable implications on choice behavior and show that preferences, attention parameters, and memory parameters can be uniquely identified from observable repeated choices.

*We are grateful to Chris Chambers, Paul Cheung, Emel Filiz-Ozbay, Semih Kara, Erkut Ozbay, Chris Turansick, Neslihan Uler, Tri Phu Vu, Kemal Yildiz, and seminar participants at the University of Maryland and Georgetown University for helpful comments. Ece gratefully acknowledges financial support from the Institute for Humane Studies through grant IHS017398.

†University of Maryland, Department of Economics, Tydings Hall 3114, 7343 Preinkert Drive, College Park, MD 20742. Email: yusufcan@umd.edu

‡The Brattle Group. Email: eceyegane@gmail.com

1 Introduction

Decision makers frequently encounter choices with an abundance of alternatives. Assessing each alternative in such complex decisions demands considerable time and effort. Over the last fifty years, theoretical and empirical research has demonstrated that many consumers effectively consider a small number of alternatives before making a final purchase decision (for an overview, see Honka et al. [2019]). In particular, the marketing literature has long recognized that the actual choice process of the consumers can be seen as a "purchase funnel" consisting of three stages: brand awareness, brand recall, and choice (e.g., Shocker et al. [1991]). One of the main findings from this literature is that awareness and recall are as important as consumption utility or preferences in influencing choices.

In contrast to the marketing literature, researchers in economics generally do not differentiate between awareness and recall, grouping both into what is termed the consideration set. In economics, the prevailing explanation for the composition of a consideration set is that it comprises alternatives that the decision maker pays attention to or is aware of. However, this explanation may refer to all of the available alternatives that the decision maker knows about, or it can pertain to the set of alternatives that comes to mind at a particular time. These two situations represent different facets of the cognitive processes affecting decision making. Specifically, increasing awareness demands grabbing the consumers' attention, while boosting recall requires strengthening memory retrieval. Thus, awareness refers to general knowledge about which alternatives are out there, whereas recall refers to which alternatives come to mind at the time of decision.

Brand awareness and brand recall together determine consumers' final consideration set. One may argue that identifying the final consideration set is sufficient for economic analysis and it is not necessary to explicitly model brand awareness and brand recall separately. However, distinguishing between how awareness and recall shape consideration sets has economically relevant practical implications. To demonstrate, consider a firm or a marketer with an aim to increase the sales of a product. Distinguishing between awareness and recall is crucial for a marketer because these two stages can be influenced by different marketing efforts. For example, by analyzing whether a product is stuck at the awareness stage or at the recall stage, a marketer can determine whether the firm should increase the brand's visibility to reach new consumers (*i.e.* boost awareness) or enhance the brand's memorability to re-engage consumers who have already been exposed to the brand (*i.e.* boost recall).[1] This exercise necessitates knowledge about the separate influence of awareness and recall on the composition of the consideration set.

Since these cognitive processes are in the mind of the decision maker, an outside analyst cannot observe which alternatives are paid attention to and which alternatives are remembered. Given the intertwined nature of attention and memory, one might wonder whether it is even possible to distinguish the effects of memory and attention on decisions from solely observed choices. Since preferences also shape choices and are also unobservable, separating attention, memory and preferences becomes an ambitious question. To answer this question, one needs a model of decision making that has both attention and memory components. In this paper, we provide such a model. In particular, we make the distinction between alternatives the decision maker knows about and alternatives that the decision maker can remember at the time of choice. We provide a parametric model in which attention and memory are captured through distinct parameters.

In our model, the decision maker maximizes a preference relation over the set of alternatives she was initially aware of and she can recall at the time of choice. Due to imperfections in the attention and memory processes, this set may be different than the entire set of alternatives, especially when the number of available alternatives is large. In this case, the decision maker may make choice errors. We attribute the choice errors that arise due to not knowing some of the alternatives exist to *the attention channel*, and the choice errors that arise due to not remembering some of the alternatives at the time of choice to *the memory channel*. Thus, the choices in our model are generated from the interaction between attention, memory, and preferences.

[1] Indeed, the 2023 Nielsen Brand Lift Report, which measures the effectiveness of marketing campaigns, ranks brand recall as the most important factor followed by baseline brand awareness. See the article titled "In emerging media, brand recall is the biggest driver of lift" at Nielsen.com.

We model attention and memory as stochastic. This modeling choice captures the observation that the set of alternatives the decision maker is aware of or can recall may change when the same choice problem is repeated. This randomness may be caused by changes in the state of the world that are not observable by the analyst analyzing choice data, but may affect the allocation of attention and memory. For example, a consumer engaging in online shopping may search for alternatives on her phone or her computer. Because of the differences in the screen size and the presentation of alternatives, the consumer may pay attention to a different subset of alternatives on her phone versus on her computer, but this information may not be available to the analyst. The changes in unobserved states can also be due to changes in cognitive factors that affect attention and memory such as emotional states or cognitive load.

Following the purchase funnel conceptualization that is popular in marketing, we model the decision making process of the decision maker as a three-stage choice procedure, each stage corresponding to one of the three elements present in our model. In the first stage, given a set of available alternatives, the decision maker initially possesses knowledge of a subset of these alternatives. We refer to this set as the *awareness set* and attribute it to the impact of attention. To describe this attention-driven process, we use the random attention rule proposed by Manzini and Mariotti [2014].[2] In this model, the composition of awareness sets is determined stochastically, based on the degree of attention the decision maker allocates to each alternative. Here, the parameter $\gamma(x)$ captures the level of attention assigned to alternative x. Thus, given a set of available alternatives, the first stage generates a probability distribution over awareness sets.

The second stage describes the memory-driven process following the random recall rule in Yegane [2022]. In the second stage, given an awareness set, the decision maker sequentially observes the alternatives in her awareness set. This assumption captures the sequential nature of information processing. We refer to this order of observing the alternatives as the *list*. We assume that the same list generates the entire set of observable choices and the list is observable to the analyst. One way to visualize this assumption is by considering the products ordered in a grocery store aisle. The assump-

[2]This parametric model is widely used in empirical applications (Abaluck and Adams-Prassl [2021], Barseghyan et al. [2021a,b]).

tion on the observability of the list assumes that the shopper follows the order presented by the store, and the how the products are placed is known by the analyst. However, due to randomness in awareness, the shopper may not observe all of the products in the aisle - she may skip some of the products. Whether a product was skipped is not assumed to be observed by the analyst, and we show this information can be revealed from the choices.

The probability of remembering an alternative is determined by that alternative's base memorability and its position in the list. Intuitively, sequentially observing the alternatives may cause the decision maker to forget some of the alternatives in her awareness set that she observed earlier. In the model, whenever the decision maker encounters a new alternative in the list, she may forget alternatives she has already seen with some probability. The parameter $q(x)$ is the probability of recalling x when the next alternative in the list is observed. This parameter, $q(x)$, captures the base memorability of an alternative x, absent of any order effects. As the decision maker continues observing the alternatives in the list, each new alternative explored creates a possibility for previously observed alternatives to be forgotten, thereby discounting the probability of remembering x further.

We refer to the set of alternatives that the decision maker was initially aware of and successfully recalled at the time of making a choice as the *consideration set*. In the third stage, the decision maker maximizes a well-behaved preference relation over the consideration set. Given a set of available alternatives, the random attention rule and the random recall rule together generate a probability distribution over possible consideration sets.

We restate the question we posed above in the context of our model. There are three possible reasons why an alternative is not chosen. First, there may be a more preferred alternative available. Second, the alternative may not be considered because the decision maker was never aware of it. Third, the alternative may not be considered even though the decision maker was aware of it at some point but she could not recall it at the time of choice. Our first set of results show that we can distinguish these three possibilities from each other. The three sets of unobservables in our model, preferences, awareness sets and consideration sets, can be uniquely identified from repeated choices when the list is observable.

In our identification exercise, we utilize choices from sets with only one or two alternatives. We show that the choice probability of x when only x is available is equal to the choice probability of x when x and y are available if and only if x is preferred to y and x appears in the list after y. Thus, even when the underlying list is unobservable, our identification results guide what type of preferences are permissible, by analyzing singleton and binary choices. Moreover, we show that if the choice probability of both x and y change from singleton sets to the set in which both x and y are available, then the underlying preference relation has to coincide with the list order. Thus, when the list is observable, we can fully reveal the preference ranking of x and y.

To better understand the relationship between attention, memory, and preferences, we study the behavioral foundations of the limited attention and memory model. In particular, we provide necessary and sufficient conditions for repeated choices associated with a list to have a limited attention and memory representation. Our characterization exercise shows that a key implication of the limited attention and memory model is that when some conditions on the preference and list ranking of x are satisfied, the event of choosing x in some set S can be divided into independent events of choosing x in subsets of set S. In other words, the act of choosing x in S can be divided into smaller sub-problems such that each of those problems can be solved independently, and the solutions of the sub-problems can be combined to form the choice probability of x in S. Our characterization exercise provides insights into the alternatives suitable for applying the "divide-and-conquer" approach.

The rest of the paper is organized as follows. Section 2 introduces the notation and formally describes the model. Section 3 provides the identification results. Section 4 provides the characterization result. Section 5 compares the choice behavior generated by the model with related models of stochastic choice. Section 6 concludes.

2 Model

Let X be a non-empty finite set representing the set of alternatives that may be available to the decision maker. Let $\mathcal{X}$ denote all nonempty subsets of X. Any element S of $\mathcal{X}$ constitutes a choice set. There is an outside option o that the decision maker can opt for when presented with any choice set S. Thus, the decision maker is effectively choosing a single alternative from $S \cup \{o\}$ when presented with the alternatives in S. We interpret the outside option as not choosing any of the alternatives in S.

We consider a decision maker who chooses from the same choice set repeatedly. We assume to observe the limiting distribution of the decision maker's repeated choices from each choice set, which we refer to as a *random choice rule*. Formally, a random choice rule p maps each choice set $S \in \mathcal{X}$ to a probability distribution over its elements and the outside option. We denote the probability of choosing alternative x from choice set S with $p(x, S)$. We assume that for all $S \in \mathcal{X}$, $x \in S$, $p(x, S) > 0$.[3] For choice sets with only a single alternative available, we abuse notation and drop the parentheses when denoting the singleton set, *i.e.* instead of $p(x, \{x\})$ we use $p(x, x)$. Similarly, we abuse notation and use $S \setminus x$ instead of $S \setminus \{x\}$.

In our model, three channels influence decision making: attention, memory, and preferences, in this particular order. We describe the decision making process of the decision maker as a three-stage choice procedure where each stage corresponds to one of the channels affecting decision making. Figure 1 provides a graphical representation of our model.

Given a choice set S, the decision maker is initially aware of a subset of alternatives $A \subseteq S$. We refer to this set A as the *awareness set*, which captures the effect of attention. To capture attention, we utilize the parametric attention rule of Manzini and Mariotti [2014], in which awareness sets are modeled probabilistically. Whether an alternative is included in the awareness set is determined by the level of attention the decision maker pays to it. The parameter $\gamma(x)$ denotes the level of attention alternative x receives. This

[3] As noted by McFadden [1973], a zero probability is empirically indistinguishable from a very small positive probability.

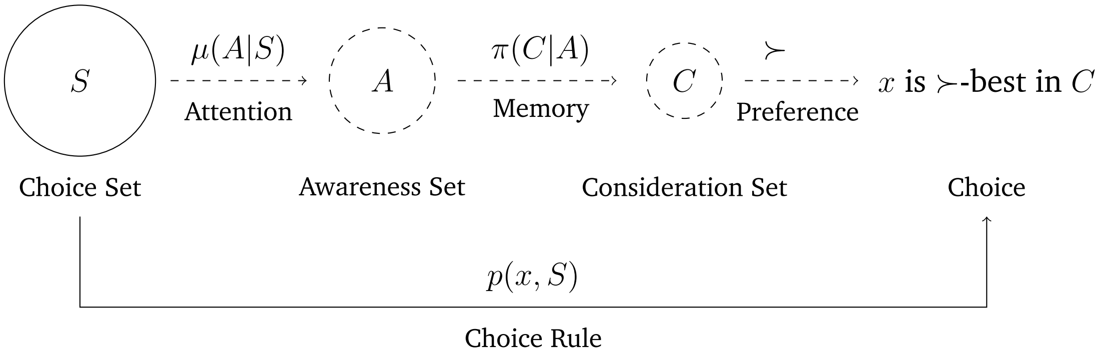

Figure 1: An illustration of our model. *Observable*: choice set and choice rule (solid line). *Unobservable*: awareness set, consideration set, and preference (dashed line).

parameter is assumed to be strictly between 0 and 1 for all alternatives in X. Given a choice set S, the awareness set can be equal to each subset of S with some probability. The probability of $A \subseteq S$ being the awareness set when the choice set is S is given by,

$$\mu(A|S) := \prod_{x \in A} \gamma(x) \prod_{y \in S \setminus A} (1 - \gamma(y))$$

In other words, the probability of the awareness set being equal to A when the choice set is S is given by probability of paying attention to every alternative in A and not paying attention to every alternative outside of A. Note that the awareness set A can be empty. In this case, we assume that the decision maker chooses the outside option o.

