Lecture 4

Attention sets

Dr. Alexander Karpov

Attention model
Masatlioglu, Nakajima, Ozbay 2012

- The revealed preference argument relies on the implicit assumption that a decision maker (DM)considers all feasible alternatives. Without the full consideration assumption, the standard revealed preference method can be misleading. It is possible that the DM prefers x to y but she chooses y when x is present simply because she does not realize that x is also available.

- For example, while using a search engine, a DM might only pay attention to alternatives appearing on the first page of the results since it takes too much time to consider all the search results. She then picks the best alternative of those on the first page, say y. It is possible that her most preferred item, x, does not appear on the first page.

Attention filters

- It has been argued that due to cognitive limitations, DMs cannot pay attention to all the available alternatives.
- A consideration set mapping Γ is an attention filter if for any s, $\Gamma(s) = \Gamma(s \setminus x)$whenever $x \notin \Gamma(s)$.
- This definition says that if an alternative does not attract the attention of the decision maker, her consideration set does not change when such an item becomes unavailable.

- Suppose she knows S is her entire feasible set. Then, she picks her consideration set $\Gamma(s)$ based optimally on her prior beliefs about the value of alternatives and the cost of inspecting them.
- A choice function c is a choice with limited attention (CLA) if there exists a complete and transitive preference $\succ$ over X and an attention filter Γ such that $c(s)$ is the $\succ$ - best element in $\Gamma(s)$.

Attention filters

Top N: A DM considers only top N alternatives according to some criterion that is different from her preference.

Top on each criterion: A DM has several criteria and considers only the best alternative(s)on each criterion (modeled as a complete and transitive binary relation).

Most popular category: A DM considers alternatives that belong to the most popular "category" in the market.

TABLE 1—TWO POSSIBLE REPRESENTATIONS FOR THE CYCLICAL CHOICE

Preference		Attention filter			
		$\{x, y, z\}$	$\{x, y\}$	$\{y, z\}$	$\{x, z\}$
$z \succ_1 x \succ_1 y$	Γ_1	xy	xy	y	xz
$x \succ_2 y \succ_2 z$	Γ_2	xyz	xy	yz	z

$$c(xyz) = x, \quad c(xy) = x, \quad c(yz) = y, \quad c(xz) = z.$$

Example

There are four products x, y, z, and t. Each of them is packed in a box. Consider a supermarket that displays these products in its two aisles according to the following rules:

(i)Each aisle can carry at most two products;

(ii) x and y cannot be placed into the same aisle because they are packed in big boxes;

(iii)The supermarket fills the first aisle first and uses the second aisle only if it is necessary;

(iv) y and z are put into the first aisle whenever they are available;

(v) t is placed in the first aisle only after all other available items are put in an aisle and the first aisle still has a space.

Example

- Consider a costumer with preference $t \succ x \succ z \succ y$ (not observable) and she visits only the first aisle and picks her most preferred item displayed in that aisle.
- Suppose all of four products are available. Then, y and z are placed in the first aisle, so z is chosen. When y becomes unavailable, then x is moved to the first aisle and is chosen. Furthermore, when z is also sold out, then x and t are placed in the first aisle, so she picks t. In sum, her choices will be $c(x\ y\ z\ t) = z$, $c(x\ z\ t) = x$ and $c(x\ t) = t$.

DEFINITION 3: *Assume c is a choice by limited attention and there are k different pairs of preference and attention filter representing c*, $(\Gamma_1, \succ_1), (\Gamma_2, \succ_2), \ldots, (\Gamma_k, \succ_k)$. *In this case*,

- *x is revealed to be preferred* to *y if $x \succ_i y$ for all i*,
- *x is revealed to attract attention at S if $\Gamma_i(S)$ includes x for all i*,
- *x is revealed* **not** *to attract attention at S if $\Gamma_i(S)$ excludes x for all i*.

