Dynamic random choice with inattention[*]

Aishwarya Harichandran[†] Saptarshi Mukherjee [‡] Nozomu Muto[§]
Ruhi Sonal[¶]

September 2025

Abstract

We model random choice when an attention-constrained decision maker (DM) chooses from a menu finite number of times. The DM's attention to each alternative in every time period may vary. We axiomatically characterize two dynamic random choice rules in this setting: an attention-constrained dynamic random utility model and a dynamic consideration set rule. Our axioms constrain the intertemporal changes in choice probabilities. We show that the attention probabilities for each time period and the preferences are identifiable.

1 Introduction

We consider an attention-constrained decision maker (DM) choosing from a menu in different time periods. Several papers have modeled random choice in static settings.[1] One of the widely accepted explanations for stochasticity is the presence of inattention. Papers such as Manzini and Mariotti (2014), Cattaneo et al. (2020), Masatlioglu and Vu (2024) model choice with attention constraints applicable to choice data that doesn't distinguish the time periods in which the choices are made. In this paper, we model stochastic choice with inattention in a dynamic setting. When choice data for multiple time periods is available, several interesting aspects of behavior can be explored:

(i) **Search and inattention:** consider a DM buying breakfast cereal every week for a month from a department store. In the first week, her attention may be drawn to brands

[*]Preliminary and incomplete. Do not cite.

[†]Department of Economics, IIM Sirmaur

[‡]Department of Humanities and Social Sciences, IIT Delhi

[§]Department of Economics, Hitotsubashi University

[¶]Department of Social Sciences and Humanities, IIIT Delhi

[1]Luce (1959), Block et al. (1959), Barberà and Pattanaik (2010), Manzini and Mariotti (2014), Manzini and Mariotti (2015), Fudenberg et al. (2015), Yildiz (2016), Ahn et al. (2017), Caplin and Martin (2018), Manzini and Mariotti (2018), Cattaneo et al. (2020) and Suleymanov (2018)

located on a prominent shelf. Empirical and experimental evidence show that shelves at eye level draw more attention (Chandon et al. (2009)). Urban (1975) shows that only a small number of products are observed at a time. In the second week, she may explore brands that she didn't notice on her previous visit. Therefore attention drawn by a brand in one period may be conditional on attention paid to it and other brands in the previous time periods. Our model captures this effect. If choice is stochastic because the DM is searching in the presence of inattention as described above, then as the time periods increase and more alternatives are observed, it is natural to expect that choice probabilities will converge to the deterministic case with one alternative being chosen with full probability. Our model implies the same and is compatible with search as the motivation for stochasticity.

(ii) **Dynamic formation of consideration sets:** In the previous example, the DM may remember a brand she observed in an earlier time period and choose it with high probability even if she has paid relatively low attention to it in the current time period. In the literature on stochastic choice[2], two-stage decision making is a well studied procedure: in the first stage, the DM *considers* a subset of alternatives from the menu and in the second stage, an alternative is chosen from the subset that is considered. Wright and Barbour (1977) is among the earliest works that introduced the idea of such consideration sets. Manzini and Mariotti (2014) explores stochastic choice through a model of random consideration sets where different subsets are considered with varying frequency. Masatlioglu and Vu (2024) models choice behavior that results from two-stage decision making and the evolution of consideration sets– they characterize a choice procedure that generates nested consideration sets. When choice data for several time periods is available, a nuanced exploration of this two-stage procedure becomes possible. In our model, previously observed alternatives remain in the consideration set in subsequent time periods with the accumulated[3] probability of the attention drawn by them. The consideration set in different time periods therefore depends on the history of attention.

(iii) **Identifying conditional attention parameters:** Closely related to the above two is the problem of identifying attention parameters for each time period from the data. As the example in (i) states, when the DM observes the same menu repeatedly, the attention drawn by an alternative in the time period t is likely to depend on the attention drawn previously by all the alternatives. A simple approach to analyzing data across time periods is to consider repeated applications of a static model in each time period. However, this approach cannot capture or identify the intertemporal nature of attention in such settings. We consider the static model in Manzini and Mariotti (2014) as the

[2]Manzini and Mariotti (2014), Masatlioglu and Vu (2024) are some of the papers that model this procedure.

[3]we discuss the procedure for computing the attention probabilities in detail in section 2. Here "accumulated" is used as an indicative term and does not mean summation.

framework for single-period choice and provide a characterization for a dynamic choice rule that is equivalent to Manzini and Mariotti (2014)'s choice rule in the absence of choice history. We identify the attention parameters with intertemporal effects– the attention parameter for period t is identified as conditional on the attention it drew in the previous time periods.

(iv) **Heterogeneity in attention:** When a DM chooses repeatedly, her attention may vary in different time periods. Factors such as cognitive constraints, time and various internal (e.g. health) and external factors (e.g. advertising) influence her attention in different time periods in varying ways. For example, she may be more time constrained on weekdays as compared to weekends, or may have been exposed to advertising that may influence her attention in some time periods (Shapiro et al. (1997)). Existing models of stochastic choice recognize the heterogeneity but this paper is the first to model stochastic choice for choice data that is disaggregated for time periods. Since our model identifies attention for each time period without restricting its correlation across time periods, it provides a framework for exploring the influence of such factors on attention over time. This is an important aspect that can provide useful feedback for each time period for marketing strategies being employed and also for further exploration of cognitive and economic decision making.