If the awareness set A is nonempty, then the decision maker observes the alternatives in A sequentially. The order in which the decision maker observes the alternatives is captured by a *list*. Formally, a list $\rhd$ is a strict total order over X. Let $\rhd(A)$ denote the ordering that is obtained from $\rhd$ by deleting the alternatives in $X \setminus A$. For any $x \in A$, $\rhd(x, A)$ denotes the number of alternatives that follow x in $\rhd(A)$. We write $x \rhd y$ to denote that x appears in the list before y. We assume that each random choice rule p is associated with a single list $\rhd$, and the list is observable to the analyst. Note that because awareness sets are stochastic, the actual order followed by the decision maker is a stochastic object. For example, if products are ordered so that x is the first alternative in the list, y is second, and z is last, then the actual order followed by the decision maker is equal to $x \rhd y \rhd z$ if the decision maker initially pays attention to all three products. However, if the decision maker does not pay attention to y, then the actual order followed by the decision maker is x and then z. The variations in the list due to randomness of awareness is not assumed to be observable. But the analyst knows that the decision maker cannot first observe z and then x.

Sequentially observing the alternatives in A may cause the decision maker to forget some of the alternatives in A. To capture memory failures, we utilize the recall procedure in Yegane [2022]. In particular, each time the decision maker observes a new alternative, this creates a possibility for previously observed alternatives to be forgotten. In addition to this order effect, each alternative x is also associated with a base memorability $q(x)$, which captures the probability of remembering x when the next item on the list is observed. We refer to $q(x)$ as the period recall probability of x and assume that it is strictly between 0 and 1. This recall procedure captures stochastic memory through an order effect and an alternative-specific memorability effect. To illustrate, suppose that there are three alternatives in the decision maker's awareness set, ordered in the list as $x \rhd y \rhd z$. The last alternative the decision maker observes is z so it is recalled with probability one. The decision maker observes z after observing y, so with probability $q(y)$ the decision maker recalls y. The decision maker observes y and then z after observing x, so x is recalled with probability $q(x)^2$. We refer to the memory probability of each alternative in the respective awareness sets as the final recall probability. Note that as opposed to the period recall probability, the final recall probabilities include the order effects on memory. We refer to the set of alternatives the decision maker was initially aware of and recalled at the time of choice as the *consideration set*. Given an awareness set A, the consideration set can be equal to subsets of A that include the last alternative in the list $\rhd(A)$ with some probability. Note that, the probability of the consideration set being equal to a subset of A that does not include the last alternative in $\rhd(A)$ is zero. In general, the probability of $C \subseteq A$ being the consideration set when the awareness set is A is given by,

$$\pi(C|A) = \prod_{x \in C} q(x)^{\rhd(x,A)} \prod_{y \in A \setminus C} (1 - q(y)^{\rhd(y,A)})$$

Thus, given an awareness set A, the probability of the consideration set being equal to C is given by the probability that each alternative in C is recalled and each alternative in $A \setminus C$ is forgotten. The generation of the consideration set from an awareness set is the second stage of the choice procedure. The exclusion of the forgotten alternatives is what we refer to as the *memory channel*. Note that by construction, being initially aware of an alternative is necessary for an alternative to be remembered later on.

In the final stage of the choice procedure, the decision maker maximizes a preference relation $\succ$ over the alternatives in the decision maker's consideration set. We assume that the decision maker's preferences can be represented by a strict total order. We now formally define our model.

Definition 1 A random choice rule p associated with list $\rhd$ has a *limited attention and memory (LAM)* representation if there exists a preference ordering $\succ$ on X, an attention function $\gamma : X \to (0,1)$, and a period recall probability function $q : X \to (0,1)$, such that for any $x \in S$ and $S \in \mathcal{X}$,

$$p(x,S) = \sum_{A \subseteq S} \underbrace{\left[\prod_{y \in A} \gamma(y) \prod_{z \in S \setminus A} (1-\gamma(z)) \right]}_{\text{awareness set is } A} \underbrace{q(x)^{\rhd(x,A)}}_{x \text{ is recalled}} \underbrace{\prod_{\{t \in A \mid t \succ x\}} (1 - q(t)^{\rhd(t,A)})}_{\text{alternatives that are preferred to } x \text{ are forgotten}}$$

Three conditions must be satisfied for an alternative $x \in S$ to be chosen in S. First, the decision maker must be initially aware of alternative x, so x must be in the awareness set. Second, alternative x must be recalled at the time of choice, so x must be in the consideration set. Finally, x must be the decision maker's most preferred alternative among the ones in her consideration set. Definition 1 formalizes this procedure. The choice probability of x in choice set S is given by the sum of probabilities of all consideration sets in which x is the decision maker's most preferred alternative. The consideration set probabilities are given by the attention and memory procedure described above.

In the limiting case where $q(x) = 1$ for all $x \in X$, LAM boils down to the model of Manzini and Mariotti [2014].[4] In the limiting case where $\gamma(x) = 1$ for all $x \in X$, the

[4]Note that if the preference relation and the list are exactly the opposite, *i.e.* for all $x, y \in X$ $x \succ y$ if and only if $y \rhd x$, then LAM makes the same predictions as Manzini and Mariotti [2014]. Thus, LAM subsumes Manzini and Mariotti [2014].

model boils down to Yegane [2022]. If $q(x) = 1$ and $\gamma(x) = 1$ for all $x \in X$, then the model makes the same predictions as rational deterministic choice.

We now provide some intuition about how LAM works. We first introduce the well-known regularity property on random choice rules. A random choice rule p satisfies regularity if for all $S \in \mathcal{X}$ with $x \in S$ and $y \notin S$,

$$p(x, S) \geq p(x, S \cup \{y\})$$

The regularity property says that adding an alternative y to a choice set cannot increase the choice probability of another alternative x. One way to think about violations of regularity is to think of them as the probabilistic equivalent of deterministic choice reversals. Suppose that the decision maker's choice behavior is deterministic, so for any $p(x, S)$ is either equal to 0 or 1. Then a violation of regularity corresponds to alternative x not being chosen in S, but it is chosen in $S \cup \{y\}$. This choice behavior violates Sen's property α and it cannot be rationalized by the standard deterministic model of choice. Similarly, violations of regularity with random choice rules cannot be rationalized by the random utility model (Block and Marschak [1960]). However, studies document systematic violations of regularity. Huber et al. [1982] is the first one to document such violations and Rieskamp et al. [2006] provide a review of regularity violations in various domains.

In what follows, we first show that LAM can rationalize regularity violations with a simple example. We then discuss under which conditions on attention and memory parameters the model predicts this behavior.

Example 1 Consider the following random choice rule, generated from LAM with $x \succ y \succ z$, $x \rhd y \rhd z$, $\gamma(x) = \gamma(y) = \gamma(z) = 0.9$, $q(x) = 0.8$, $q(y) = q(z) = 0.9$.[5]

S	$p(x, S)$	$p(y, S)$	$p(z, S)$	$p(o, S)$
$\{x, y\}$	0.74	0.25	-	0.01
$\{x, y, z\}$	0.61	0.33	0.06	0

[5] Note that we round the numbers to the nearest hundredth. Thus, $p(o, \{x, y, z\})$ is a small but positive number.

When z is added to the choice set $\{x, y\}$, the choice probability of y increases from 0.25 to 0.33. Therefore, this random choice rule violates regularity. The explanation behind this prediction is as follows. As z appears in the list after x and y, adding it to the choice set $\{x, y\}$ decreases the final recall probability of x and y when the awareness set is $\{x, y, z\}$. This decrease in the final recall probability of y causes a decrease in the choice probability of y. On the other hand, the decrease in the final recall probability of x causes an increase in the choice probability of y. This is because x is preferred to y, so x must be not paid attention to or paid attention to and forgotten for y to be chosen, and the presence of z decreases the probability of y being considered by pushing its location in the list. When the effect on the final recall probability of x is stronger, the model predicts a regularity violation. Thus, the model rationalizes regularity violations via changes in recall probabilities.

To better understand the interaction of attention and memory in generating this result, we now inquire for which attention and memory parameter values the model predicts a regularity violation in y when moving from set $\{x, y\}$ to $\{x, y, z\}$ given that $x \succ y \succ z$, $x \rhd y \rhd z$. We first focus on for which parameter values $p(y, \{x, y, z\}) - p(y, \{x, y\})$ is strictly positive. We then discuss how the size of the regularity violation changes in response to the changes in the parameters.

First of all, whether there is a regularity violation in this setup depends only on parameters $q(x), q(y), \gamma(x)$. Because z is the last alternative, it's period recall probability does not affect choices. Parameters $\gamma(y)$ and $\gamma(z)$ affect the size of the regularity violation, but they do not affect whether there is a violation. Essentially, both $p(y, \{x, y\})$ and $p(y, \{x, y, z\})$ are linear in $\gamma(y)$. As long as $\gamma(z)$ is strictly positive, the model creates the trade-off between the final recall probability of x and y described above, even when $\gamma(z)$ is very small.

In Figure 2, we fix a value for $\gamma(x)$, and then demonstrate the $q(x)$ and $q(y)$ combinations for which the model predicts a regularity violation in y. The purple area corresponds to $q(x)$ and $q(y)$ combinations that predict a regularity violation when $\gamma(x) = 0.75$. When $\gamma(x) = 0.95$, $q(x)$ and $q(y)$ combinations in the purple and blue areas predict a regularity violation. The first observation is that $q(y)$ must be sufficiently

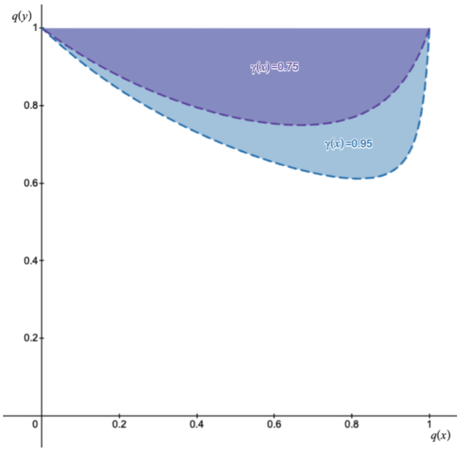

Figure 2: $q(x)$-$q(y)$ combinations that predict a regularity violation in y

high, otherwise the model does not predict a regularity violation in y. If alternative y is not memorable, then the decrease in the choice probability of y when z is added to the choice set becomes very high.

The second observation is, given $\gamma(x)$ and $q(y)$, a regularity violation in y is observed when $q(x)$ take moderate values. This is because when $q(x)$ is very high, then the final recall probability of x is very close to 1, regardless of where x is placed in the list. Thus, the decrease in the final recall probability of x due to the addition of z becomes negligible and the model does not predict a regularity violation. On the other hand, when $q(x)$ is sufficiently small, the decrease in the final recall probability of x due to the addition of z also becomes very small. Thus, for sufficiently small $q(x)$ values, the model does not predict a regularity violation.

The third observation is that if $\gamma(x)$ decreases, then the sets of $q(x)$ and $q(y)$ combinations for which the model predicts a regularity violation shrinks. To see why, consider the limiting case in which $\gamma(x) = 0$. Then, in the choice set $\{x, y\}$, $\{y\}$ is the only awareness set assigned strictly positive probability in which y can be chosen. Similarly, in the choice set $\{x, y, z\}$, $\{y\}$ and $\{y, z\}$ are the two awareness sets assigned strictly positive probability in which y can be chosen. In awareness set $\{y\}$, alternative y is recalled with probability one, thus it is chosen with probability one. However, in awareness set $\{y, z\}$, as z appears in the list after y, alternative y is recalled with probability $q(y)$ which is less than one. Thus, adding z to the choice set unambiguously decreases the choice probability of y in the limiting case where the decision maker is completely unaware of alternative x. When $\gamma(x)$ is in $(0,1)$ and the choice set is $\{x,y,z\}$, the decision maker faces awareness sets $\{x,y,z\}$ and $\{y,z\}$ with positive probability. The choice probability of y in awareness set $\{x,y,z\}$ may be higher than the choice probability of y in awareness set $\{x,y\}$ as z increases the probability that x is forgotten. As $\gamma(x)$ increases, the probability of facing the awareness set $\{x,y,z\}$ increases while the probability of facing the awareness set $\{y,z\}$ decreases, so lower $q(y)$ values may generate regularity violations. In other words, whether there is a regularity violation critically depends on the decision maker being initially aware of alternative x, but forgetting it due to the presence of another alternative z.