Stochastic Choice and Consideration Sets. Manzini, Mariotti, 2014

- We consider a boundedly rational agent who maximises a preference relation but makes random choice errors due to imperfect attention.
- The source of choice errors in our model is the agent’s failure to consider all feasible alternatives. For example, a consumer buying a new PC is not aware of all the latest models and specifications; a time-pressured doctor overlooks the relevant disease for the given set of symptoms; an ideological voter deliberately ignores some candidates independently of their policies.

Model

Once a consideration set C (A) has been formed, a final choice is made by maximising a preference relation over C (A), which we assume to be standard (complete and transitive).

Each alternative a is considered with a probability γ(a), the attention parameter relative to alternative a. For example, γ(a) may indirectly measure the degree of brand awareness for a product, or the willingness of an agent to seriously evaluate a political candidate.

The assumption that γ(a) is menu independent is a substantive one.

Model

- We allow the agent not to pick any alternative from a menu, so we also assume the existence of a default alternative a^* (e.g. walking away from the shop, abstaining from voting, exceeding the time limit for a move in a game of chess).
- Denote $X^* = X \cup \{a^*\}$ and $A^* = A \cup \{a^*\}$ for all A.

Random choice rule

- p (a; A) denotes the probability that the alternative $a \epsilon A^*$ is chosen when the possible choices (in addition to the default a*) faced by the agent are the alternatives in A. Note that a* is the action taken when the menu is empty, so that $p(a^*; \emptyset) = 1$.

Random consideration set rule

- A random consideration set rule is a random choice rule, for which there exists a pair $(\succ;\gamma)$, where $\succ$ is a strict total order on X and γ is a map γ: X $\longrightarrow$(0; 1), such that

$$p_{\succ,\gamma}(a, A) = \gamma(a) \prod_{b \in A : b \succ a} (1 - \gamma(b)) \;\; for\ all\ A \in \mathcal{D},\ for\ all\ a \in A$$

Theorem

A random choice rule satisfies i-Asymmetry and i-Independence if and only if it is a random consideration set rule $p_{\succ,\gamma}$. Moreover, both $\succ$ and γ are unique, that is, for any random choice rule $p_{\succ',\gamma'}$ such that rule $p_{\succ',\gamma'}$ =p we have $(\succ',\gamma')= (\succ,\gamma)$.

i-Asymmetry says that if b is not neutral for a in a menu, then a must be neutral for b in the same menu. Note how this axiom rules out randomness due to 'utility errors', while it is compatible with 'consideration errors'. It is a stochastic analog of a property of rational deterministic choice: if the presence of b determines whether a is chosen, then b is better than a, and therefore the presence of a cannot determine whether b is chosen.

i-Independence states that the impact of an alternative on another cannot depend on which other alternatives are present in the menu.

Menu-Dependent Stochastic Feasibility Brady and Rehbeck, 2016

Consider a researcher with scanner data on consumer choice from repeated visits to a grocery store. In addition, each store supplies the researcher with the list of offered alternatives. However, there is random variation of alternatives *available* to consumers that is unknown to the researcher. For example, the researcher may not know if alternatives were sold out, a delivery was delayed, or an aisle was closed. A rational consumer's choices will depend on the available alternatives. Therefore, random variation in availability will cause a rational consumer's choices to appear stochastic to the researcher. Hence, stochastic feasibility induces a stochastic choice function.

Random Conditional Choice Set Rule (RCCSR)

We consider a full support probability distribution π on D. Thus, there is a positive probability each $A \in D$ is feasible. We call $F(A)$ the *feasible set*. When $D = 2^X$, $\pi(A)$ represents the probability that A is feasible in X. For a menu A, the probability of facing the feasible set $B \subseteq A$ is

$$Pr(F(A) = B) = \frac{\pi(B)}{\sum_{C \subseteq A} \pi(C)}.$$

If B is not a subset of A, then $Pr(F(A) = B) = 0$. Thus, the probability of facing a given feasible set is conditioned on the offered menu.

Random conditional choice set rule

- We allow $F(A)$ to be empty, in which case the agent chooses the default option $x*$. Therefore, $P(x*,A)$ is the probability that $F(A)$ is empty.

