BEHAVIORAL ECONOMICS · LECTURE 3

Stochastic choice

Niels Boissonnet – Bielefeld University

Master Economics – Winter 2026/27

Before we start: a quick vote

Which do you prefer?

A €14 for sure.

B €30 with probability $\frac{1}{2}$, otherwise nothing.

Before we start: a quick vote

Which do you prefer?

A €14 for sure.

B €30 with probability $\frac{1}{2}$, otherwise nothing.

Questions

- Write down your answer. **How sure are you?**
- Would you give the **same answer** next week? After lunch?

Learning objectives

By the end of this lecture you should be able to:

- Define a **stochastic choice function** and list reasons why choices may be random.
- Define the **random utility model** (RUM) and prove that it satisfies **regularity**.
- Explain why a population of rational individuals can produce **intransitive** choice probabilities.
- State **Luce's axiom** (IIA), derive the **logit** formula and compute logit probabilities.
- Explain the **red bus** / **blue bus** problem and how probit, nested and mixed logit address it.

Roadmap

Choices are random

Random utility

Luce’s model and logit

Wrapping up

Where we are in the course

The assumptions of the Robinson model

0. **Consistent preferences:** Robinson's choices can be described by a utility function u
1. **Costless optimization:** he always finds the option that maximizes u
2. **Correct beliefs & expected utility:** he knows p and maximizes expected utility
3. **Absolute, self-regarding consumption:** only his own (b, y) enters u
4. **Stable tastes:** the same u in both periods

Take-away

Lecture 2 described choices by a choice function: one menu, one choice. But the **same** person facing the **same** menu does not always choose the same option. Today: a theory of **choice probabilities**. Randomness can come from tastes (random utility) or from imperfect optimization (noise, Assumption 1).

1

Choices are random

The same person, facing the same menu, does not always choose the same option.

Evidence: same question, different answers

Repeated choices

- Mosteller and Nogee (1951): the probability of taking a bet rises **smoothly** with its payoff.
- Hey and Orme (1994): the same 100 pairs of lotteries, presented twice on different days.
  - About **a quarter** of the choices change.
- Agranov and Ortoleva (2017): the same question **three times in a row**, on one screen.
  - Many subjects choose differently, and say they do it **on purpose**.

Evidence: same question, different answers

Repeated choices

- Mosteller and Nogee (1951): the probability of taking a bet rises **smoothly** with its payoff.
- Hey and Orme (1994): the same 100 pairs of lotteries, presented twice on different days.
  - About **a quarter** of the choices change.
- Agranov and Ortoleva (2017): the same question **three times in a row**, on one screen.
  - Many subjects choose differently, and say they do it **on purpose**.

Choice data are **probabilistic**. A deterministic model ($C(A) = \arg\max_A u$) treats every switch as a violation of WARP. We need models that predict **how often** each option is chosen.

Stochastic choice functions

Definition

Let X be a finite set of alternatives. A **stochastic choice function** p assigns to each menu $A \subseteq X$ a probability distribution over A:

$$p(x \mid A) \geq 0, \qquad \sum_{x \in A} p(x \mid A) = 1.$$

$p(x \mid A)$ = probability that x is chosen from A.

- Deterministic choice is the special case $p(x \mid A) \in \{0, 1\}$.
- For binary menus we write $p(x, y) = p(x \mid \{x, y\})$.

Where do the data come from?

- **Individual** data: the same person, many repetitions.
- **Population** data: market shares, many people facing the same menu.

Stochastic choice functions

Definition

Let X be a finite set of alternatives. A **stochastic choice function** p assigns to each menu $A \subseteq X$ a probability distribution over A:

$$p(x \mid A) \geq 0, \qquad \sum_{x \in A} p(x \mid A) = 1.$$

$p(x \mid A)$ = probability that x is chosen from A.

- Deterministic choice is the special case $p(x \mid A) \in \{0, 1\}$.
- For binary menus we write $p(x, y) = p(x \mid \{x, y\})$.

Where do the data come from?

- **Individual** data: the same person, many repetitions.
- **Population** data: market shares, many people facing the same menu.

Careful

Population data cannot tell individual randomness from **heterogeneity**: "30% buy x" fits both.

Why might choices be random?

Six stories

1. **Random tastes**: preferences differ across people, or fluctuate over time and context.
2. **Noisy evaluation**: the DM perceives or computes utilities with error.
3. **Mistakes** (“trembles”): she occasionally picks the wrong option by accident.
4. **Random attention**: she does not see all options (Lecture 6).
5. **Heuristics**: rules of thumb applied at random.
6. **Deliberate randomization**: hedging, exploring, avoiding regret.

Why might choices be random?