Another example of the above effects on choice is a DM searching for a movie to watch every day on an OTT platform such as Netflix. On any day, she does not observe all the available movies. She observes a few and chooses a movie from the set that she has observed in the current or previous time periods. Over time, she learns which are the available movies and may spend less time searching before making a choice. Our model is applicable to the above example. The probability with which a movie is chosen will depend on not just the probability of considering it in that time period, but also on whether or not it was observed in previous time periods. While ours is the first paper to model inattention in the setting of dynamic choice, papers that are most closely related to ours is Manzini and Mariotti (2014) which we use as the framework for single-period choice. In terms of the economic intuition Masatlioglu and Vu (2024) which models the expansion of consideration sets using *attention filters* followed a similar reasoning. Their characterization is time-invariant and is applicable to unobserved heterogeneous preferences and consideration sets. The evolution of consideration sets is also modeled by Fershtman and Pavan (2025) in a search and learning framework.

There is a growing literature on dynamic choice behavior that includes a variety of approaches such as structural models, dynamic optimization and standard choice theoretic approaches. Aguirregabiria and Mira (2010) provides a survey of discrete dynamic choice approaches. Frick et al. (2019) develop a model of dynamic random utility where utilities are allowed to exhibit serial correlation over time. They also show that the imposition of Bayesian rationality can restrict randomness and explore time-consistency of preferences. Kashaev et al. (2023) model dynamic random utility in the broad framework of random utility models (RUM)- their model generalizes static RUM to dynamic settings and allows for preference heterogeneity and correlation across time. Our paper models choice with inattention in a dynamic setting which to the best of our knowledge hasn't been explored previously. The model also provides a framework for understanding the intertemporal formation of consideration sets.

We introduce and characterize the **dynamic attention rule (DAR).** The rule is sensitive to the time periods in which the DM makes her choice. It states that for any set A, the probability of choosing an alternative $x \in A$ in time period t is the joint probability that x drew attention in any of the time periods up to t and the probability that alternatives ranked higher than x did not draw attention until time period t. Intuitively, the rule is similar to the static random consideration set rule (RCS rule) in Manzini and Mariotti (2014). We note that if there is only one time period, DAR and RCS rules are equivalent. In order to capture the intertemporal effects of attention on choice, we introduce the following axioms: dynamic regularity (DR) (i), (ii) and (iii). These axioms constrain the changes in choice probability over time. DR (i) is a dynamic version of menu independence and of Manzini and Mariotti (2014)'s i-asymmetry axiom. DR (ii) requires that if an alternative impacts the choice probability of another in any time period, then it also impacts it in future time periods. The motivation for this axiom is the idea that it is possible that the DM does not pay attention to an alternative y that is better than some x for some time periods. However, if in any time period he notices y such that it affects the choice probability of x then in future time periods, the effect may increase or decrease but never returns to the state where y had no effect on x.

DR (iii) requires that choice probabilities of an alternative from a singleton set i.e. in the absence of any other alternative must be non-decreasing over time. Note that since the reason for random choice in our model is inattention, if the DM is fully attentive to a singleton menu, she must choose the alternative with full probability. When the DM is repeatedly observing the same menu and is inattentive, then in the absence of any other alternatives there is no reason within the framework of this model for her to choose with lower probability. In addition to the above dynamic regularity axioms, we introduce a dynamic menu independence (DMI) as another axiom. This axiom is a dynamic version of standard menu independence axioms in the stochastic choice literature (Manzini and Mariotti (2014), Bhattacharya et al. (2021)). DAR is fully characterized with DR and DMI. The primitives of the rule: the attention parameters for each time period and the underlying binary relation are identified. We present the axioms and the characterization in section 2 in detail.

In the following example, we compare DAR to Manzini and Mariotti (2014)'s rcs rule to demonstrate the granular identification of attention parameters for each time period given the history of choice that DAR enables.

Example 1 *Suppose that choice data disaggregated for time periods* $t = 1, 2$ *is available for a DM facing the menu* $\{x, y\}$*:*

t=1	*t=2*
$P_1(x, \{x\}) = 0.4$	$P_2(x, \{x\}) = 0.5$
$P_1(y, \{y\}) = 0.25$	$P_2(y, \{y\}) = 0.6$
$P_1(x, \{x, y\}) = 0.3$	$P_2(x, \{x, y\}) = 0.2$
$P_1(y, \{x, y\}) = 0.5$	$P_2(y, \{x, y\}) = 0.6$

For $t = 1$ *both Manzini and Mariotti (2014) and DAR explain the above data as choice generated by a strict binary relation* $\succ$ *where* $y \succ x$ *and the attention probabilities for* x *and* y *are identified as* $\delta(x) = P_1(x, \{x\}) = 0.4$ *and* $P_1(y, \{y\}) = \delta(y) = 0.25$ *respectively. For* $t = 2$*, Manzini and Mariotti (2014) identifies the probabilities as* $\delta(x) = P_2(x, \{x\}) = 0.5$ *and* $\delta(y) = P_2(y, \{y\}) = 0.6$*. Manzini and Mariotti (2014) do not model choice in different time periods therefore the attention probabilities are identified independently of the same. DAR identifies the total attention drawn by each alternative until each time period and also the conditional attention in every time period given the attention drawn previously. This allows attention and consequently the formation of consideration sets to be modeled as related across time periods. The following table shows the conditional attention probabilities as calculated in our model:*

t=1	*t=2*
$\delta_1(x, \{x\}) = 0.4$	$\delta_2(x, \{x\}) = 0.167$
$\delta_1(y, \{y\}) = 0.25$	$\delta_2(y, \{y\}) = 0.467$

Note that although $P_2(x, \{x, y\}) < P_1(x, \{x, y\})$ *Manzini and Mariotti (2014) identifies a higher attention probability for both* x *and* y *in* $t = 2$ *as compared to* $t = 1$*. This points to the intuition that both the alternatives drew greater attention in* $t = 2$*. On the contrary, using DAR we find that the conditional attention to* x *in* $t = 2$ *is lower than the attention it drew in* $t = 1$ *i.e.* $\delta_2(x) < \delta_1(x)$*. Therefore, we have an alternative explanation for the reduction in choice probability of* x *in* $t = 2$*: it did not draw the same quantity of attention in* $t = 2$ *as it did previously while* y *drew more attention than in* $t = 1$*. DAR enables us to interpret such intertemporal changes in attention when dynamic choice data is available. Further, we identify the total attention drawn by each alternative until each time period which is dependent on the conditional attention.*

In section 2 we discuss the model, the axioms and the characterizing result. In section 3 we provide some observations about the relationship between dynamic choice and search.