We now discuss how the size of the violation changes in response to a change in model parameters, conditional on observing a regularity violation. Due to the intuition described above, the size of the regularity violation is increasing in $q(y)$ and $\gamma(x)$, while initially increasing and then decreasing in $q(x)$. Conditional on observing a regularity violation, the size of the violation is increasing in $\gamma(y)$ and $\gamma(z)$. This is because an increase in $\gamma(y)$ creates a level effect and boosts the choice probability of y in every choice set. On the other hand, an increase in $\gamma(z)$ causes the final recall probability effects in alternatives x and y to be stronger, thereby boosting the size of the violation.

3 Identification

In this section, we assume that the random choice rule p associated with list $\rhd$ has a limited attention and memory representation. We inquire whether we can identify the preference relation, the attention function and the period recall probability function from $(p,\rhd)$. We first define a binary relation based on singleton and binary choices:

$$xPy \text{ if one of the following is observed: } (i)\, p(x,x) = p(x,\{x,y\}), \text{ or}$$
$$(ii)\, p(y,y) \neq p(y,\{x,y\}) \text{ and } x \rhd y.$$

The only instance in which the choice probability of x stays constant when another alternative is added to the choice set is if x is preferred to the new alternative and the new alternative does not affect the probability of x being recalled, which means that the new alternative appears in the list before x. In this case, x is chosen in the choice set $\{x, y\}$ if the decision maker is initially aware of it. Thus, $P(i)$ reveals that x is preferred to y.[6] Similarly, if y appears in the list after x, then adding x to the choice set does not affect the final recall probability of y. If adding x changes the choice probability of y, then it must be that the choice probability of y is affected by whether x is remembered, which reveals that x is preferred to y.

If alternative x is the only available alternative, then it is recalled with probability one if the decision maker initially paid attention to it. Thus, the choice probabilities from singleton sets reveal the attention probabilities.

$$\gamma(x) = p(x, x) \tag{1}$$

If alternative x is revealed to be preferred to y and $x \rhd y$, we define

$$q(x) = \frac{p(x, \{x, y\}) - p(x, x)(1 - p(y, y))}{p(x, x)p(y, y)} \tag{2}$$

Since x is preferred to y and x appears in the list before y, x is chosen in $\{x, y\}$ if x is the only alternative in the awareness set or both x and y are in the decision maker's awareness set and x is recalled. To reveal $q(x)$, we need to isolate the choice probability of x when the awareness set is $\{x, y\}$ in $p(x, \{x, y\})$. Thus, we subtract the choice probability of x when the awareness set is $\{x\}$ in choice set $\{x, y\}$, which is given by $p(x, x)(1 - p(y, y))$. What remains is the probability of recalling x in the set $\{x, y\}$ conditional on both x and y being in the awareness set.

At this point, we wish to emphasize that recall failures affect choices only when the preference order and the list order coincide. To see why, consider the situation in

[6] Equivalently, we can write $P(i)$ as $p(x, x) = p(x, \{x, y\})$ and $y \rhd x$. However, the latter part is implied by the former because if x appears in the list before y, then the choice probability of x has to change when y is added to the choice set. See the discussion following Axiom 1.

which the preference relation and the list are exactly the opposite: For all $x, y \in X$, $x \succ y$ and $y \rhd x$. In this case, the stochastic choice function is independent of the period recall probability function $q(.)$. In any awareness set, the decision maker recalls the last alternative in the list, which is her most preferred alternative in that awareness set, with probability one. Therefore, the decision maker chooses the alternative she recalls with probability one in any awareness set. This means that the decision maker's choices are entirely governed by the initial awareness of the alternatives. Recall failures do not affect choices because only less preferred alternatives are forgotten. Now suppose that the preference relation $\succ$ and the list $\rhd$ are exactly opposite except for the ranking of two alternatives $i, j \in X$ such that $i \succ j$ and $i \rhd j$. Now consider an awareness set that includes i and j but does not include any of the alternatives that are preferred to i. In this awareness set, any alternative that is inferior to j is not chosen, alternative i is chosen if it is recalled and alternative j is chosen if i is forgotten. Now consider an awareness set that includes at least one alternative that is preferred to i. Then, the last alternative in the list is preferred to i, and is recalled with probability one, so neither i nor j can be chosen in such awareness sets. Hence, the stochastic choice function is independent of all $q(.)$ parameters except $q(i)$. In this case, we can reveal $q(i)$ from Equation 2. In other words, alternative i's period recall probability $q(i)$ appears in the model only if there exists another alternative j such that $i \succ j$ and $i \rhd j$, and in this case Equation 2 can be used to identify $q(i)$. Thus, we can uniquely reveal $q(i)$ as long as it is relevant for the model.

3.1 Identification with Limited Data

The identification strategy we described above explains a method to uniquely identify the preference relation as well as the attention and memory parameters from choice sets when only one or two alternatives are available. We now discuss a way of identifying the underlying preference relation from choice sets with three or more available alternatives. For a choice set $S \in \mathcal{X}$, let $\rhd_1$ denote the first alternative in the associated list $\rhd(S)$. Note that the first alternative in the list $\rhd(S)$ depends on choice set S. We use the simplified notation $\rhd_1$ as long as it does not create any confusion. We consider choice set variations in which the first alternative $\rhd_1$ in the list $\rhd(S)$ is removed from choice set S. For any $x \in S \setminus \rhd_1$ define a binary relation $\overline{P}$,

$$x\overline{P}\rhd_1 \text{ if } p(x,S) = p(x, S \setminus \rhd_1)$$

$$\rhd_1\overline{P}x \text{ if } p(x,S) \neq p(x, S \setminus \rhd_1)$$

As $\rhd_1$ is the first alternative in the list, removing it does not change the final recall probability of any of the alternatives in $S \setminus \rhd_1$ in any awareness set. Thus, if an alternative x is preferred to $\rhd_1$, its choice probability must remain unchanged when $\rhd_1$ is removed. If an alternative is inferior to $\rhd_1$, then it can be chosen either if $\rhd_1$ is not initially paid attention to, or if it is forgotten. Thus, the choice probability of such alternatives increase when $\rhd_1$ is removed. This means that we can fully reveal the ranking of $\rhd_1$ within S from choice sets S and $S \setminus \rhd_1$. For any $x \in S$ that is inferior to $\rhd_1$, we can reveal its period recall probability by calculating the ratio in part (i) of Axiom 6.[7]

4 Characterization

In this section, we provide necessary and sufficient conditions for a random choice rule and an associated list $\rhd$ to have a limited attention and memory representation. In the identification section, we provide a method for identifying the unobservable variables that shape the behavior of the decision makers from the observables, *i.e.,* we show that we can identify the preference relation, the attention function and the period recall probability function from the random choice rule and the associated list. However, the identification results are applicable if the random choice rule is generated from the limited attention and memory model. Our characterization exercise addresses this gap by presenting testable conditions that link the model to the identification results.

The first axiom of our characterization result says that if alternative y is followed by

[7]Note that the least preferred alternative's period recall probability is not relevant for the model. This is because the last alternative in the list is recalled with probability one, so the least preferred alternative can only be chosen if it is the last one in the list.

another alternative x in the list, then the choice probability of y must decrease when x is added to the choice set $\{y\}$.

Axiom 1 (Asymmetry) If $y \triangleright x$, then $p(y, y) > p(y, \{x, y\})$.

If alternative y is followed by another alternative x in the list, then in the awareness set $\{x, y\}$, the presence of alternative x decreases the final recall probability of alternative y. Hence, adding x to the choice set $\{y\}$ decreases the choice probability of y, regardless of the preference ranking of x and y. Note that Axiom 1 implies we cannot have $p(y, y) = p(y, \{x, y\})$ and $p(x, x) = p(x, \{x, y\})$ at the same time, which ensures that P is asymmetric.

The next axiom says that the revealed preference relation has to be transitive.

Axiom 2 (Transitivity) P is transitive.

Due to the definition of P, the implication of transitivity is not straightforward. To illustrate, suppose that xPy and yPz. Axiom 2 indicates that either $p(x, x) = p(x, \{x, z\})$ or $p(z, z) \neq p(z, \{x, z\})$ and $x \triangleright z$. Thanks to Axiom 1, we know more about this. Assume that both these revelations originated from part (i) of the definition of P, *i.e.*, $p(x, x) = p(x, \{x, y\})$ and $p(y, y) = p(y, \{y, z\})$. Then, Axiom 1 yields $z \triangleright y \triangleright x$. The transitivity of the list order implies that we must have $p(x, x) = p(x, \{x, z\})$. Hence, xPz follows from part (i) (not part (ii) which would imply that $p(z, z) \neq p(z, \{x, z\})$ and $x \triangleright z$).

Similarly, if xPy and yPz both follow from part (ii) of the definition of P, *i.e.*, $p(y, y) \neq p(y, \{x, y\})$, $p(z, z) \neq p(z, \{y, z\})$ and $x \triangleright y \triangleright z$, then Axiom 1 and the transitivity of $\triangleright$ imply that $p(x, x) \neq p(x, \{x, z\})$. Therefore, xPz must follow from part (ii) of the definition, which requires that $p(z, z) \neq p(z, \{x, z\})$ and $x \triangleright z$.

Note that P is complete by definition. For any two alternative x and y, either $p(x, x) = p(x, \{x, y\})$ or $p(x, x) \neq p(x, \{x, y\})$. If $p(x, x) = p(x, \{x, y\})$, then xPy by $P(i)$. If $p(x, x) \neq p(x, \{x, y\})$ and $y \triangleright x$, then yPx by $P(ii)$. If $p(x, x) \neq p(x, \{x, y\})$ and $x \triangleright y$, then we also need to check the relation between $p(y, y)$ and $p(y, \{x, y\})$. If $p(y, y) = p(y, \{x, y\})$ then yPx by $P(i)$. If $p(y, y) \neq p(y, \{x, y\})$, then xPy follows from $P(ii)$, since $x \triangleright y$. Thus, Axiom 1 and 2 together guarantee that P is a linear order.

The next axiom guarantees that the impact of forgetting an alternative x on the choice probability of x in binary choice sets is consistent across two inferior alternatives y and z. In other words, forgetting x influences its choice probability uniformly in binary choice sets, independent of the specific inferior options available.

Axiom 3 (Consistent Recall) If xPy, z and $x \rhd y, z$ then,

$$\frac{p(x, \{x, y\}) - p(x, x)(1 - p(y, y))}{p(x, x)p(y, y)} = \frac{p(x, \{x, z\}) - p(x, x)(1 - p(z, z))}{p(x, x)p(z, z)} > 0$$

If there are two or more alternatives that are inferior to x and follow x in the list, then the impact of forgetting x on the choice probability of x in binary choice sets must be consistent across two inferior alternatives. Axiom 3 addresses this by guaranteeing that the period recall probability of x implied by Equation 2 from each of the inferior alternatives must be the same. In other words, Axiom 3 ensures that the period recall probability function is well-defined.

Additionally, we rule out always-forgotten and always-recalled alternatives by assumption, so each period recall probability must be in $(0, 1)$. Axiom 3 handles this assumption in the following manner. First, notice that Axiom 3 has $q(x) > 0$ directly embedded in. Second, to see that $q(x) < 1$ is implied by Axiom 3, we can rewrite Equation 2 as,

$$q(x) = 1 + \frac{p(x, \{x, y\}) - p(x, x)}{p(x, x)p(y, y)}$$

As $x \rhd y$, Axiom 1 implies that $p(x, x) > p(x, \{x, y\})$ and thus $q(x) < 1$.