Six stories

1. **Random tastes**: preferences differ across people, or fluctuate over time and context.
2. **Noisy evaluation**: the DM perceives or computes utilities with error.
3. **Mistakes** ("trembles"): she occasionally picks the wrong option by accident.
4. **Random attention**: she does not see all options (Lecture 6).
5. **Heuristics**: rules of thumb applied at random.
6. **Deliberate randomization**: hedging, exploring, avoiding regret.

Different stories put **different restrictions** on p. Today: the oldest and most used one, **random utility**, and its workhorse, **logit**.

2

Random utility

Preferences are random, but the decision maker always maximizes the preference she has at the moment of choice.

The random utility model (RUM)

Definition (Block and Marschak, 1960)

p is a **random utility model** if there is a probability distribution $\mathbb{P}$ over strict preference orders $\succ$ on X such that

$$p(x \mid A) = \mathbb{P}(\{\succ : \ x \succ y \text{ for all } y \in A \backslash \{x\}\}).$$

- Equivalently: random utilities $U_x = V_x + \varepsilon_x$ (no ties), and x is chosen iff U_x is the largest in A.
- V_x = systematic part, ε_x = random part.

Example: three types

$X = \{\text{yoga}, \text{bourbon}, \text{reading}\}$

$\frac{1}{2}$: yoga $\succ$ bourbon $\succ$ reading

$\frac{1}{4}$: bourbon $\succ$ yoga $\succ$ reading

$\frac{1}{4}$: reading $\succ$ yoga $\succ$ bourbon

The random utility model (RUM)

Definition (Block and Marschak, 1960)

p is a **random utility model** if there is a probability distribution $\mathbb{P}$ over strict preference orders $\succ$ on X such that

$$p(x \mid A) = \mathbb{P}(\{\succ : \ x \succ y \text{ for all } y \in A \backslash \{x\}\}).$$

- Equivalently: random utilities $U_x = V_x + \varepsilon_x$ (no ties), and x is chosen iff U_x is the largest in A.
- V_x = systematic part, ε_x = random part.

Example: three types

$X = \{\text{yoga}, \text{bourbon}, \text{reading}\}$

$\frac{1}{2}$: yoga $\succ$ bourbon $\succ$ reading

$\frac{1}{4}$: bourbon $\succ$ yoga $\succ$ reading

$\frac{1}{4}$: reading $\succ$ yoga $\succ$ bourbon

$p(\text{yoga} \mid X) = \frac{1}{2}$ $\quad$ $p(\text{yoga}, \text{bourbon}) = \frac{3}{4}$

$p(\text{bourbon} \mid X) = \frac{1}{4}$ $\quad$ $p(\text{yoga}, \text{reading}) = \frac{3}{4}$

$p(\text{reading} \mid X) = \frac{1}{4}$ $\quad$ $p(\text{bourbon}, \text{reading}) = \frac{3}{4}$

What does random utility rule out?

The preferences do not depend on the menu

- $p(\text{yoga}, \text{bourbon}) = \alpha$: a share α of preferences rank yoga $\succ$ bourbon.
- Now add *reading* to the menu. The preferences stay the same; only the menu changes.
- Suppose a share $\beta > 0$ has reading $\succ$ yoga $\succ$ bourbon. They now choose reading.
- Yoga keeps only the preferences that rank it above **both** options:

$$p(\text{yoga} \mid \{\text{yoga}, \text{bourbon}, \text{reading}\}) \leq \alpha - \beta < \alpha.$$

In the example: $\alpha = \frac{3}{4}$, $\beta = \frac{1}{4}$, and $p(\text{yoga} \mid X) = \frac{1}{2}$.

What does random utility rule out?

The preferences do not depend on the menu

- $p(\text{yoga}, \text{bourbon}) = \alpha$: a share α of preferences rank yoga $\succ$ bourbon.
- Now add *reading* to the menu. The preferences stay the same; only the menu changes.
- Suppose a share $\beta > 0$ has reading $\succ$ yoga $\succ$ bourbon. They now choose reading.
- Yoga keeps only the preferences that rank it above **both** options:

$$p(\text{yoga} \mid \{\text{yoga}, \text{bourbon}, \text{reading}\}) \leq \alpha - \beta < \alpha.$$

In the example: $\alpha = \frac{3}{4}$, $\beta = \frac{1}{4}$, and $p(\text{yoga} \mid X) = \frac{1}{2}$.

Adding an option can take choosers away from yoga, never bring new ones. This property is called **regularity**.

Every RUM satisfies regularity

Regularity.

p satisfies **regularity** if for all menus $A \subseteq B$ and $x \in A$: $p(x \mid B) \leq p(x \mid A)$.