2 Model

Let X be the set of all alternatives. A decision maker (DM) makes a choice from a menu $A \subseteq X$ in each time period $t = 1, 2, \ldots n$. Let $T = \{1, 2, \ldots, n\}$ denote the set of finite time periods. In any $t \in T$, the DM may decide not to choose any alternative. In this case we say that the default x^* is chosen. Let X^* be the relevant domain such that every $A \subseteq X^*$ contains x^*.

Definition 1 (Dynamic stochastic choice rule) *A dynamic choice rule is a collection of functions* $\{P_t | t = 1, 2, \ldots n\}$ *such that for all* t,

(i) $P_t : X^* \times 2^X \to [0, 1]$

(ii) $P_t(x, A) \in (0, 1)$ *for all* $x \in A, A \subseteq X^*, x \neq x^*$

(iii) $\sum_{x \in A} P_t(x, A) = 1$ *for all* $A \subseteq X^*$.

The influence of attention on stochastic choice has been widely modeled in static settings. We introduce the following dynamic stochastic choice rule that captures inattention in the framework of dynamic choice:

Definition 2 (Dynamic attention rule (DAR)) *A dynamic stochastic choice rule* $\{P_t\}$ *is a dynamic attention rule if* $\exists$ *a strict total order* $\succ$ *over* X *and for all* t, $\exists$ $\delta_t : X \to (0, 1)$ *such that for any* $x \in A$, $A \subseteq X^*$,

$$P_t(x, A) = \Delta_t(x) \prod_{y \in A \setminus \{x\}; y \succ x} (1 - \Delta_t(y))$$

where $\Delta_t(x) = 1 - \prod_{t=1}^{n} (1 - \delta_t(x))$.

The above rule states that the choice probability of an alternative x in any time period t is the probability that x drew attention in some time period upto (including) t and alternatives that $\succ$ ranks above x did not draw the DM's attention.

For a single time period the above rule coincides with the model in Manzini and Mariotti (2014). However, repeated application of Manzini and Mariotti (2014) in each time period would yield different binary relations and attention parameters for every t, as demonstrated in example 1. In DAR, we uniquely identify a time-invariant binary relation that represents the decision maker. In addition, the attention parameters Δ_t are dynamic- it denotes the cumulative attention paid to an alternative over all previous time periods. These attention parameters can be disaggregated into single-period conditional attention parameters for each time period that are distinct from those identified in Manzini and Mariotti (2014)'s one-period model.

Dynamic attention and formation of consideration sets: We interpret $\Delta_t(x)$ as the probability of the event that the decision maker has paid attention to x in some time period upto t. We denote the probability that attention is paid to an alternative in time period t by

$\delta_t : X \to (0,1)$. Note that δ_k, δ_l for some $k, l \leq t$ may be correlated. If for some $x \in X$, δ_k, δ_l are independent for all $k, l \leq t$ then:

$$\Delta_t(x) = 1 - \prod_{t=1}^{n}(1 - \delta_t(x))$$

i.e. the probability that the DM pays attention to x at some time period upto t is the probability that it is not the case that he does not pay attention to x in any of the time period till t. Note that here Δ_t is non-decreasing in t.

Using the above, we can find out the probability of formation of a consideration set in different time periods. For example, let $A = \{a, b, c, d\}$ and $C = \{a, b\}$. The probability of $C \subseteq A$ being considered in $t = 1$ is:

$$\Delta_1(a)\Delta_1(b) = [1 - (1 - \delta_1(a))][1 - (1 - \delta_1(b))] = \delta_1(a)\delta_1(b)$$

In $t = 2$, $C \subseteq A$ is considered with the probability:

$$\Delta_2(a)\Delta_2(b) = [1 - (1 - \delta_1(a))(1 - \delta_2(a))][1 - (1 - \delta_1(b))(1 - \delta_2(b))]$$

In $t = 2$, the probability of considering C is not independent of the attention drawn by the alternatives in C in $t = 1$. Similarly, the probability of considering C in every t can be computed. We calculate the attention probabilities as shown above in DAR and show that these are uniquely identifiable. Next, we introduce the axioms that characterize the rule.

Axiom 1 *Dynamic Regularity (DR) For any $x, y \in A$, $A \subseteq X^*$, and $t \in \{1, 2, \ldots, n\}$*

(i) For any $x, y \in A$, $A \subseteq X^*$, and $t \in \{1, 2, \ldots, n\}$

$$\frac{P_t(x, A \setminus y)}{P_t(x, A)} \neq 1 \implies \frac{P_t(y, A \setminus x)}{P_t(y, A)} = 1$$

(ii) if $\frac{P_t(x,A\setminus y)}{P_t(x,A)} \neq 1$ then $\frac{P_{t+k}(x,A\setminus y)}{P_{t+k}(x,A)} \neq 1$ for all $k = 1, 2, \ldots, (n - t)$

(iii) $\frac{P_{t+1}(x,\{x\})}{P_t(x,\{x\})} > 1$ for all $t < n$

According to DR (i), if the removal of an alternative y from a set A containing both x and y affects the choice probability of x, then the removal of x from the same set A must not affect the choice probability of y. The ratio $\frac{P_t(x,A\setminus y)}{P_t(x,A)}$ is interpreted as the "impact" of y on x similar to Manzini and Mariotti (2014). This axiom is an adaptation of the i-Asymmetry axiom in Manzini and Mariotti (2014) to a dynamic setting.