The next axiom relates the choice probability of an alternative in some choice set with at least three alternatives to that alternative's choice probability in subsets of that larger choice set. The idea behind the axiom is to isolate the choice probability of an alternative when the decision maker's awareness set includes every available alternative from the observed choice probability in a given choice set. The goal is to isolate the change in the choice probability of an alternative due to memory limitations from those due to attention limitations, in response to systematic choice set variations. Isolating the portion of the choice probability in which every option is paid attention to essentially shuts down the attention channel, allowing us to focus on the memory channel effects. To illustrate, suppose that $(p, \triangleright)$ has a limited attention and memory representation. Then,

$$\begin{aligned} & p(x,\{x,y,z\}) - \frac{p(x,\{x,y\})p(x,\{x,z\})}{p(x,x)} \\ =& \gamma(x)\gamma(y)\gamma(z)q(x)^{\triangleright(x,\{x,y,z\})}\left[\prod_{\{t\in\{y,z\}|t\succ x\}}(1-q(t)^{\triangleright(t,\{x,y,z\})}) - \prod_{\{t\in\{y,z\}|t\succ x\}}(1-q(t)^{\triangleright(t,\{x,t\})})\right] \end{aligned} \tag{3}$$

Notice that the difference between $p(x,\{x,y,z\})$ and $\frac{p(x,\{x,y\})p(x,\{x,z\})}{p(x,x)}$ is equal to the choice probability of x when the decision maker's awareness set includes every available alternative minus an additional term that constructs the choice probability of x from the choice probability of x from sets $\{x,y\}$ and $\{x,z\}$, conditional on the decision maker being initially aware of every alternative. Notice that if x is the most preferred alternative in $\{x,y,z\}$, then

$$p(x,\{x,y,z\}) - \frac{p(x,\{x,y\})p(x,\{x,z\})}{p(x,x)} = 0$$

Thus, the choice probability of the most preferred alternative x can be written as the multiplication of two binary sets constructed in a way to cover every alternative in the ternary set and making sure x is available in both of the smaller sets.[8] In other words, if x is the decision maker's most preferred alternative in $\{x,y,z\}$, then the act of choosing x in $\{x,y,z\}$ can be divided into smaller sub-problems of $\{x,y\}$ and $\{x,z\}$, such that each of those problems can be solved independently, and the solutions of the sub-problems can be combined to form the choice probability of x in $\{x,y,z\}$. Thus, we can "divide-and-conquer" choosing alternative x in choice set $\{x,y,z\}$.

[8]Note that we need to divide the multiplication of binary choice probabilities by $p(x,x)$ because the outside option is always available, so it is double counted in sets $\{x,y\}$ and $\{x,z\}$.

Now suppose that $y \succ x$,

$$p(x, \{x, y, z\}) - \frac{p(x, \{x, y\})p(x, \{x, z\})}{p(x, x)} =$$
$$\gamma(x)\gamma(y)\gamma(z)q(x)^{\triangleright(x,\{x,y,z\})} \left[(1 - q(y)^{\triangleright(y,\{x,y,z\})}) - (1 - q(y)^{\triangleright(y,\{x,y\})})\right]$$

Because y is preferred to x, the choice probability of x also depends on the probability that y is forgotten, which can disrupt the divide-and-conquer procedure. However, if the probability that y is forgotten is the same in $\{x, y\}$ and $\{x, y, z\}$, then,

$$p(x, \{x, y, z\}) - \frac{p(x, \{x, y\})p(x, \{x, z\})}{p(x, x)} = 0$$

Thus, if the probability that y is forgotten is the same in $\{x, y\}$ and $\{x, y, z\}$, then we can divide-and-conquer choosing alternative x in choice set $\{x, y, z\}$ into smaller problems of choosing x in $\{x, y\}$ and $\{x, z\}$. Summarizing this yields the following observation.

Observation 1 *If x is revealed to be the most preferred option in $\{x, y, z\}$ or if y is revealed to be preferred to x, and y follows z in the list, then*

$$p(x, \{x, y, z\}) - \frac{p(x, \{x, y\})p(x, \{x, z\})}{p(x, x)} = 0$$

Note that Observation 1 is essentially a path independence condition as in Plott [1973]. In particular, the condition isolates memory channel effects and then characterizes when the change in choice probability due to memory channel effects satisfies path independence.

Our next axiom extends the idea behind Observation 1 to sets with more than three alternatives. Note that when there are only three available alternatives, dividing the choice set into smaller sub-problems is straightforward, as the only way to do it is to divide the choice set into two binary choice sets. However, when there are more than three alternatives, the division into sub-problems is not straightforward. Unlike the standard path independence condition, which essentially says that the division does not matter, the division matters within our model because of the memory channel effects. In particular, when a new alternative is added to a choice set, the addition of this alternative decreases the recall probability of alternatives that appear earlier in the list. This effect causes the choice probability of those alternatives that are pushed back to decrease. However, for inferior alternatives to be chosen, the decision maker has to forget the alternatives she prefers. Thus, pushing the alternatives in the list actually may boost the choice probability of some of the inferior alternatives, because the recall probability of some of the better alternatives decreases. When a divide-and-conquer procedure is applied, the division of the choice set into smaller subsets matter because of these memory channel effects.

In what follows, we describe a way of dividing any choice set S for any alternative $x \in S$ into smaller sub-problems to minimize the memory channel effects described above. This requires varying the choice set such that the alternatives that appear earlier in the list are removed, so that the changes in the recall probabilities of other alternatives are minimized.

Let $\rhd_1$ and $\rhd_2$ denote the first and second alternative in list $\rhd$, respectively. For a given list $\rhd$, choice set S and alternative $x \neq \rhd_1$, consider the following sets: $\{\rhd_1, x\}, S \setminus \rhd_1$. If $x = \rhd_1$, then consider the sets: $\{\rhd_2, x\}, S \setminus \rhd_2$. We now define an auxiliary object depending on x and the list, which captures the alternative that will be removed in the choice set variations we consider:

$$\rhd^x = \begin{cases} \rhd_1 & \text{if } x \neq \rhd_1 \\ \rhd_2 & \text{if } x = \rhd_1 \end{cases}$$

To illustrate the idea behind the next axiom, we start with an example that applies the reasoning underlying Observation 1 to sets with four alternatives. Consider a set with four elements: $S = \{x, y, z, t\}$ and suppose that $y \succ x \succ z \succ t$ and $y \rhd x \rhd z \rhd t$. Since x is not the first option in the list, we consider the choice probability of x in sets $\{x, y\}$, $\{x, z, t\}$ and then compare it with the choice probability of x in $\{x, y, z, t\}$.

$$
\begin{aligned}
p(x,\{x,y,z,t\}) - \frac{p(x,\{x,y\})p(x,\{x,z,t\})}{p(x,x)} &= \gamma(x)\gamma(y)\gamma(z)\gamma(t)q(x)^2[(1-q(y)^3)-(1-q(y))] \\
&+ \gamma(x)\gamma(y)\gamma(z)(1-\gamma(t))q(x)[(1-q(y)^2)-(1-q(y))] \\
&+ \gamma(x)\gamma(y)(1-\gamma(z))\gamma(t)q(x)[(1-q(y)^2)-(1-q(y))]
\end{aligned}
$$

Recall that we aim to isolate the portion of the choice probability of x that's related to when the decision maker's awareness set encompasses all available alternatives. However, the expression above includes terms related to the choice probability of x when the awareness set is $\{x,y,z\}$ or $\{x,y,t\}$. Notice that the additional terms have the following structure:

$$
\begin{aligned}
p(x,\{x,y,z\}) - \frac{p(x,\{x,y\})p(x,\{x,z\})}{p(x,x)} &= \gamma(x)\gamma(y)\gamma(z)q(x)[(1-q(y)^2)-(1-q(y))] \\
p(x,\{x,y,t\}) - \frac{p(x,\{x,y\})p(x,\{x,t\})}{p(x,x)} &= \gamma(x)\gamma(y)\gamma(t)q(x)[(1-q(y)^2)-(1-q(y))]
\end{aligned}
$$

Rewriting the additional terms yields:

$$
\begin{aligned}
p(x,\{x,y,z,t\}) - \frac{p(x,\{x,y\})p(x,\{x,z,t\})}{p(x,x)} &= \gamma(x)\gamma(y)\gamma(z)\gamma(t)q(x)^2[(1-q(y)^3)-(1-q(y))] \\
&+ (1-\gamma(t))\left[p(x,\{x,y,z\}) - \frac{p(x,\{x,y\})p(x,\{x,z\})}{p(x,x)}\right] \\
&+ (1-\gamma(z))\left[p(x,\{x,y,t\}) - \frac{p(x,\{x,y\})p(x,\{x,t\})}{p(x,x)}\right]
\end{aligned}
$$

To generalize Observation 1 to choice sets with four or more alternatives, we need to handle the additional terms that appear in the expression. In what follows, we define a function f_x that extends the idea behind Equation 3 to sets with four or more alternatives by systematically isolating the awareness set with all available alternatives. We then provide an axiom which extends the idea behind Observation 1 to sets with more than three alternatives.

Let $\mathcal{X}^x$ denote the elements of $\mathcal{X}$ that includes x. For each alternative x, define $f_x : \mathcal{X}^x \to [0, 1)$ recursively as

$$f_x(S) = p(x, S) - \frac{p(x, \{\rhd^x, x\})p(x, S \setminus \rhd^x)}{p(x, x)} - \sum_{\{x, \rhd^x\} \subset T \subset S} f_x(T) \prod_{k \in S \setminus T} p(o, \{k\})$$

Note that the summation term in $f_x(T)$ equals zero if $|T| = 3$. The summation term cancels out the additional terms that are unrelated to the awareness set that includes every alternative. Rewriting the expression from the example with four alternatives above,

$$\begin{aligned} p(x, \{x, y, z, t\}) - \frac{p(x, \{x, y\})p(x, \{x, z, t\})}{p(x, x)} &= \gamma(x)\gamma(y)\gamma(z)\gamma(t)q(x)^2[(1 - q(y)^3) - (1 - q(y))] \\ &+ (1 - \gamma(t))f_x(\{x, y, z\}) + (1 - \gamma(z))f_x(\{x, y, t\}) \\ &= \gamma(x)\gamma(y)\gamma(z)\gamma(t)q(x)^2[(1 - q(y)^3) - (1 - q(y))] \\ &+ \sum_{\{x,y\} \subset T \subset \{x,y,z,t\}} f_x(T) \prod_{k \in \{x,y,z,t\} \setminus T} p(o, \{k\}) \end{aligned}$$

Recall that Observation 1 identifies for which S and $x \in S$ values $f_x = 0$. We are now ready to state the axiom, which describes the conditions under which path independence is satisfied, according to the choice set variations described above. In other words, the axiom identifies the general conditions under which $f_x(S) = 0$.

Axiom 4 (Limited Path Independence) If $xP\rhd_1$ or $x = \rhd_1$ and $xP\rhd_2$, then $f_x(S) = 0$.

By the construction of the choice set variation of S, if $x \neq \rhd_1$ the recall probabilities of all the alternatives except $\rhd_1$ are the same in S and $S \setminus \rhd_1$. Thus, if $xP\rhd_1$, then we can divide-and-conquer choosing x in S. Similarly, if $x = \rhd_1$, then the recall probabilities of all the alternatives in $S \setminus \rhd_2$ except x are the same in S and $S \setminus \rhd_2$. Therefore, if $xP\rhd_2$, then we can divide-and-conquer choosing x in S.

The next axiom says that the choice probability of the outside option in any set S

can be written as the multiplication of its choice probabilities in singleton sets of the elements of S.

Axiom 5 (Path Independence for Outside Option) For any $S \subseteq X$,

$$p(o, S) = \prod_{x \in S} p(o, x)$$

In other words, we can divide-and-conquer choosing the outside option in any choice set, as it is not affected by the final recall probability effects that violates this procedure.

As 4 hints at, not every alternative can satisfy path independence. The next axiom describes how the choice probabilities change when an alternative does not satisfy path independence. Define $\mathbf{q}(x) = \frac{p(x,\{x,y\})-p(x,x)(1-p(y,y))}{p(x,x)p(y,y)}$. Thus, $\mathbf{q}(x)$ is the observable counterpart of the period recall probability of x, $q(x)$, as discussed in Section 3. Let $L(x, S)$ denote the lower counter set of x in set S, *i.e.*, $L(x, S) = \{y \in S \mid xPy\} \cup \{x\}$.