Every RUM satisfies regularity

Regularity.

p satisfies **regularity** if for all menus $A \subseteq B$ and $x \in A$: $p(x \mid B) \leq p(x \mid A)$.

Result

Every random utility model satisfies regularity.

Every RUM satisfies regularity

Regularity.

p satisfies **regularity** if for all menus $A \subseteq B$ and $x \in A$: $p(x \mid B) \leq p(x \mid A)$.

Result

Every random utility model satisfies regularity.

Proof

If x is the $\succ$-best element of B, it is also the $\succ$-best element of every $A \subseteq B$ that contains it. So

$$\{\succ\text{: } x \text{ best in } B\} \subseteq \{\succ\text{: } x \text{ best in } A\},$$

and the probability of the smaller event is smaller. $\square$

Is regularity enough?

Characterization

- With **three** alternatives (and data on every menu): p is a RUM **if and only if** it satisfies regularity (Block and Marschak, 1960).
- With **four or more**: regularity is necessary but **insufficient**. The axiom requires the **Block–Marschak polynomials** to be nonnegative (Falmagne, 1978)

$$\mathbb{P}(y \succ x \text{ iff } y \in X \setminus A) = \sum_{B \supseteq A} (-1)^{|B \setminus A|} p(x \mid B) \geq 0$$

Why three is special

With $X = \{x, y, z\}$, the RUM needs the probability that both $y \succ x$ and $z \succ x$ to be nonnegative:

$$1 - p(x, \{x, y\}) - p(x, \{x, z\}) + p(x \mid X) \geq 0.$$

This follows from regularity: $p(y \mid X) \leq 1 - p(x, y)$ and $p(z \mid X) \leq 1 - p(x, z)$. Add them up.

Rational individuals, intransitive population

A Condorcet population

Shares $\frac{1}{3}$ each: $x \succ y \succ z$, $y \succ z \succ x$, $z \succ x \succ y$.

Rational individuals, intransitive population

A Condorcet population

Shares $\frac{1}{3}$ each: $x \succ y \succ z,\ y \succ z \succ x,$ $z \succ x \succ y.$

$$p(x, y) = p(y, z) = p(z, x) = \frac{2}{3},$$

and each option has probability $\frac{1}{3}$ in X.

Rational individuals, intransitive population

A Condorcet population

Shares $\frac{1}{3}$ each: $x \succ y \succ z,\ y \succ z \succ x,$ $z \succ x \succ y.$

$$p(x, y) = p(y, z) = p(z, x) = \frac{2}{3},$$

and each option has probability $\frac{1}{3}$ in X.

Weak stochastic transitivity (WST)

If $p(x, y) \geq \frac{1}{2}$ and $p(y, z) \geq \frac{1}{2}$, then $p(x, z) \geq \frac{1}{2}$.

Rational individuals, intransitive population

A Condorcet population

Shares $\frac{1}{3}$ each: $x \succ y \succ z$, $y \succ z \succ x$, $z \succ x \succ y$.

$$p(x, y) = p(y, z) = p(z, x) = \frac{2}{3},$$

and each option has probability $\frac{1}{3}$ in X.

Weak stochastic transitivity (WST)

If $p(x, y) \geq \frac{1}{2}$ and $p(y, z) \geq \frac{1}{2}$, then $p(x, z) \geq \frac{1}{2}$.

Every individual is a perfect utility maximizer, yet "the majority" prefers x to y, y to z and z to x. A RUM can **violate WST**.

- Intransitive **aggregate** data are not evidence of irrational **individuals**.
- Tversky (1969) found WST violations **within** subjects, with lotteries.

3

Luce’s model and logit

One axiom, independence of irrelevant alternatives, gives the logit formula. It is elegant, tractable, and too restrictive when options are similar.

Luce's choice axiom

Luce's axiom (IIA)

For all menus $A \subseteq B$ and $x, y \in A$ (with positive probabilities):

$$\frac{p(x \mid A)}{p(y \mid A)} = \frac{p(x \mid B)}{p(y \mid B)}.$$

The odds of x against y do not depend on the other options: **independence of irrelevant alternatives**.

Luce's choice axiom

Luce's axiom (IIA)

For all menus $A \subseteq B$ and $x, y \in A$ (with positive probabilities):

$$\frac{p(x \mid A)}{p(y \mid A)} = \frac{p(x \mid B)}{p(y \mid B)}.$$

The odds of x against y do not depend on the other options: **independence of irrelevant alternatives**.

Theorem (Luce, 1959)

If all choice probabilities are positive, p satisfies IIA **iff** there is $w : X \to (0, \infty)$ with

$$p(x \mid A) = \frac{w(x)}{\sum_{y \in A} w(y)}.$$

Moreover, w is unique up to a positive factor.