DR (ii) and DR (iii) restrict the choice probabilities across time periods. DR (ii) requires that if y impacts the choice probability of x in time period t, it must impact it in all subsequent time periods. This rules out situations in which the presence of y influences x only in some time periods and then stops impacting the choice probability in further time periods. Once an alternative influences the choice probability of another alternative, the axiom requires that it continues to affect the probabilities in future time periods although the extent of this effect (if measurable) may vary. Intuitively, this places a restriction on the DM forgetting the alternatives. DR (iii) requires that the choice probability of any x from the singleton set $\{x\}$ is non-decreasing over time. The intuition for this axiom is that since choice is random due to inattention, when a decision maker observes a menu multiple times, the attention and therefore the choice probability of an alternative in the absence of others, will not fall over time.

Axiom 2 (Dynamic menu independence (DMI)) *For any $x, y \in X^*$, for all $t = 1, 2, \ldots, n$,*

$$\frac{P_t(x, A \setminus y)}{P_t(x, A)} = \frac{P_t(x, B \setminus y)}{P_t(x, B)}$$

for all $A, B \subseteq X^*$ such that $x, y \in A \cap B$.

DMI is an adaptation of the i-independence axiom in Manzini and Mariotti (2014) to a dynamic setting. This is a standard independence axiom based on Plott (1973)'s path independence. It requires that the impact of y on x is independent of the menu in which x, y occur.

Lemma 1 *If a dynamic stochastic choice rule $\{P_t | t = 1, 2, \ldots n\}$ satisfies DR (i) and DMI, then for any $x \in A$, $A \subseteq X^*$,*

(i) $\frac{P_t(x, A\setminus y)}{P_t(x, A)} \geq 1$ *for any $y \in A$, $x \neq y$.*

(ii) $\frac{P_t(x, A\setminus y)}{P_t(x, A)} > 1 \iff \frac{P_t(y, A\setminus x)}{P_t(y, A)} = 1$

Proof. Let P_t be a dynamic stochastic choice rule that satisfies DR (i) and DMI. Consider $x, y \in X^*$, $x \neq y$. Suppose that $\frac{P_t(x, A\setminus y)}{P_t(x, A)} < 1$. Let's consider the default x^*. By DMI,

$$\frac{P_t(x^*, \{x\})}{P_t(x^*, \{x, y\})} = \frac{P_t(x^*, \phi)}{P_t(x^*, \{y\})}$$

which implies

$$P_t(x^*, \{x\}))P_t(x^*, \{y\}) = P_t(x^*, \{x, y\})$$

therefore, $(1 - P_t(x, \{x\}))(1 - P_t(y, \{y\})) = 1 - P_t(x, \{x, y\}) - P_t(y, \{x, y\})$ which implies

$$P_t(x, \{x\}) + P_t(y, \{y\}) - P_t(x, \{x\})P_t(y, \{y\}) = P_t(x, \{x, y\}) + P_t(y, \{x, y\}) \tag{1}$$

By DR (i), $\frac{P_t(x,A\setminus y)}{P_t(x,A)} < 1 \implies \frac{P_t(y,A\setminus x)}{P_t(y,A)} = 1$. By DMI, $\frac{P_t(y,\{y\})}{P_t(y,\{x,y\})} = \frac{P_t(y,\{y\})}{P_t(y,\{x,y\})} = 1$. Therefore, $P_t(y, \{x, y\}) = P_t(y, \{y\})$. Using this in (1), we get $P_t(x, \{x\})(1 - P_t(y, \{y\})) = P_t(x, \{x, y\})$. Since $P_t(z, A) \in (0, 1)$ for all $z \in A$, $A \subseteq X^*$, we get $\frac{P_t(x,\{x\})}{P_t(x,\{x,y\})} = \frac{1}{1-P_t(y,\{y\})} > 1$. By DMI, $\frac{P_t(x,A\setminus y)}{P_t(x,A)} > 1$ for any A such that $x, y \in A$, which contradicts $\frac{P_t(x,A\setminus y)}{P_t(x,A)} < 1$. Thus, (i) holds.

Next, notice that (i) and DR (i) together imply

$$\frac{P_t(x, A \setminus y)}{P_t(x, A)} > 1 \implies \frac{P_t(y, A \setminus x)}{P_t(y, A)} = 1$$

Suppose $\frac{P_t(y,A\setminus x)}{P_t(y,A)} = \frac{P_t(x,A\setminus y)}{P_t(x,A)} = 1$ then (1) would imply that $P_t(x, \{x\})P_t(y, \{y\}) = 0$. Since choice probabilities are strictly positive, this is a contradiction. ■

Theorem 1 *A dynamic stochastic choice rule $\mathcal{P}$ is a DAR if and only if P satisfies DR and DMI.*