Axiom 6 (Proportional Path Independence Bias) (i) If $\rhd_1 Px$, and $|L(x, S)| \geq 2$ then

$$\frac{f_x(S)}{\sum_{y \in L(x,S)} f_y(S)} = \mathbf{q}(x)^{\rhd(x,S)}$$

(ii) If $x = \rhd_1$, $\rhd_2 Px$, and $|L(x, S)| \geq 2$ then

$$\frac{f_x(S)}{\sum_{y \in L(x,S)} f_y(S)} = \mathbf{q}(x)^{\rhd(x,S)-1}$$

First, consider the case with $\rhd_1 Px$, which means that x is not the first alternative in list $\rhd$. In this case, the sets used in $f_x(S)$ is $\{\rhd_1, x\}, S \setminus \rhd_1\}$. Thus, the final recall probability of every alternative in $S \setminus \rhd_1$ is the same in $S \setminus \rhd_1$ and S. However, because $\rhd_1 Px$ and the final recall probability of $\rhd_1$ is different in sets $\{x, \rhd_1\}$ and S, so we cannot divide-and-conquer choosing x in S.

$$f_x(S) = \prod_{z \in S} \gamma(z)\ q(x)^{\rhd(x,S)} \prod_{\{y \in S \setminus \{\rhd_1\} | yPx\}} (1 - q(y)^{\rhd(y,S)}) \left[(1 - q(1)^{\rhd(\rhd_1,S)}) - (1 - q(1)^{\rhd(\rhd_1,\{x,\rhd_1\})})\right]$$

Because P is transitive, $\rhd_1$ is also preferred to any alternative that is in the lower counter set of x. Thus, the divide-and-conquer structure is violated for any alternative in the lower counter set of x. Moreover, given that x follows $\rhd_1$ in the list, the final recall probability of $\rhd_1$ in $\{x, \rhd_1\}$ is independent from x. Therefore, the violation in the divide-and-conquer structure is the same in every alternative in the lower counter set of x, and this effect cancels out. Now suppose that $x = \rhd_1$ and $\rhd_2 P x$. The intuition behind this case is similar to the first, however, notice that $f_x(S)$ is evaluated using the sets $\{\{\rhd_2, x\}, S \setminus \rhd_2\}$, while every alternative y in $S \setminus x$ is evaluated using the sets $\{\{\rhd_1, y\}, S \setminus \rhd_1\}$. The difference in the sets used causes an additional term to appear in the sum over the lower counter set.

We are now ready to state the characterization theorem.

Theorem 1 A random choice rule p associated with list $\rhd$ has a limited attention and memory representation if and only if $(p, \rhd)$ satisfies Axiom 1-6.

5 Related Literature

In this section, we compare our model with well-known parametric models of stochastic choice. Figure 4 illustrates the relationship between our model and seven well-known stochastic choice models on the Marschak-Machina triangle in a domain of two alternatives, x and y, and the outside option o. This is the simplest environment we can utilize to study the differences and similarities between these models. We show that even in this simple environment, LAM makes distinct predictions from related parametric models of stochastic choice.

The Marschak-Machina triangle is a tool used in decision theory to represent stochastic choices. In the standard setup, each corner represents a different alternative. The edges of the triangle represent the probability mixtures of the three outcomes. The existence of the outside option requires a slight modification of the triangle, as the outside option is always available. To highlight that there is no instance in which we observe choices when the outside option is not available, the edge between x and y is denoted by a dotted line in 3.

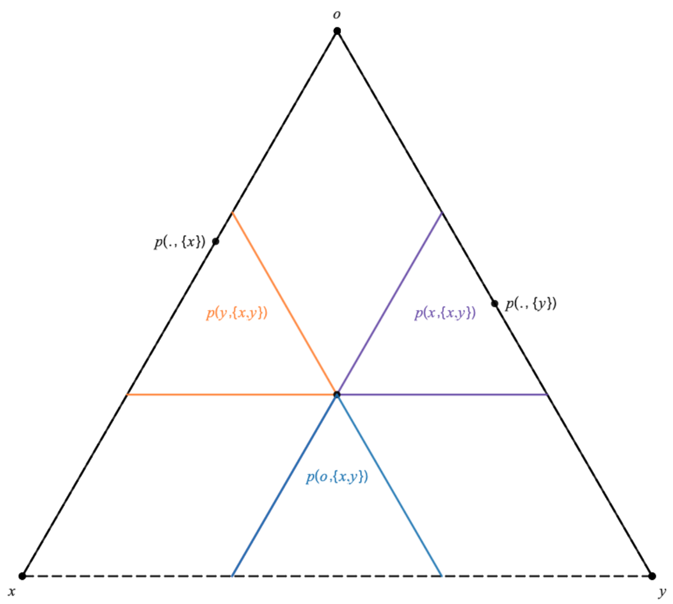

Figure 3: The distance between point $p(., \{x\})$ and corner o (x) correspond to $p(x, x)$ $(p(o, x))$. The distance between point $p(., \{y\})$ and corner o (y) correspond to $p(y, y)$ $(p(o, y))$. Given a point inside the Marschak-Machina triangle, the denoted lengths correspond to the choice probabilities of alternatives x, y and the outside option o when the choice set is $\{x, y\}$.

In Figure 3, choosing the outside option with probability one is depicted by the top corner of the triangle. We then fix stochastic choice for singleton sets and denote these by the black dots on the edges of the triangle. The distance between the corner o and the black point labeled by $p(., \{x\})$ represents $p(x, x)$, which is the choice probability of x when x is the only alternative.[9] The distance between the corner x and $p(., \{x\})$ represents $p(o, x)$, the choice probability of o when only x is available. Similarly, the distance between corner o and the black point labeled by $p(., \{y\})$ is equal to $p(y, y)$ and the distance between corner y and the black point labeled by $p(., \{y\})$ is equal to $p(o, y)$. The points inside the Marschak-Machina triangle capture choice probabilities from the choice set $\{x, y\}$. Consider the black point in the middle of the triangle. We draw three lines that pass through the black dot that are parallel to the edges of the triangle. The length of each of the two lines (denoted in orange) connecting the black point to the o-y edge of the triangle corresponds to $p(y, \{x, y\})$. Similarly, the purple lines connecting the black point to the o-y edge of the triangle corresponds to $p(x, \{x, y\})$ and the blue lines connecting the black point to the x-y edge of the triangle corresponds to $p(o, \{x, y\})$.

[9]Note that $p(x, x)$ is equivalent to $p(x, \{x, o\})$ as the outside option is available as part of any choice set.

In what follows, we fix the singleton choice probabilities as shown in Figure 4. This figure identifies all possibilities for $p(\cdot, \{x, y\})$ for each model given that all models agree on these singleton choices. The thicker green line (labeled as LAM) in Figure 4 denotes all possible $p(\cdot, \{x, y\})$ belonging to our model when x is ranked above y and listed before y, given the singleton choices. Hence, this line captures the explanatory power of our model given these singleton choices.[10]

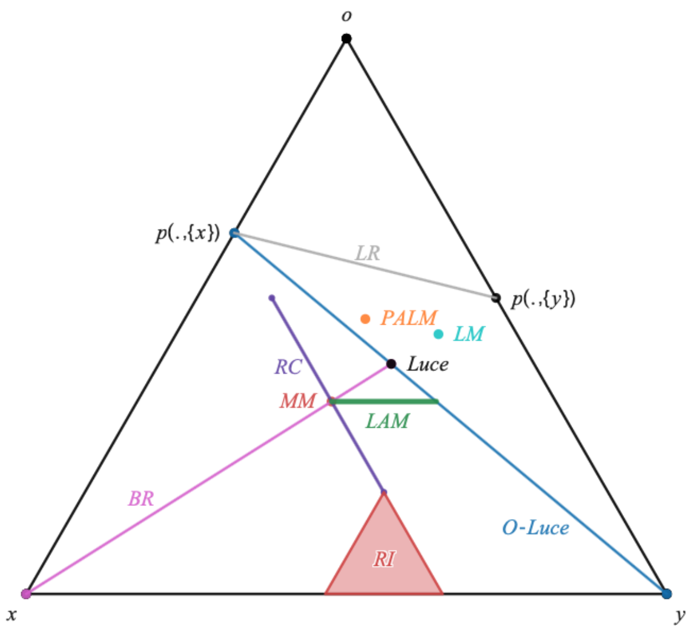

Figure 4: Given $x \succ y$, and choice data on singleton sets, this figure identifies all possibilities for $p(\cdot, \{x, y\})$ for each model of interest. The predictions of LAM are highlighted by the green (thicker) line.

The first set of models we consider is the standard Luce model (see Luce [1959]) and its generalizations. We first modify the Luce model so that we can accommodate the outside option. This definition is also adapted by Echenique et al. [2018]. The Luce model is characterized by the Luce Independence of Irrelevant Alternatives (Luce IIA) condition, which implies the ratio of choice probabilities of two alternatives are the same across choice sets. Hence, the modified Luce can accommodate only one point in Figure 4, denoted by the black point labeled Luce.[11] Echenique et al. [2018] proposes a model

[10]If x were listed after y, then LAM makes the same predictions as Manzini and Mariotti [2014].

[11]There are other models that make the same prediction as the Luce model in this simple setting, as long as each available alternative chosen with positive probability. See the binary focal model of Kovach and Tserenjigmid [2022a], the nested logit model characterized by Kovach and Tserenjigmid [2022b], the stochastic semi-order model of Horan [2021], the general Luce model of Echenique and Saito [2019], the

(PALM) in which the decision maker evaluates available alternatives following a priority order, capturing perception. The likelihood of selecting a particular alternative hinges on the probability of not choosing a previously perceived alternative. Additionally, the probability of opting for an alternative is determined by its relative utility according to Luce. If none of the perceived alternatives is chosen, then the decision maker opts for the outside option. Echenique et al. [2018] uses violations of Luce's IIA to reveal the perception priority of alternatives. We assume that x is perceived higher than y. Then, PALM can accommodate one possible choice, which is denoted by PALM in Figure 4.[12] Tserenjigmid [2021] is another generalization of the Luce model in which the Luce weights depend on the order of the alternatives. In the model, removing a lower ranked alternative does not cause Luce IIA violations. We denote the prediction of this model under the assumption that x is ranked above y in Figure 4, denoted by O-Luce. Since x is the top-ranked alternative, the relative choice probability of x and o does not change after removing y. Hence, all these models are independent of LAM.[13]

The second set of parametric models is about capturing limited attention. One of the seminal papers in this literature is Manzini and Mariotti [2014].[14] This model satisfies two conditions: i) adding an inferior alternative does not affect the choice probability of more-preferred alternatives, and ii) choice set variations in which the most preferred alternative is removed respect Luce IIA. As mentioned above, Manzini and Mariotti [2014] is a special case of our model. There is only one choice probability satisfying these conditions, which is marked by the red dot labeled MM in Figure 4. This point also belongs to our model and captured by the situation in which x is ranked above y and y is listed before x. The model of Manzini and Mariotti [2014] is also subsumed by the models of Brady and Rehbeck [2016] and Aguiar [2017]. Brady and Rehbeck

Luce rule with limited consideration of Ahumada and Ülkü [2018], the preference-oriented Luce model of Doğan and Yıldız [2021], and the categorical consideration model of Honda [2021].

[12] If x and y are perceived simultaneously, then PALM makes the same prediction as Luce. If y is perceived before x, then the choice probability of y is the same as Luce, but the choice probability of x (outside option) is higher (lower) than Luce, which does not overlap with any other models depicted in the triangle.

[13] In the limiting case of LAM where $q(x) = 0$, the model satisfies Luce IIA when alternative y is removed. This limiting case can be captured by the Ordered Luce model.

[14] See Horan [2019] for an axiomatic characterization of Manzini and Mariotti [2014] model when there is no default alternative. Demirkan and Kimya [2020] study a generalization of this model in which γ is the menu-dependent. Gibbard [2021] allows heterogeneous preferences within Manzini and Mariotti [2014]'s framework.

	State 1	State 2	State 3
x	v_x	0	0
y	0	v_y	0
o	R	R	R
Prior probability	s_1	s_2	$1 - s_1 - s_2$

Table 1: The state-payoff structure assumed in the rational inattention model for comparison with the limited attention and memory model.

[2016] assumes a distribution over possible consideration sets in which each possible subset has a positive probability. In this model, we must have $\frac{p(o,\{x,y\})}{p(y,\{x,y\})} = \frac{p(o,y)}{p(y,y)}$, *i.e.*, removing the most preferred alternative respects Luce IIA. Since this model does not require the first condition, it is more general than Manzini and Mariotti [2014]. The points that can be accommodated by Brady and Rehbeck [2016] are indicated by the purple line (labeled as BR) in Figure 4. On the other hand, the random categorization model of Aguiar [2017] accommodates multiple possibilities (labeled as RC) in Figure 4. Note that if $p(\cdot, \{x, y\})$ belongs to this model, $p(x, \{x, y\})$ must be equal to $p(x, x)$. These models are not only different from each other but also distinct from LAM since their predictions do not overlap with the green line except at MM.