Random conditional choice set rule

A *random conditional choice set rule* (RCCSR) is a random choice rule $P_{\succ,\pi}$ for which there exists a pair ($\succ$,π), where $\succ$ is a strict preference ordering on X and $\pi : D \rightarrow (0, 1)$ a full support probability distribution over D, such that for all $A \in D$ and for all $a \in A$

$$P_{\succ,\pi}(a, A) = \frac{\sum_{B \in A_a} \pi(B)}{\sum_{C \subseteq A} \pi(C)}.$$

Thus, $P_{\succ,\pi}(a,A)$ is the probability that a is the best item available when offered menu A. Menu-dependence is clear since $Pr(F(A) = B)$ is conditioned on the subsets of the offered menu.

Sequential independence

- Let the *default option* be $x* \notin X$. The default option is available for each menu and can be interpreted as "not chosen". When the menu is empty, the default option $x*$ is always chosen, so $P(x*, \emptyset) = 1$.

We say that alternative *b* is *sequentially independent* from alternative *a* in menu *{a,b}* if

$$P(b, A) = P\big(b, A \setminus \{a\}\big)P\big(A^* \setminus \{a\}, A\big).$$

the term P(A$*$ \{a}A) is the probability a is not available in A.

Sequential independence

- We see that the probability *b* is chosen from the set $A \setminus \{a\}$ is the same as the probability *b* is chosen from *A* conditional on *a* being sold out.

$$P(b, A \setminus \{a\}) = \frac{P(b, A)}{1 - P(a, A)}.$$

Identification

- Now suppose that choice data is generated by an RCCSR. We focus on how to recover $\pi(\cdot)$. First, consider the case that D = 2X. We note that $\pi(\emptyset) = P(x*, X)$ since $x*$ is chosen when the feasible set is empty and probabilities are unconditional in X. Using this, for an arbitrary singleton menu of $a \in X$ we have

$$P(x^*, \{a\}) = \frac{\pi(\emptyset)}{\pi(\{a\}) + \pi(\emptyset)} \Rightarrow \pi(\{a\}) = \frac{P(x^*, X)}{P(x^*, \{a\})} - P(x^*, X).$$

Example (Grocery Store)

Consider a researcher with scanner data of a consumer from several grocery stores. The alternatives of interest are apples (*a*), bananas (*b*), and carrots (*c*). Here the set of alternatives is $X = \{a, b, c\}$ and $D = 2X$. Suppose we observe choice from all possible nonempty menus given by Table 1.

Table 1: Grocery Store Stochastic Choice Data

	$\{a,b,c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\{a\}$	$\{b\}$	$\{c\}$
a	$7/20$	$1/3$	$1/2$	0	$1/2$	0	0
b	$11/20$	$1/2$	0	$11/13$	0	$3/4$	0
c	$1/20$	0	$1/4$	$1/13$	0	0	$1/2$

- It would be reasonable to think correlation exists between which objects are feasible. Correlation would mean that $Pr(a \in F(A) | b \in F(A)) \neq Pr(a \in F(A))$ for some $a, b \in A$ with $a \neq b$. We note that a random consideration set rule does not allow these effects.
- One can use this data and our revealed preference relation to find that $a \succ b \succ c$ and that the π system is given by

$$\pi(\emptyset) = \frac{1}{20} \quad \pi(\{a\}) = \frac{1}{20} \quad \pi(\{b\}) = \frac{3}{20} \quad \pi(\{c\}) = \frac{1}{20}$$

$$\pi(\{a,b\}) = \frac{1}{20} \quad \pi(\{a,c\}) = \frac{1}{20} \quad \pi(\{b,c\}) = \frac{8}{20} \quad \pi(\{a,b,c\}) = \frac{4}{20}.$$

- Suppose that a researcher observes *b* is available when the agent chooses from *X*. Now, the researcher can back out the probability that *a* was also in the feasible set since

$$P(a \in F(X) \mid b \in F(X)) = \frac{\pi(\{a,b\}) + \pi(\{a,b,c\})}{\pi(\{a,b\}) + \pi(\{a,b,c\}) + \pi(\{b,c\}) + \pi(\{b\})} = \frac{5}{16}$$

- but $P(a \in F(X)) = 7/20$. This may suggest that apples and bananas are substitutes since apples are less likely to be available given bananas are still available.