Proof. Let P be a dynamic stochastic choice rule that satisfies DR and DMI. Consider a menu $A \subseteq X^*$ and $x \in A$. Let $A^x = \{z \in A : \frac{P_t(x,A\setminus z)}{P_t(x,A)} \neq 1, y \neq x\}$. For any $z \in A \setminus A^x$,

$$\frac{P_t(x, A \setminus z)}{P_t(x, A)} = 1$$

which implies $P_t(x, A \setminus z) = P_t(x, A)$. Similarly, we iteratively remove each $z \in A \setminus A^x$ since it does not affect the choice probability of x in A. Therefore, $P_t(x, A) = P_t(x, A^x)$. Now, pick any $w \in A^x$. By construction of A^x, $\frac{P_t(x,A\setminus w)}{P_t(x,A)} \neq 1$. By Lemma 1, $\frac{P_t(x,A\setminus w)}{P_t(x,A)} > 1$. Using DR (ii), $\frac{P_{t'}(x,A\setminus w)}{P_{t'}(x,A)} > 1$ for all $t' \geq t$. By DMI, $\frac{P_t(x,A\setminus w)}{P_t(x,A)} = \frac{P_t(x,\{x\})}{P_t(x,\{x,w\})}$. Further, DMI also implies

$$\frac{P_t(x^*, \{x\})}{P_t(x^*, \{x, w\})} = \frac{P_t(x^*, \phi)}{P_t(x^*, \{w\})}$$

which implies

$$\frac{1 - P_t(x, \{x\})}{1 - P_t(x, \{x, w\}) - P_t(w, \{x, w\})} = \frac{1}{1 - P_t(w, \{w\})}$$

Note that by DR (i), $\frac{P_t(w,\{w\})}{P_t(w,\{x,w\})} = 1$. Using this in the above equation, we get

$$P_t(x, \{x, w\}) = P_t(x, \{x\})(1 - P_t(w, \{w\}))$$

Using the above expression for $P_t(x, \{x, w\})$ in $\frac{P_t(x,A)}{P_t(x,A\setminus w)} = \frac{P_t(x,\{x,w\})}{P_t(x,\{x\})}$, we get

$$\frac{P_t(x, A)}{P_t(x, A \setminus w)} = 1 - P_t(w, \{w\})$$

We re-write the above as follows

$$P_t(x, A) = P_t(x, A \setminus w)[1 - P_t(w, \{w\})]$$

Repeating the above steps for each $w' \in A^x \setminus w$, we get

$$P_t(x, A) = P_t(x, \{x\}) \prod_{w \in A^x} [1 - P_t(w, \{w\})] \tag{2}$$

Define $\succ$ as follows: for any $x, y \in X$,

$$y \succ x \iff \frac{P_t(x, \{x, y\})}{P_t(x, \{x\})} \neq 1$$

By DR (i) and (ii) $\succ$ is complete and does not vary with t. By Lemma 1, $\succ$ is strict. We show that $\succ$ is transitive: for all $x, y, z \in X^*$, by DMI,

$$\frac{P_t(x, \{x, y, z\})}{P_t(x, \{x, z\})} = \frac{P_t(x, \{x, y\})}{P_t(x, \{x\})} = \lambda_{ab}$$

From the above, we can write $P_t(x, \{x, y\}) = \lambda_{ab} P_t(x, \{x\})$ and $P_t(x, \{x, y, z\}) = \lambda_{xz}\lambda_{xy} P_t(x, \{x\}))$. Now consider a particular $a, b, c \in A$, $A \in X^*$. By DMI,

$$\frac{P_t(x^*, \phi)}{P_t(x^*, \{a\})} = \frac{P_t(x^*, \{b\})}{P_t(x^*, \{a, b\})} = \frac{P_t(x^*, \{c\})}{P_t(x^*, \{a, c\})} = \frac{P_t(x^*, \{b, c\})}{P_t(x^*, \{a, b, c\})}$$

which implies

$$P_t(x^*, \{a, b\}) = P_t(x^*, \{a\}) P_t(x^*, \{b\}) \tag{3}$$

$$P_t(x^*, \{a, b, c\}) = P_t(x^*, \{a\}) P_t(x^*, \{b, c\}) = P_t(x^*, \{a\}) P_t(x^*, \{b\}) P_t(x^*, \{c\}) \tag{4}$$

Since $\sum x \in A \subseteq X^* P_t(x, A) = 1$, we use (3), (4) and the fact that $P_t(x^*, \{x\}) = 1 - P_t(x, \{x\})$ for all $x \in A$ to write the following for $A = \{a, b, c\}$ and its binary subsets. Note that x^* is included in A and its subsets by the domain assumption.

$$\lambda_{ac}\lambda_{ab} P_t(a, \{x\}) + \lambda_{bc}\lambda_{ba} P_t(b, \{b\}) + \lambda_{ca}\lambda_{cb} P_t(c, \{c\}) + (1 - P_t(a, \{a\}))(1 - P_t(b, \{b\}))(1 - P_t(c, \{c\})) = 1$$

$$\lambda_{ab} P_t(a\{a\}) + \lambda_{ba} P_t(b, \{b\}) + (1 - P_t(a, \{a\}))(1 - P_t(b, \{b\})) = 1$$

$$\lambda_{ab} P_t(a\{a\}) + \lambda_{ba} P_t(b, \{b\}) + (1 - P_t(a, \{a\}))(1 - P_t(b, \{b\})) = 1$$

$$\lambda_{ac} P_t(a\{a\}) + \lambda_{ca} P_t(c, \{c\}) + (1 - P_t(a, \{a\}))(1 - P_t(c, \{c\})) = 1$$