We now discuss how LAM is related to the rational inattention model. In the rational inattention model, the decision maker has imperfect information about the payoff resulting from each available alternative but she can acquire information at a cost to resolve the uncertainty regarding the payoffs. The cost of acquiring information is given by the Shannon mutual information between prior and posterior beliefs, multiplied by a parameter capturing the unit cost of acquiring information. The distinction between the modeling approaches of the rational inattention model and LAM requires making certain assumptions about the underlying states of the world and the payoff structure in order to compare the predictions of the two models. In particular, we fix the state space and the payoff structure of the rational inattention as shown in Table 1.

Following Matějka and McKay [2015], we depict the outside option o as an alternative that pays the same payoff regardless of the underlying state. We also assume that $R > 0$. Given the prior state probabilities, the parameter capturing the unit cost of acquiring information, and R, choices from the singleton choice sets pin down the specific

v_x and v_y values that explain the singleton choices. We then use these v_x and v_y values to calculate the model's prediction of choice probabilities when both x and y are available (in addition to the outside option). We repeat this exercise for the possible values of prior state probabilities, the unit cost of acquiring information and R can take, while ensuring that the singleton choices can be supported. The red triangle (labeled as RI) in Figure 4 denote all possible predictions for the binary choice set, given the choices from the singleton choice sets and varying the unit cost of acquiring information and the R value. We would like to highlight that LAM makes distinct predictions from the rational inattention model in this setting.[15]

The third set of parametric models we consider is the literature on choice from lists. Rubinstein and Salant [2006] introduced a model in which the decision maker chooses from lists. They also explored random choice rules where the list is stochastic. Their model satisfies regularity, which is violated in our model. Yegane [2022] is the only model in this literature studying limited memory. To accommodate the outside option, we consider a modified version of this model where the outside option can be chosen with a positive probability. We assume that the last alternative is always the outside option and considered with probability one. To make the comparison with LAM consistent, we assume x appears before y in the list ($x \rhd y \rhd o$).[16] This implies that removing the first alternative in the list does not generate violations of Luce IIA between y and o. On the other hand, introducing y reduces the choice probability of x and the relative choice probabilities of x and o. The model of Yegane [2022] makes a strong prediction: There is only one possible point in Figure 4, denoted by light blue dot labeled LM. Notice that even though LAM combines the models of Manzini and Mariotti [2014] and Yegane [2022], the predictions of LAM is not a convex combination of these models.

Yildiz [2016] considers a decision maker who chooses an alternative by recursively making binary comparisons until the choice set is exhausted. The order of comparison is captured by a list. Yildiz [2016] characterizes random choice rules that are rational-

[15]As an extension, we modified the payoff structure in Table 1 by allowing the payoff of x in state 2 and the payoff of y in state 1 distinct from zero. We then repeated the exercise described above. In this case, the area corresponding to the possible predictions of the rational inattention model becomes very large, and encompasses most of the models depicted in Table 1 above, including LAM.

[16]If the list order is given by $y \rhd x \rhd o$, then the prediction of Yegane [2022] for the given singleton choice probabilities corresponds to a point on the purple line capturing Aguiar [2017].

izable by this process, referred to as list rational (LR) random choice rules. In LR, the list is endogenously determined from the observed choice behavior. In addition, the LR model does not have an outside option. To make the comparison consistent, we introduce an outside option and assume the list exogenously given, which is $x \triangleright y \triangleright o$. Under this list order, LR can accommodate the choices that are indicated by the light gray line (labeled as LR) in Figure 4. Hence, LAM is distinct from the LR model.

6 Conclusion

In this paper, we provide a parametric model to study the interactions between attention, memory, preferences, and choices. Our model conceptually distinguishes between the allocation of attention to different alternatives and the accessibility of alternatives from memory. In our model, when an alternative is not selected, it can be attributed to one of three potential reasons. The first is the presence of a more favored alternative. The second reason is that the alternative might not have been taken into account because the decision maker was unaware of it. The third possibility is that even though the decision maker was aware of the alternative at some point, she could not recall it at the time of choice. Our results show that we can distinguish these three possibilities from each other. To better understand the relationship between attention, memory and preferences, we study the behavioral foundations of the limited attention and memory model. In our future work, we plan to apply our model to better understand the interaction between attention, memory and preferences in economically relevant applications.

References

J. Abaluck and A. Adams-Prassl. What do consumers consider before they choose? identification from asymmetric demand responses. *The Quarterly Journal of Economics*, 136(3):1611–1663, 2021.

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

A. Ahumada and L. Ülkü. Luce rule with limited consideration. *Mathematical Social Sciences*, 93:52–56, 2018.

L. Barseghyan, M. Coughlin, F. Molinari, and J. C. Teitelbaum. Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048, 2021a.

L. Barseghyan, F. Molinari, and M. Thirkettle. Discrete choice under risk with limited consideration. *American Economic Review*, 111(6):1972–2006, 2021b.

H. Block and J. Marschak. Random orderings and stochastic theories of responses. In *Economic Information, Decision, and Prediction*, pages 172–217. Springer, 1960.

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

Y. Demirkan and M. Kimya. Hazard rate, stochastic choice and consideration sets. *Journal of Mathematical Economics*, 87:142–150, 2020.

S. Doğan and K. Yıldız. Odds supermodularity and the luce rule. *Games and Economic Behavior*, 126:443–452, 2021.

F. Echenique and K. Saito. General luce model. *Economic Theory*, 68(4):811–826, 2019.

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

P. Gibbard. Disentangling preferences and limited attention: Random-utility models with consideration sets. *Journal of Mathematical Economics*, 94:102468, 2021.

E. Honda. Categorical consideration and perception complementarity. *Economic Theory*, 71:693–716, 2021.

E. Honka, A. Hortaçsu, and M. Wildenbeest. Empirical search and consideration sets. In *Handbook of the Economics of Marketing*, volume 1, pages 193–257. Elsevier, 2019.

S. Horan. Random consideration and choice: A case study of âdefaultâ options. *Mathematical Social Sciences*, 102:73–84, 2019.

S. Horan. Stochastic semi-orders. *Journal of Economic Theory*, 192:105171, 2021.

J. Huber, J. W. Payne, and C. Puto. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, 9 (1):90–98, 1982.

M. Kovach and G. Tserenjigmid. The focal luce model. *American Economic Journal: Microeconomics*, 14(3):378–413, 2022a.

M. Kovach and G. Tserenjigmid. Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy*, 130(9):2411–2461, 2022b.

R. D. Luce. *Individual choice behavior: A theoretical analysis*. New York, Wiley, 1959.

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

F. Matějka and A. McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298, 2015.

D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*. Academic Press, New York, 1973.

C. R. Plott. Path independence, rationality, and social choice. *Econometrica: Journal of the Econometric Society*, pages 1075–1091, 1973.

J. Rieskamp, J. R. Busemeyer, and B. A. Mellers. Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, 44(3): 631–661, 2006.

A. Rubinstein and Y. Salant. A model of choice from lists. *Theoretical Economics*, 1(1): 3–17, 2006.

A. D. Shocker, M. Ben-Akiva, B. Boccara, and P. Nedungadi. Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Letters*, 2:181–197, 1991.

G. Tserenjigmid. The order-dependent luce model. *Management Science*, 67(11):6915–6933, 2021. doi: 10.1287/mnsc.2020.3866.

E. Yegane. Stochastic choice with limited memory. *Journal of Economic Theory*, 205: 105548, 2022.

K. Yildiz. List-rationalizable choice. *Theoretical Economics*, 11(2):587–599, 2016.

A Proof of Theorem 1

(Necessity) The necessity of all axioms is already outlined in the main text.
(Sufficiency) Assume that $(p, \rhd)$ is a random choice rule with an associated list satisfying Axiom 1 - 6. Define a binary relation P on X as follows: xPy if one of the following is observed:
(i) $p(x,x) = p(x,\{x,y\})$, or
(ii) $p(y,y) \neq p(y,\{x,y\})$ and $x \rhd y$.

We first show that P is asymmetric. For a contradiction, suppose that there exists $x, y \in X$ with xPy and yPx. Without loss of generality, let $x \rhd y$. Then yPx must be revealed from $P(i)$, due to the definition of P. Thus, $p(y,y) = p(y,\{x,y\})$. Therefore, xPy can only follow from $P(i)$, which requires that $p(x,x) = p(x,\{x,y\})$, but this is ruled out by Axiom 1 as $x \rhd y$. Hence, P is asymmetric. P is complete by definition. For any two alternative x and y, either $p(x,x) = p(x,\{x,y\})$ or $p(x,x) \neq p(x,\{x,y\})$. If $p(x,x) = p(x,\{x,y\})$, then xPy by $P(i)$. If $p(x,x) \neq p(x,\{x,y\})$ and $y \rhd x$, then yPx by $P(ii)$. If $p(x,x) \neq p(x,\{x,y\})$, $x \rhd y$, and $p(y,y) = p(y,\{x,y\})$ then yPx by $P(i)$. If $p(x,x) \neq p(x,\{x,y\})$, $x \rhd y$, and $p(y,y) \neq p(y,\{x,y\})$, then xPy follows from $P(ii)$. Transitivity of P follows from Axiom 2. Therefore, P is a linear order. For all $x \in X$, define

$$\gamma(x) = p(x,x)$$

If alternative xPy and $x \rhd y$, define

$$q(x) = \frac{p(x,\{x,y\}) - p(x,x)(1 - p(y,y))}{p(x,x)p(y,y)}$$

Note that $q(x)$ is well-defined and strictly positive, because if there exists $y, z \in X$ such that xPy, z and $x \rhd y, z$, then Axiom 3 implies

$$\frac{p(x,\{x,y\}) - p(x,x)(1 - p(y,y))}{p(x,x)p(y,y)} = \frac{p(x,\{x,z\}) - p(x,x)(1 - p(z,z))}{p(x,x)p(z,z)} > 0$$

Moreover, we can rewrite $q(x)$ as

$$q(x) = 1 + \frac{p(x,\{x,y\}) - p(x,x)}{p(x,x)p(y,y)}$$

Thus, Axiom 1 implies that $q(x) < 1$.

For any set $S \in \mathcal{X}$ with $|S| \geq 2$, we provide a proof by strong induction on the cardinality of the set S.

(Basis Step): First, we show that the choices are consistent with the representation for the basis case $|S| = 2$. If xPy and $x \rhd y$,

$$p(x, \{x, y\}) = \gamma(x)\gamma(y)q(x) + \gamma(x)(1 - \gamma(y))$$

By Axiom 5,

$$\begin{aligned} p(o, \{x, y\}) &= p(o, x)p(o, y) \\ &= (1 - \gamma(x))(1 - \gamma(y)) \end{aligned}$$

Since choice probabilities in any choice set must sum up to 1,

$$p(y, \{x, y\}) = \gamma(x)\gamma(y)(1 - q(x)) + \gamma(y)(1 - \gamma(x))$$

If xPy and $y \rhd x$,

$$\begin{aligned} p(x, \{x, y\}) &= \gamma(x) \\ p(y, \{x, y\}) &= \gamma(y)(1 - \gamma(x)) \\ p(o, \{x, y\}) &= (1 - \gamma(x))(1 - \gamma(y)) \end{aligned}$$

(Induction Hypothesis:) Now suppose that for any S with $|S| \leq n$, the representation holds. **(Inductive Step):** In what follows, to simplify the notation, we denote the final recall probability of x in the awareness set S when the list is $\rhd$ by $Q(x, S, \rhd)$,

$$Q(x, S, \rhd) = q(x)^{\rhd(x,S)}$$

Since the random choice rule p is associated with a single, fixed list $\rhd$, to simplify the notation, we fix the list as $\rhd$ and use the notation $Q(x, S) = Q(x, S, \rhd)$ as long as the simplification does not create any confusion. Note that if $S \cap T = \{x\}$

$$Q(x, S)Q(x, T) = Q(x, S \cup T)$$

Let $\mu(S|T)$ denote the probability that the awareness set equals S when the set of available alternatives is T,

$$\mu(S|T) = \prod_{x \in S} \gamma(x) \prod_{y \in T \setminus S} (1 - \gamma(y))$$

We will utilize the next two lemmas show that the $(p, \rhd)$ satisfying Axiom 1-6 is consistent with the representation. Let $\rhd_1$ and $\rhd_2$ denote the first and second alternative in list $\rhd$ respectively. Define the function 1_{xPy} where $x, y \in X$ as follows,

$$1_{xPy}(i) : \begin{cases} i & \text{if } xPy \\ 1 & \text{otherwise} \end{cases}$$