Random choice as behavioral optimization
Gul, Natenzon, Pesendorfer, 2014

Subjects routinely violate the weak axiom of revealed preference. Often, these violations occur in a manner inconsistent with any deterministic theory.

In fact, empirical and experimental studies almost always interpret individual choice behavior as probabilistic.

Behavioral optimization

- Gul, Natenzon, Pesendorfer, 2014 explore random choice as a theory of *behavioral optimization*, that is, not as a model of measurement error but as a model of a consumer whose rationality is constrained by behavioral limitations such as limited cognitive abilities or limited attention.

The Luce rule

- Each option s has a Luce value, v_s, so that the probability of choosing s from a set A containing s is

$$\rho_s(A) := \frac{v_s}{\sum_{t \in A} v_t}.$$

The Luce rule

- We can interpret the Luce value as a measure of desirability: s is *stochastically preferred* to t if, for any set A that contains neither s nor t, the agent is more likely to choose s from A ∪ {s} than t from A ∪ {t}. Luce values represent this stochastic preference: s is stochastically preferred to t if and only if $v_s \geq v_t$.

Systematic violations of the Luce model

Debreu (1960) anticipated the best known violation and identified the main shortcoming of Luce's model: consider two items s_1 and s_2 that are very similar (a yellow bus and a red bus) and a third dissimilar option t (a train). Then, it may be that each item is chosen with probability 1/2 from every two-element subset of $\{s_1\ s_2\ t\}$, but t is chosen from {s1 s2 t} more frequently than each of the other two options. It is easy to check that this behavior cannot be generated (nor approximated) by any Luce rule. The problem that Debreu's example identifies is more generally referred to as the "duplicates problem" in the discrete choice estimation literature.

Attribute rule

This model, the *attribute rule*, addresses the shortcomings of the Luce model but retains Luce's idea that choice is governed by desirability values. It does so by reinterpreting the choice objects as *bundles of attributes.* Attributes, or at least their relevance, are subjective; they are properties of the decision maker and not of the choice objects.

Let Z be the collection of attributes, let X_s be the set of attributes that s has and

$$X(A) = \bigcup_{s \in A} X_s.$$

Attribute rule

Let $X(A) := \{x \in Z \mid \eta^x(A) > 0\}$ for all $A \in \mathcal{A}_+$.

An *attribute value* is a function w : Z → R++ that measures the desirability of attribute x.

$$w(X) := \sum_{x \in X} w_x$$

We say that η is *simple* if it is equal to 0 or 1 for all x,s. The choice rule ρ is a (complete) attribute rule if there exists a (complete) attribute system (w, η) such that

$$\rho_s(A) = \sum_{x \in X(A)} \frac{w_x}{w(X(A))} \cdot \frac{\eta_s^x}{\eta^x(A)}.$$

- In an attribute rule, the decision maker first chooses a relevant attribute according to a Luce-type formula and then picks one option that has that attribute according to another Luce-type formula. The attribute rule reduces to a Luce rule when no pair of alternatives shares a common attribute.

Example

Let A = {r,s,t} and assume there are three attributes, Z ={1, 2, 3}. Each attribute value is 1, that is, w_x = 1 for all x ∈ Z. Option r has attributes 1, 3, s has attributes 1, 2, and t has attributes 2, 3. In particular, $\eta_r^1 = \eta_s^2 = \eta_t^3 = 4$, $\eta_r^3 = \eta_s^1 = \eta_t^2 = 1$, and $\eta_r^2 = \eta_s^3 = \eta_t^1 = 0$. This attribute system represents the choice rule ρ such that

$$\rho_r(\{r, s\}) = \rho_s(\{s, t\}) = \rho_t(\{r, t\}) = 3/5.$$

Problem 1

Let X be a finite set. A **random choice function** is a function $P : 2^X \setminus \emptyset \to [0,1]^X$ such that $P(A)(x) \geq 0$ and $\sum_{a \in A} P(A)(a) = 1$. For notational ease, let $P_A = P(A)$. That is, a random choice function takes a menu of options and outputs a probability distribution over the menu, where $P_A(x)$ denotes the probability that x is chosen from menu A.