We re-write the above as

$$P_t(a, \{a\}) P_t(b, \{b\}) + P_t(a, \{a\}) P_t(c, \{c\}) + P_t(b, \{b\}) P_t(c, \{c\})$$

$$= (1 - \lambda_{ab}\lambda_{ac}) P_t(a, \{a\}) + (1 - \lambda_{bc}\lambda_{ba}) P_t(b, \{b\}) + (1 - \lambda_{ca}\lambda_{cb}) P_t(c, \{c\}) + P_t(a, \{a\}) P_t(b, \{b\}) P_t(c, \{c\}), \tag{5}$$

$$(1-\lambda_{ab})P_t(a,\{a\})+(1-\lambda_{ba})P_t(b,\{b\})=P_t(a,\{a\})P_t(b,\{b\})$$

$$(1-\lambda_{ac})P_t(a,\{a\})+(1-\lambda_{ca})P_t(c,\{c\})=P_t(a,\{a\})P_t(c,\{c\})$$

$$(1-\lambda_{bc})P_t(b,\{b\})+(1-\lambda_{cb})P_t(c,\{c\})=P_t(b,\{b\})P_t(c,\{c\})$$

Substituting the above three equations in (5) and re-arranging,

$$(1-\lambda_{ab})(1-\lambda_{ac})P_t(a,\{a\})+(1-\lambda_{ba})(1-\lambda_{bc})P_t(b,\{b\})+(1-\lambda_{ca})(1-\lambda_{cb})P_t(c,\{x\})$$

$$=P_t(a,\{a\})P_t(b,\{b\})P_t(c,\{c\}) \tag{6}$$

Now suppose $a\succ b$, $b\succ c$ and $c\succ a$. Therefore, $\lambda_{ba}>1$ and $\lambda_{cb}>1$. By DR (i), $\lambda_{ab}=1=\lambda_{bc}$. Further, $\lambda_{ac}>1$. Using Lemma 1 (ii), this implies $\lambda_{ca}=1$. Substituting in (6), we get $0=P_t(a,\{a\})P_t(b,\{b\})P_t(c,\{c\})>0$ which is a contradiction. Therefore, $\succ$ is transitive.

Using the definition of $\succ$ and the construction of A^x, we rewrite (1) as

$$P_t(x,A)=P_t(x,\{x\})\prod_{w\succ x:w\in A}[1-P_t(w,\{w\})] \tag{7}$$

Now, set $\delta_1(x)=P_1(x,\{x\})$. Note that for $t=1$, $\Delta_1(x)=\delta_1(x)$. Let

$$\delta_2(x)=\frac{P_2(x,\{x\})-P_1(x,\{x\})}{1-P_1(x,\{x\})}$$

DR (iii) and $P_t(x,A)\in(0,1)$ for all $x\in A\subseteq X^*$ imply $\delta_2(x)>0$. Since $P_1(x,\{x\})=\Delta_1(x)=\delta_1(x)$,

$$\delta_2(x)=\frac{P_2(x,\{x\})-\Delta_1(x,\{x\})}{1-\Delta_1(x)}$$

and let $\Delta_2(x)=1-(1-\delta_1(x))(1-\delta_2(x))$. Re-arranging the terms in (8), we get:

$$P_2(x,\{x\})=\delta_2(x)+P_1(x,\{x\})-\delta_2(x)P_1(x,\{x\})$$

We re-write the above as

$$P_2(x,\{x\})=1-[1-\delta_2(x)-P_1(x,\{x\})+\delta_2(x)P_1(x,\{x\})]$$

which implies

$$P_2(x,\{x\})=1-(1-P_1(x,\{x\}))(1-\delta_2(x))$$

using $\delta_1(x)=P_1(x,\{x\})$ in the above, we get

$$P_2(x,\{x\})=1-(1-\delta_1(x))(1-\delta_2(x))$$

Note that since $\Delta_2(x) = 1 - (1 - \delta_1(x))(1 - \delta_2(x))$,

$$P_2(x, \{x\}) = \Delta_2(x)$$

where $\Delta_2(x)$ is identified. Next, we similarly set $\delta_3(x)$ as follows

$$\delta_3(x) = \frac{P_3(x, \{x\}) - \Delta_2(x,)}{1 - \Delta_2(x)}$$

which implies

$$P_3(x, \{x\}) = \Delta_2(x) + \delta_3(x)[1 - \Delta_2(x)]$$

similar to the steps above for $P_2(x, \{x\})$), we re-write

$$P_3(x, \{x\}) = 1 - [1 - \Delta_2(x) - \delta_3(x)(1 - \Delta_2(x))]$$

which implies

$$P_3(x, \{x\}) = 1 - [(1 - \delta_3(x))(1 - \Delta_2(x))]$$

Setting $\Delta_3(x) = 1 - [(1 - \delta_3(x))(1 - \Delta_2(x))]$, we get

$$P_3(x, \{x\}) = \Delta_3(x)$$

Similarly, we construct $\delta_4(x), \delta_5(x), \dots \delta_k(x)$ for all $k \leq t$. In general, for any $k \leq t$, let

$$\delta_k(x) = \frac{P_k(x, \{x\}) - \Delta_{k-1}(x)}{1 - \Delta_{k-1}(x)}$$

where $\Delta_l(x) = 1 - \prod_{l \leq k}(1 - \delta_l(x))$ for any $l \leq t$. This implies

$$P_k(x, \{x\}) = \Delta_{k-1}(x) + \delta_k(x)[1 - \Delta_{k-1}(x)] \tag{8}$$

Simplifying the above using the expression for Δ in terms of δ:

$$P_k(x, \{x\}) = \delta_k(x)[1 - \{1 - \prod_{l<k}(1 - \delta_l(x))\}] + 1 - \prod_{l<k}(1 - \delta_l(x))$$