Lemma 1 *Suppose that $(p, \rhd)$ has a limited attention and memory representation. For any $S \subseteq X$ with $|S| \geq 3$ and any $x \in S$,*

$$\frac{p(x, \{x, \rhd_i\})p(x, S \setminus \rhd_i)}{p(x,x)} - \sum_{\{x,\rhd_i\} \subset T \subset S} f_x(T) \prod_{k \in S \setminus T} p(o,k) \tag{4}$$
$$= \mu(S|S)Q(x,S) \prod_{y \in S \setminus \rhd_i : yPx} (1 - Q(y,S)) 1_{\rhd_i Px}(1 - Q(\rhd_i, \{x, \rhd_i\}))$$
$$+ \sum_{T \subset S} \mu(T|S)Q(x,T) \prod_{y \in T : yPx} (1 - Q(y,T))$$

$$\textit{where } \rhd_i = \begin{cases} \rhd_1 & \textit{if } x \neq \rhd_1 \\ \rhd_2 & \textit{if } x = \rhd_1 \end{cases}$$

Proof. We provide a proof by strong induction on the cardinality of the set S.
(Basis Step:) Suppose that $|S| = 3$. Note that if S is a singleton or a binary set, $f_x(S) = 0$. Hence,

$$\sum_{\{x,\rhd_i\} \subseteq T \subset S} f_x(T) \prod_{k \in S \setminus T} p(o,k) = 0$$

Let $S = \{x, y, \rhd_i\}$, then

$$\frac{p(x, \{x, \rhd_i\})p(x, S \setminus \rhd_i)}{p(x,x)} = \frac{p(x, \{x, \rhd_i\})p(x, \{x, y\})}{p(x,x)}$$

$$\begin{aligned} &\frac{p(x, \{x, \rhd_i\})p(x, \{x, y\})}{p(x,x)} \\ &= \gamma(x)\gamma(y)\gamma(i)Q(x, \{x,y\})Q(x, \{x, \rhd_i\})1_{yPx}(1 - Q(y, \{x,y\}))1_{\rhd_i Px}(1 - Q(\rhd_i, \{x, \rhd_i\})) \\ &\quad + \gamma(x)(1-\gamma(y))\gamma(i)Q(x, \{x, \rhd_i\})1_{\rhd_i Px}(1 - Q(\rhd_i, \{x, \rhd_i\})) \\ &\quad + \gamma(x)\gamma(y)(1-\gamma(i))Q(x, \{x,y\})1_{yPx}(1 - Q(y, \{x,y\})) \\ &\quad + \gamma(x)(1-\gamma(y))(1-\gamma(i)) \end{aligned}$$

Plugging in

$$Q(y, \{x,y\}) = Q(y, \{x, y, \rhd_1\}) = Q(y, S)$$
$$Q(x, \{x,y\})Q(x, \{x, \rhd_i\}) = Q(x, \{x, y, \rhd_1\}) = Q(x, S)$$

$$\begin{aligned}
\frac{p(x,\{x,\rhd_i\})p(x,\{x,y\})}{p(x,x)} =& \mu(S|S)Q(x,S)1_{yPx}(1-Q(y,S))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\})) \\
&+\mu(x,\rhd_i|S)Q(x,\{x,\rhd_i\})1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\})) \\
&+\mu(x,y|S)Q(x,\{x,y\})1_{yPx}(1-Q(y,\{x,y\})) \\
&+\mu(x|S) \\
&=\mu(S|S)Q(x,S)1_{yPx}(1-Q(y,S))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\})) \\
&+\sum_{T\subset\{x,y,\rhd_i\}}\mu(T|x,y,\rhd_i)Q(x,T)\prod_{y\in T:yPx}(1-Q(y,T)) \\
&=\mu(S|S)Q(x,S)\prod_{y\in S\setminus\rhd_i:yPx}(1-Q(y,S))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\})) \\
&+\sum_{T\subset S}\mu(T|S)Q(x,T)\prod_{y\in T:yPx}(1-Q(y,T))
\end{aligned}$$

(Induction Hypothesis:) Now suppose that for any S with $|S|\leq n$, Equation 4 holds. **(Inductive Step):** Consider S with $|S|=n+1$. Note that

$$\frac{p(x,\{x,\rhd_i\})}{p(x,x)}=\gamma(i)Q(x,\{x,\rhd_i\})1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\}))+1-\gamma(i)$$

$$\begin{aligned}
&\frac{p(x,\{x,\rhd_i\})p(x,S\setminus\rhd_i\})}{p(x,x)}-\sum_{\{x,\rhd_i\}\subset T\subset S}f_x(T)\prod_{k\in S\setminus T}p(o,k) \\
&=\sum_{T\subseteq S\setminus\rhd_i}(1-\gamma(i))\mu(T|S\setminus\{\rhd_i\})Q(x,T)\prod_{y\in T:yPx}(1-Q(y,T)) \\
&-\sum_{T\subseteq S\setminus\rhd_i}\gamma(i)\mu(T|S\setminus\{\rhd_i\})Q(x,T\cup\{\rhd_i\})\prod_{y\in T\setminus\{\rhd_i\}:yPx}(1-Q(y,T))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_1\})) \\
&-\sum_{\{x,\rhd_i\}\subset T\subset S}f_x(T)\prod_{k\in S\setminus T}p(o,k) \\
&=\sum_{T\subseteq S\setminus\rhd_i}\mu(T|S)Q(x,T)\prod_{y\in T:yPx}(1-Q(y,T)) \\
&-\sum_{\{x,\rhd_i\}\subseteq T\subseteq S}\mu(T|S)Q(x,T)\prod_{y\in T\setminus\{\rhd_i\}:yPx}(1-Q(y,T))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\})) \\
&-\sum_{\{x,\rhd_1\}\subset T\subset S}f_x(T)\prod_{k\in S\setminus T}p(o,k)
\end{aligned}$$

Using the definition of $f_x(T)$ and the induction hypothesis,

$$\begin{aligned}
f_x(T) &= p(x,T) - \sum_{K\subset T}\mu(K|T)Q(x,K)\prod_{y\in K:yPx}(1-Q(y,K)) \\
&\quad - \mu(T|T)Q(x,T)\prod_{y\in T\setminus\{\rhd_i\}:yPx}(1-Q(y,T))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\})) \\
&= \sum_{K\subseteq T}\mu(K|T)Q(x,K)\prod_{y\in K:yPx}(1-Q(y,K)) \\
&\quad - \sum_{K\subset T}\mu(K|T)Q(x,K)\prod_{y\in K:yPx}(1-Q(y,K)) \\
&\quad - \mu(T|T)Q(x,T)\prod_{y\in T\setminus\{\rhd_i\}:yPx}(1-Q(y,T))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\})) \\
&= \mu(T|T)Q(x,T)\prod_{y\in T\setminus\{\rhd_i\}:yPx}(1-Q(y,T))1_{\rhd_i Px}(1-Q(\rhd_i,T)-(1-Q(\rhd_i,\{x,\rhd_i\})))
\end{aligned}$$

Thus,

$$\begin{aligned}
&\frac{p(x,\{x,\rhd_i\})p(x,S\setminus\{\rhd_i\})}{p(x,x)} - \sum_{\{x,\rhd_i\}\subset T\subset S} f_x(T)\prod_{k\in S\setminus T}p(o,k) = \sum_{T\subseteq S\setminus\rhd_i}\mu(T|S)Q(x,T)\prod_{y\in T:yPx}(1-Q(y,T)) \\
&\quad - \sum_{\{x,\rhd_i\}\subseteq T\subseteq S}\mu(T|S)Q(x,T)\prod_{y\in T\setminus\{\rhd_i\}:yPx}(1-Q(y,T))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\})) \\
&- \sum_{\{x,\rhd_i\}\subset T\subset S}\mu(T|T)\prod_{k\in S\setminus T}p(o,k)Q(x,T)\prod_{y\in T\setminus\{\rhd_i\}:yPx}(1-Q(y,T))1_{\rhd_i Px}[1-Q(\rhd_i,T)-(1-Q(\rhd_i,\{x,\rhd_i\}))] \\
&= \sum_{T\subseteq S\setminus\rhd_i}\mu(T|S)Q(x,T)\prod_{y\in T:yPx}(1-Q(y,T)) - \sum_{\{x,\rhd_i\}\subseteq T\subset S}\mu(T|S)Q(x,T)\prod_{y\in T\setminus\{\rhd_i\}:yPx}(1-Q(y,T)) \\
&\quad + \mu(S|S)Q(x,S)\prod_{y\in S\setminus\{\rhd_i\}:yPx}(1-Q(y,S))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\})) \\
&\quad = \sum_{T\subset S}\mu(T|S)Q(x,T)\prod_{y\in T:yPx}(1-Q(y,T)) \\
&\quad + \mu(S|S)Q(x,S)\prod_{y\in S\setminus\{\rhd_i\}:yPx}(1-Q(y,S))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\}))
\end{aligned}$$

■

Lemma 2 *Suppose that for any T with $|T|\leq n$, choices are consistent with the representation. Then, for any S with $|S|=n+1$,*

$$\sum_{x\in S\cup\{o\}} f_x(S) = \sum_{x\in S:\rhd_i Px}\mu(S|S)Q(x,S)\prod_{y\in S\setminus\{\rhd_i\}:yPx}(1-Q(y,S))[(1-Q(\rhd_i,S))-(1-Q(\rhd_i,\{x,\rhd_i\}))]$$

Proof. By the definition of $f_x(S)$,

$$p(x,S)-f_x(S)=\frac{p(x,\{x,\rhd_i\})p(x,S\setminus\{\rhd_i\})}{p(x,x)}+\sum_{\{x,\rhd_i\}\subseteq T\subset S}f_x(T)\prod_{k\in S\setminus T}p(o,k)$$

By Lemma 1,

$$\sum_{x\in S\cup\{o\}}[p(x,S)-f_x(S)]=\sum_{x\in S\cup\{o\}}\sum_{T\subset S}\mu(T|S)Q(x,T)\prod_{y\in T:yPx}(1-Q(y,T))$$
$$+\sum_{x\in S\cup\{o\}}\mu(S|S)Q(x,S)\prod_{y\in S\setminus\{\rhd_i\}:yPx}(1-Q(y,S))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\}))$$

Note that

$$\sum_{x\in S\cup\{o\}}\sum_{T\subset S}\mu(T|S)Q(x,T)\prod_{y\in T:yPx}(1-Q(y,T))+\sum_{x\in S\cup\{o\}}\mu(S|S)Q(x,S)\prod_{y\in S:yPx}(1-Q(y,S))=1$$

Thus,

$$\sum_{x\in S\cup\{o\}}[p(x,S)-f_x(S)]=1-\sum_{x\in S\cup\{o\}}\mu(S|S)Q(x,S)\prod_{y\in S:yPx}(1-Q(y,S))$$
$$+\sum_{x\in S\cup\{o\}}\mu(S|S)Q(x,S)\prod_{y\in S\setminus\{\rhd_i\}:yPx}(1-Q(y,S))1_{\rhd_i Px}(1-Q(\rhd_i,\{x,\rhd_i\}))$$

Plugging in

$$\sum_{x\in S\cup\{o\}}p(x,S)=1$$

$$1-\sum_{x\in S\cup\{o\}}f_x(S)=1-\sum_{x\in S:\rhd_i Px}\mu(S|S)Q(x,S)\prod_{y\in S\setminus\{\rhd_i\}:yPx}(1-Q(y,S))[(1-Q(\rhd_i,S))-(1-Q(\rhd_i,\{x,\rhd_i\}))]$$
$$\sum_{x\in S\cup\{o\}}f_x(S)=\sum_{x\in S:\rhd_i Px}\mu(S|S)Q(x,S)\prod_{y\in S\setminus\{\rhd_i\}:yPx}(1-Q(y,S))[(1-Q(\rhd_i,S))-(1-Q(\rhd_i,\{x,\rhd_i\}))]$$

■

Recall that by the induction hypothesis, the representation holds for any set T with $|T|\leq n$.