A random choice function admits a **Luce representation** if there exists a set of weights $\{w(x) \geq 0 : x \in X\}$ such that

$$P_A(x) = \begin{cases} \frac{w(x)}{\sum_{a \in A} w(a)} & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

A random choice function satisfies **Random Independence of Irrelevant Alternatives** if

$$\frac{P_A(x)}{P_A(y)} = \frac{P_B(x)}{P_B(y)}$$

whenever $P_A(x), P_A(y), P_B(x), P_B(y) > 0$.

Problem from http://eml.berkeley.edu/~dahn/ProblemSet1.pdf

(b) Prove or provide a counterexample to the following statement: If P admits a Luce representation, then it satisfies Random IIA.

(c) Suppose $P_A(a) > 0$ whenever $a \in A$. Prove that if P satisfies Random IIA, then P admits a Luce representation. (Hint: Consider a candidate for w.)

(d) For $A, B \subseteq X$, define $P_A(B)$ as $\sum_{b \in B} P_A(B)$. Suppose that P admits a Luce representation. Prove that if $C \subseteq B \subseteq A$, then $P_A(C) = P_B(C) P_A(B)$. Interpret this condition.

Problem 2

- Consider "Stochastic Choice and Consideration Sets. Manzini, Mariotti, 2014" model. Assume that preference relation is not a linear order, but a weak order. If consideration set consist of several indifferent alternatives then DM chooses one of them with equal probabilities.
- Rationalise examples from "Random choice as behavioral optimization. Gul, Natenzon, Pesendorfer, 2014" by generalised Manzini-Mariotti model.

Examples

EXAMPLE 5: To define an attribute rule that is consistent with Example 3 above, let $Z = \{x, y\}$, where $x = A$ is the bus-attribute and $y = \{t\}$ is the train-attribute. Let $w_x = 3$, $w_y = 2$ and let η be the simple intensity such that $\eta_s^x = 1$ if and only if $s \in x = A$ and $\eta_s^y = 1$ if and only if $s = t$. Then, $\rho_{s_i}(\{s_i, t\}) = 0.6$ and $\rho_{s_i}(At) = 0.2$, as required.

EXAMPLE 6: For an attribute rule that is consistent with Example 4, let $Z = \{x, y\}$, where $x = \{r, s\}$ and $y = \{r, t\}$. Set $w_x = w_y = 1$ and let η be the simple attribute intensity such that $\eta_r^x = \eta_s^x = 1$, $\eta_r^y = \eta_t^y = 1$. Then, $\rho_r(\{r, s\}) = \rho_r(\{r, t\}) = 3/4$ and $\rho_r(\{r, s, t\}) = 1/2$, as required.

- Example 8

$$\rho_r\big(\{r, s\}\big) = \rho_s\big(\{s, t\}\big) = \rho_t\big(\{r, t\}\big) = 3/5.$$

Problem 3

Consider example from Manzini, Mariotti, 2014

Example 2 *Let* $\gamma(a) = \frac{4}{9}$, $\gamma(b) = \frac{1}{2}$ *and* $\gamma(c) = \frac{9}{10}$ *with* $a \succ b \succ c$. *We have:*

$$
\begin{aligned}
p_{\succ,\gamma}(b, \{b, c\}) &= \frac{1}{2} \\
p_{\succ,\gamma}(c, \{a, c\}) &= \frac{9}{10}\frac{5}{9} = \frac{1}{2}
\end{aligned}
$$

but also

$$p_{\succ,\gamma}(b, \{a, b\}) = \frac{1}{2}\frac{5}{9} = \frac{5}{18} < \frac{1}{2}$$

violating Weak stochastic transitivity.

Is it possible to rationalise this choice by Gul, Natenzon, Pesendorfer, 2014 model?