Simplifying the above, we get

$$P_k(x, \{x\}) = 1 - \prod_{l \leq k}(1 - \delta_l(x)) = \Delta_k(x)$$

Therefore, for any time period t and any $x \in X$:

$$\delta_t(x) = \frac{P_t(x, \{x\}) - \Delta_{t-1}(x)}{1 - \Delta_{t-1}(x)}$$

where $\Delta_t(x) = 1 - (1 - \delta_t(x))(1 - \Delta_{t-1}(x))$ and

$$P_t(x, \{x\}) = \Delta_t(x) \tag{9}$$

The above holds for all $x \in X$. Using (9) in (7) for x, and for all w, $w \succ x$, we get

$$P_t(x, A) = \Delta_t(x) \prod_{w \succ x : w \in A} (1 - \Delta_t(w))$$

We now show necessity of DR (i), (ii), (iii) and DMI. Suppose that P is a DAR and $\succ$ is a strict, complete binary relation. Consider any $x, y \in A \subseteq X^*$. Without loss of generality, assume $y \succ x$. By the definition of DAR, for any t,

$$\frac{P_t(x, A \setminus y)}{P_t(x, A)} = \frac{\Delta_t(x) \prod_{z \succ x; z \in A \setminus y} (1 - \Delta_t(z))}{\Delta_t(x) \prod_{z \succ x; z \in A} (1 - \Delta_t(z))} = \frac{1}{(1 - \Delta_t(y))} > 1$$

since $\Delta_t(y) \in (0, 1)$. Since $y \succ x$, for any $t = 1, 2, \ldots, n$

$$\frac{P_t(y, A \setminus x)}{P_t(y, A)} = \frac{\Delta_t(y) \prod_{z \succ y; z \in A \setminus x} (1 - \Delta_t(z))}{\Delta_t(y) \prod_{z \succ y; z \in A} (1 - \Delta_t(z))} = 1.$$

Therefore, DR (i) holds. Without loss of generality, suppose that $y \succ x$ for some $x, y \in A \subseteq X^*$ and $\frac{P_t(x, A \setminus y)}{P_t(x, A)} \neq 1$ for some t. Using the definition of the DAR in the above,

$$\frac{P_t(x, A \setminus y)}{P_t(x, A)} = \frac{\Delta_t(x) \prod_{z \succ x; z \in A \setminus y} (1 - \Delta_t(z))}{\Delta_t(x) \prod_{z \succ x; z \in A} (1 - \Delta_t(z))} = \frac{1}{(1 - \Delta_t(y))}$$

Now, consider some $k \in \{1, 2, 3, \ldots, (n - t)\}$. Suppose that

$$\frac{P_{t+k}(x, A \setminus y)}{P_t(x, A)} = 1$$

By the definition of DAR,

$$\frac{\Delta_{t+k}(x) \prod_{z \succ x; z \in A \setminus y} (1 - \Delta_{t+k}(z))}{\Delta_{t+k}(x) \prod_{z \succ x; z \in A} (1 - \Delta_{t+k}(z))} = \frac{1}{(1 - \Delta_{t+k}(y))} \neq 1$$

which contradicts our assumption. Therefore, DR (ii) holds.

Consider any $x \in X^*$ and $t < n$. By DAR, $P_t(x, \{x\}) = 1 - \prod_{k \leq t} (1 - \delta_k(x))$ and $P_{t+1}(x, \{x\}) = 1 - \prod_{k' \leq t+1} (1 - \delta_{k'}(x))$. Since $\delta_k, \delta_{k'} \in (0, 1)$ for all k, k', we get $P_{t+1}(x, \{x\}) > P_t(x, \{x\})$. Therefore, DR (iii) is necessary.

Notice that for any $x, y \in A$, $A \subseteq X^*$ such that $y \succ x$, by the definition of DAR, $\frac{P_t(x, A \setminus y)}{P_t(x, A)} = 1$ for all $A \subseteq X^*$, or $\frac{P_t(x, A \setminus y)}{P_t(x, A)} = \frac{1}{(1 - \Delta_t(y))}$ for all $A \subseteq X^*$. Therefore, for any $A, B \in X^*$ such that $x, y \in A \cap B$, DMI is satisfied. ■

3 Discussion

In this section we outline some observations and potential applications of DAR.

Choice probabilities as cumulative attention: Note that the proof of theorem 1 provides a decomposition of $P_t(x, \{x\})$ in terms of the cumulative attention prior to t (Δ_{t-1}) and the conditional attention parameter for time period t. In general, for any t, we rewrite (8) as:

$$P_t(x, \{x\}) = \Delta_{t-1}(x) + \delta_t(x)[1 - \Delta_{t-1}(x)] = \delta_t(x) + \Delta_{t-1}(x) - \delta_t(x)\Delta_{t-1}(x)$$

where Δ_{t-1} and δ_t are the cumulative and conditional attention parameters as described earlier. The above computes the choice probability of x when no other alternative is available as the sum of the probability that x attracts the DM's attention either in time period t i.e. $\delta_t(x)$ or earlier i.e. $\Delta_{t-1}(x)$. The term $\delta_t(x)\Delta_{t-1}(x)$ is subtracted to rule out double counting of the attention probability for the tth time period. Consider any t, the above explains choice probability of x in isolation depending on past attention and the attention in t. Since Δ_{t-1} is identifiable from data of $1, 2, \ldots, t-1$ a seller (or marketeer) can easily adapt the above to form an expectation of the probability that x is sold in t, where only $\delta_t(x)$ is unknown. Further, different marketing policies may be tested based on the identified trends in δ_t, Δ_t as t varies.