Now consider a set $S \subseteq X$ with $|S| = n+1$. If $xP\rhd_1$ or $x = \rhd_1$ and $xP\rhd_2$, then by Axiom 4,

$$\begin{aligned} f_x(S) &= p(x,S) - \frac{p(x,\{x,\rhd_i\})p(x,S\setminus\{\rhd_i\})}{p(x,x)} - \sum_{\{x,\rhd_i\}\subset T\subset S} f_x(T) \prod_{k\in S\setminus T} p(o,k) \\ &= 0 \\ p(x,S) &= \frac{p(x,\{x,\rhd_i\})p(x,S\setminus\{\rhd_i\})}{p(x,x)} + \sum_{\{x,\rhd_i\}\subset T\subset S} f_x(T) \prod_{k\in S\setminus T} p(o,k) \end{aligned}$$

By Lemma 1,

$$\begin{aligned} p(x,S) &= \mu(S|S)Q(x,S) \prod_{y\in S\setminus\{\rhd_i\}:yPx} (1-Q(y,S)) 1_{\rhd_i P x}(1-Q(\rhd_i,\{x,\rhd_i\})) \\ &+ \sum_{T\subset S} \mu(T|S)Q(x,T) \prod_{y\in T:yPx} (1-Q(y,T)) \\ &= \mu(S|S)Q(x,S) \prod_{y\in S\setminus\{\rhd_i\}:yPx} (1-Q(y,S)) \\ &+ \sum_{T\subset S} \mu(T|S)Q(x,T) \prod_{y\in T:yPx} (1-Q(y,T)) \\ &= \sum_{T\subseteq S} \mu(T|S)Q(x,T) \prod_{y\in T:yPx} (1-Q(y,T)) \end{aligned}$$

Thus, Axiom 4 ensures that choices are consistent with the representation if $xP\rhd_1$ or $x = \rhd_1$ and $xP\rhd_2$. What remains to show is that the choices are consistent with the representation if $\rhd_1 Px$ or $x = \rhd_1$ and $\rhd_2 Px$.

Axiom 4 implies that for any alternative $x \in S$ that is revealed to be preferred to $\rhd_1$, $f_x(S) = 0$. From Axiom 5, for any S, $f_o(S) = 0$. Thus,

$$\sum_{x\in S\cup\{o\}} f_x(S) = \sum_{x\in S:\rhd_1 Px} f_x(S) + f_{\rhd_1(S)} \tag{5}$$

First, consider the case with $x = \rhd_1$ and $\rhd_2 P\rhd_1$. By Axiom 6,

$$\frac{f_{\rhd_1}(S)}{\sum_{y\in L(\rhd_1,S)} f_y(S)} = q(1)^{\rhd(\rhd_1,S)-1}$$

Lemma 2 together with Equation 5 implies,

$$
\begin{aligned}
&f_{\rhd_1}(S)\\
&= q(1)^{\rhd(\rhd_1,S)-1}\left[\sum_{x\in S:\rhd_i Px}\mu(S|S)Q(x,S)\prod_{y\in S\setminus\{\rhd_i\}:yPx}(1-Q(y,S))\left[(1-Q(\rhd_i,S))-(1-Q(\rhd_i,\{x,\rhd_i\}))\right]\right]\\
&\qquad= q(1)^{\rhd(\rhd_1,S)-1}\mu(S|S)\Bigg[\sum_{x\in S\setminus\{\rhd_1\}:\rhd_1 Px}Q(x,S)[(1-Q(\rhd_1,S))\\
&\qquad\qquad-(1-Q(\rhd_1,\{x,\rhd_1\}))]\prod_{y\in S\setminus\{\rhd_1\}:yPx}(1-Q(y,S))\\
&\qquad+Q(\rhd_1,S)[(1-Q(\rhd_2,S))-(1-Q(\rhd_2,\{\rhd_1,\rhd_2\}))\prod_{y\in S\setminus\{\rhd_2\}:yPx}(1-Q(y,S))]\Bigg]
\end{aligned}
$$

Since $Q(\rhd_2,\{\rhd_1,\rhd_2\})=1$

$$
\begin{aligned}
f_{\rhd_1}(S) &= q(1)^{\rhd(\rhd_1,S)-1}\mu(S|S)\Bigg[\sum_{x\in S\setminus\{\rhd_1\}:\rhd_i Px}Q(x,S)\left[(1-Q(\rhd_1,S))\right.\\
&\quad\left.-(1-Q(\rhd_1,\{x,\rhd_1\}))\prod_{y\in S\setminus\{\rhd_1\}:yPx}(1-Q(y,S))+Q(\rhd_1,S)\prod_{y\in S:yP\rhd_1}(1-Q(y,S))\right]\Bigg]\\
&= q(1)^{\rhd(\rhd_1,S)-1}\mu(S|S)\prod_{y\in S:yP\rhd_1}(1-Q(y,S))\Bigg[Q(\rhd_1,S)\\
&\quad+\sum_{x\in S\setminus\{\rhd_1\}:\rhd_1 Px}Q(x,S)\left[(1-Q(\rhd_1,S))-(1-Q(\rhd_1,\{x,\rhd_1\}))\right]\prod_{y\in S:\rhd_1 PyPx}(1-Q(y,S))\Bigg]
\end{aligned}
$$

For any $x\in S\setminus\{\rhd_1\}$, $Q(\rhd_1,\{x,\rhd_1\})=q(1)$,

$$
\begin{aligned}
f_{\rhd_1}(S) &= q(1)^{\rhd(\rhd_1,S)-1}\mu(S|S)\prod_{y\in S:yP\rhd_1}(1-Q(y,S))\Bigg[Q(\rhd_1,S)\left[(1-Q(\rhd_1,S))\right.\\
&\qquad\left.-(1-Q(\rhd_1,\{x,\rhd_1\}))\right)\sum_{x\in S\setminus\{\rhd_1\}:\rhd_1 Px}Q(x,S)\prod_{y\in S:\rhd_1 PyPx}(1-Q(y,S))\Bigg]
\end{aligned}
$$

Denote the last alternative in list $\rhd$ with $\rhd_n$, so $Q(\rhd_n,S)=1$. Notice that, if $\rhd_n Pz$ for some $z \in S$, then

$$\mu(S|S)Q(x,S)\left[(1-Q(\rhd_1,S))-(1-Q(\rhd_1,\{x,\rhd_1\}))\right]\prod_{y\in S\setminus\{\rhd_1\}:yPx}(1-Q(y,S))=0$$

Enumerate the alternatives that are strictly worse that $\rhd_1$ and at least as good as $\rhd_n$ so that $\rhd_1 P x_1 P x_2 ... P x_k = \rhd_n$

$$\begin{aligned}\sum_{i=1}^{k} Q(x_i,S)\prod_{j\in\{1,2,\ldots,k\}:x_jPx_i}(1-Q(x_j,S)) &= Q(x_k,S)(1-Q(x_{k-1},S))\prod_{j<k-1}(1-Q(x_j,S))\\ &\quad + Q(x_{k-1},S)\prod_{j<k-1}(1-Q(x_j,S))\\ &\quad + \sum_{i=1}^{k-2} Q(x_i,S)\prod_{j\in\{1,2,\ldots,k\}:x_jPx_i}(1-Q(x_j,S))\end{aligned}$$

Since $Q(x_k,S)=1$,

$$\begin{aligned}\sum_{i=1}^{k} Q(x_i,S)\prod_{j\in\{1,2,\ldots,k\}:x_jPx_i}(1-Q(x_j,S)) &= \prod_{j<k-1}(1-Q(x_j,S))\\ &\quad + \sum_{i=1}^{k-2} Q(x_i,S)\prod_{j\in\{1,2,\ldots,k\}:x_jPx_i}(1-Q(x_j,S))\end{aligned}$$

If we keep iterating until the sum is exhausted,

$$\sum_{i=1}^{k} Q(x_i,S)\prod_{j\in\{1,2,\ldots,k\}:x_jPx_i}(1-Q(x_j,S))=1$$

Thus,

$$\begin{aligned}f_{\rhd_1}(S) &= q(1)^{\rhd(\rhd_1,S)-1}\mu(S|S)\prod_{y\in S:yP\rhd_1}(1-Q(y,S))\Big[Q(\rhd_1,S)+(1-Q(\rhd_1,S))-(1-Q(\rhd_1,\{x,\rhd_1\}))\Big]\\ &= q(1)^{\rhd(\rhd_1,S)-1}\mu(S|S)\prod_{y\in S:yP\rhd_1}(1-Q(y,S))Q(\rhd_1,\{x,\rhd_1\})\\ &= Q(\rhd_1,S)\mu(S|S)\prod_{y\in S:yP\rhd_1}(1-Q(y,S))\end{aligned}$$

Combining this with the definition of $f_{\rhd_1}(S)$ and Lemma 1,

$$f_{\rhd_1}(S) = \mu(S|S)Q(\rhd_1, S) \prod_{y\in S: yP\rhd_1} (1 - Q(y,S))$$
$$= p(\rhd_1, S) - \mu(S|S)Q(\rhd_1, S)(1 - Q(\rhd_2, \{\rhd_1, \rhd_2\})) - \sum_{T\subset S} \mu(T|S)Q(\rhd_1, T) \prod_{z\in T: zPx} (1 - Q(z,T))$$

$Q(\rhd_2, \{\rhd_1, \rhd_2\}) = 1$ implies that

$$p(\rhd_1, S) = \mu(S|S)Q(\rhd_1, S) \prod_{y\in S: yP\rhd_1} (1 - Q(y,S) + \sum_{T\subset S} \mu(T|S)Q(\rhd_1, T) \prod_{z\in T: zPx} (1 - Q(z,T))$$
$$= \sum_{T\subseteq S} \mu(T|S)Q(\rhd_1, T) \prod_{z\in T: zPx} (1 - Q(z,T))$$

Now suppose that $\rhd_1 Px$. First, consider the most preferred alternative x among the ones with $f_x(S) > 0$.

$$\sum_{x\in S\cup\{o\}} f_x(S) = \sum_{x\in S:\rhd_1 Px} f_x(S)$$

By Axiom 6,

$$\frac{f_x(S)}{\sum_{y\in L(x,S)} f_y(S)} = q(x)^{\rhd(\rhd_x, S)}$$

Lemma 2 implies,

$$f_x(S)$$
$$= q(x)^{\rhd(x,S)} \left[\sum_{x\in S:\rhd_1 Px} \mu(S|S)Q(x,S) \prod_{y\in S\setminus\{\rhd_1\}: yPx} (1 - Q(y,S)) \left[(1 - Q(\rhd_1, S)) - (1 - Q(\rhd_1, \{x, \rhd_1\}))\right] \right]$$
$$= q(x)^{\rhd(x,S)} \mu(S|S) \left[(1 - Q(\rhd_1, S)) - (1 - (1 - Q(\rhd_1, \{\rhd_1, x\})))\right]$$
$$\left[\sum_{x\in S:\rhd_1 Px} Q(x,S) \prod_{y\in S\setminus\{\rhd_1\}: yPx} (1 - Q(y,S)) \right]$$
$$= q(x)^{\rhd(x,S)} \mu(S|S) \left[(1 - Q(\rhd_1, S)) - (1 - q(1))\right] \prod_{y\in S: yP\rhd_1} (1 - Q(y,S))$$

From Lemma 1 and Axiom 6,

$$
\begin{aligned}
f_x(S) &= p(x,S) - \mu(S|S)Q(x,S)(1-Q(\rhd_1,\{\rhd_1,x\})) + \sum_{T\subset S}\mu(T|S)Q(x,T)\prod_{z\in T:zPx}(1-Q(z,T)) \\
p(x,S) &= f_x(S) + \mu(S|S)Q(x,S)(1-Q(\rhd_1,\{\rhd_1,x\})) + \sum_{T\subset S}\mu(T|S)Q(x,T)\prod_{z\in T:zPx}(1-Q(z,T)) \\
&= \mu(S|S)Q(x,S)\left[(1-Q(\rhd_1,S))-(1-Q(\rhd_1,\{\rhd_1,x\}))\right]\prod_{y\in S:yP\rhd_1}(1-Q(y,S)) \\
&\quad + \mu(S|S)Q(x,S)(1-Q(\rhd_1,\{\rhd_1,x\})) \\
&\quad + \sum_{T\subset S}\mu(T|S)Q(x,T)\prod_{z\in T:zPx}(1-Q(z,T)) \\
&= \mu(S|S)Q(x,S)(1-Q(\rhd_1,S))\prod_{y\in S:yP\rhd_1}(1-Q(y,S)) \\
&\quad + \sum_{T\subset S}\mu(T|S)Q(x,T)\prod_{z\in T:zPx}(1-Q(z,T)) \\
&= \mu(S|S)Q(x,S)\prod_{y\in S:yPx}(1-Q(y,S)) \\
&\quad + \sum_{T\subset S}\mu(T|S)Q(x,T)\prod_{z\in T:zPx}(1-Q(z,T)) \\
p(x,S) &= \sum_{T\subseteq S}\mu(T|S)Q(x,T)\prod_{z\in T:zPx}(1-Q(z,T))
\end{aligned}
$$

which completes the proof.