Convergence as t increases: The choice probabilities converge to the deterministic case in DAR at t increases. In order to see this consider any $x \in A$. By the definition of DAR, $P_t(x, A) = \Delta_t(x) \prod_{y \in A \setminus \{x\}; y \succ x}(1 - \Delta_t(y))$. As $t \to \infty$, DR(iii) implies $\Delta_t(x), \Delta_t(y) \to 1$. Therefore, $P_t(x, A) \to 0$ if $y \succ x$ for any $y \in A \setminus \{x\}$. Note that $P_t(y, A) \to 1$ as $t \to \infty$ if $y \succ x$ for all $x \in A \setminus \{y\}$. As the number of time periods increase, the choice probabilities of all alternatives except the $\succ -max$ alternative converges to 0. The probability of the $\succ -max$ alternative converges to 1. Clearly, in the limit choice is converging to the deterministic case where one alternative is chosen with full probability. This observation is intuitive as over time, the DM eventually has searched and paid attention to all the alternatives and comes close to choosing the best (according to $\succ$) alternative with probabilities close to 1.

A Independence of axioms

DR (i): Let P_t be a dynamic stochastic choice rule such that for all $x \in A$, $A \subseteq X$,

$$P_t(x, A) = \delta_t(x) \prod_{y \in A \setminus \{x\}} \delta_t(y)$$

where $\delta_t : X \to [0, 1]$ are attention probabilities in time period t. This rule violates DR(i) but satisfies DR(ii),(iii) and DMI.

DR (ii): Consider the following dynamic stochastic choice rule: for all $x \in A$, $A \subseteq X$,

$$P_t(x, A) = \begin{cases} \delta_t(x) \prod_{y \succ x | y \in A} (1 - \delta_t(y)), & t \leq k \\ \delta_t(x), & t > k \end{cases}$$

for some $k < n$ where $\delta_t : X \times (0, 1)$ are the attention probabilities in time period t and $\succ$ is complete total order. This rule satisfies DR(i), DR (iii) and DMI, but violates DR(ii) for $t > k$.

DR (iii): A simple version of Manzini and Mariotti (2014)'s rcs-rule demonstrates the independence of DR(iii): for all $x \in A$, $A \subseteq X$,

$$P_t(x, A) = \delta(x) \prod_{y \succ x | y \in A \setminus \{x\}}$$

where $\delta : X \to (0, 1)$ are time-invariant attention probabilities and $\succ$ is a complete total order. The above rule satisfies DR(i)y, DR (ii) and DMI which are similar axioms to those in Manzini and Mariotti (2014). It does not satisfy DR (iii) which is crucial in our model.

DMI: Consider the following dynamic stochastic choice rule: for all $x \in A$, $A \subseteq X$,

$$P_t(x, A) = \frac{\delta_t(x)}{|A|}$$

where $\delta_t : X \to (0, 1)$ is the attention probability of the alternatives in time t. Notice that For any A, B such that $|A| \neq |B|$, DMI will be violated. However, this rule satisfies DR(i), DR (ii) and (iii).

References

Aguirregabiria, V. and P. Mira (2010): "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156, 38–67, structural Models of Optimization Behavior in Labor, Aging, and Health.

Ahn, D. S., F. Echenique, K. Saito, et al. (2017): "On path independent stochastic choice." *Theoretical Economics*, 13, 61–85.

Barberà, S. and P. K. Pattanaik (2010): "Falmagne and the rationalizability of stochastic choices in terms of random orderings," *Econometrica*, 118, 274–299.

Bhattacharya, M., S. Mukherjee, and R. Sonal (2021): "Frame-based stochastic choice rule," *Journal of Mathematical Economics*, 97, 102553.

Block, H. D., J. Marschak, et al. (1959): "Random orderings and stochastic theories of response," Tech. rep., Cowles Foundation for Research in Economics, Yale University.

Caplin, A. and D. Martin (2018): "Framing as information design," *www.martinonline.org*.

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

Chandon, P., J. W. Hutchinson, E. T. Bradlow, and S. H. Young (2009): "Does In-Store Marketing Work? Effects of the Number and Position of Shelf Facings on Brand Attention and Evaluation at the Point of Purchase," *Journal of Marketing*, 73, 1–17.

Fershtman, D. and A. Pavan (2025): "Searching for "Arms": Experimentation with Endogenous Consideration Sets," *Working Paper*.

Frick, M., R. Iijima, and T. Strzalecki (2019): "Dynamic Random Utility," *Econometrica*, 87, 1941–2002.

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

Kashaev, N., V. H. Aguiar, M. Plávala, and C. Gauthier (2023): "Dynamic and Stochastic Rational Behavior," .

Luce, R. (1959): *Individual choice behavior: a theoretical analysis*, John Wiley and Sons.

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

——— (2015): "Modelling imperfect attention," *Working paper 744, Queen Mary University, London.*

——— (2018): "Dual random utility maximisation," *Journal of Economic Theory*, 177, 162–182.

Masatlioglu, Y. and T. P. Vu (2024): "Growing attention," *Journal of Economic Theory*, 222, 105929.

Plott, C. R. (1973): "Path independence, rationality and social choice," *Econometrica*, 41, 1075–1091.

Shapiro, S., D. J. Macinnis, and S. E. Heckler (1997): "The Effects of Incidental Ad Exposure on the Formation of Consideration Sets," *Journal of Consumer Research*, 24, 94–104.

Suleymanov, E. (2018): "Stochastic attention and search," Tech. rep., Working Paper, University of Michigan.

Urban, G. L. (1975): "Perceptor: A Model for Product Positioning," *Manage. Sci.*, 21, 858–871.

Wright, P. and F. Barbour (1977): "Phased decision strategies: sequels to an initial screening," *Starr, M., Zeleny, M. (Eds.), Studies in Management Sciences, Multiple Criteria Decision Making.*, 21, 91–109.

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