Proof of Luce's theorem

($\Leftarrow$) Luce form implies IIA

$\frac{p(x \mid A)}{p(y \mid A)} = \frac{w(x)}{w(y)}$ does not depend on A.

Proof of Luce's theorem

($\Leftarrow$) Luce form implies IIA

$\dfrac{p(x \mid A)}{p(y \mid A)} = \dfrac{w(x)}{w(y)}$ does not depend on A.

($\Rightarrow$) IIA implies Luce form

Set $w(x) = p(x \mid X) > 0$. IIA with $B = X$: for all $x, y \in A$, $p(x \mid A) = p(y \mid A)\,\dfrac{w(x)}{w(y)}$.

Sum over $x \in A$: $1 = \dfrac{p(y \mid A)}{w(y)} \sum_{x \in A} w(x)$, so $p(y \mid A) = \dfrac{w(y)}{\sum_{x \in A} w(x)}$.

Proof of Luce's theorem

($\Leftarrow$) Luce form implies IIA

$\dfrac{p(x \mid A)}{p(y \mid A)} = \dfrac{w(x)}{w(y)}$ does not depend on A.

($\Rightarrow$) IIA implies Luce form

Set $w(x) = p(x \mid X) > 0$. IIA with $B = X$: for all $x, y \in A$, $p(x \mid A) = p(y \mid A)\,\dfrac{w(x)}{w(y)}$.

Sum over $x \in A$: $1 = \dfrac{p(y \mid A)}{w(y)} \displaystyle\sum_{x \in A} w(x)$, so $p(y \mid A) = \dfrac{w(y)}{\sum_{x \in A} w(x)}$.

Uniqueness

If w and w' both work: $\dfrac{w(x)}{w(y)} = \dfrac{p(x \mid \{x, y\})}{p(y \mid \{x, y\})} = \dfrac{w'(x)}{w'(y)}$, so $w' = c\,w$ with $c > 0$.

The multinomial logit model

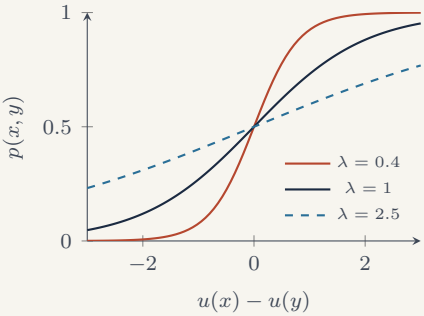
Logit

Write $w(x) = e^{u(x)/\lambda}$, with noise $\lambda > 0$:

$$p(x \mid A) = \frac{e^{u(x)/\lambda}}{\sum_{y \in A} e^{u(y)/\lambda}},$$

Log-odds:
$\log\left[p(x \mid A)/p(y \mid A)\right] = \big(u(x) - u(y)\big)/\lambda.$

- Logit **is** a RUM: $U_x = u(x) + \lambda\varepsilon_x$ with ε_x i.i.d. Gumbel (McFadden, 1974).
- $\lambda \to 0$: maximization; $\lambda \to \infty$: uniform choice.
- Rational inattention foundation (Matějka and McKay, 2015): Lecture 9.



Mistakes are likely when options are **close**, rare when they are far apart (as in Mosteller and Nogee).

The red bus / blue bus problem (Debreu, 1960)

Commuting

A commuter takes the car or the blue bus with probability $\frac{1}{2}$ each. The city adds a **red** bus, identical except for the colour.

The red bus / blue bus problem (Debreu, 1960)

Commuting

A commuter takes the car or the blue bus with probability $\frac{1}{2}$ each. The city adds a **red** bus, identical except for the colour.

	car	blue bus	red bus
Logit (IIA)	$1/3$	$1/3$	$1/3$
Common sense	$1/2$	$1/4$	$1/4$

The red bus / blue bus problem (Debreu, 1960)

Commuting

A commuter takes the car or the blue bus with probability $\frac{1}{2}$ each. The city adds a **red** bus, identical except for the colour.

	car	blue bus	red bus
Logit (IIA)	$1/3$	$1/3$	$1/3$
Common sense	$1/2$	$1/4$	$1/4$

IIA ignores **similarity**: a new option takes market share from **all** others in proportion, not mainly from its close substitutes.

- Debreu's original example: two recordings of the same Debussy piece and one of a Beethoven symphony.

Beyond logit: allowing for similarity

Three random utility models

- **Probit**: $\varepsilon \sim \mathcal{N}(0, \Sigma)$, errors can be **correlated**. Binary choice: $p(x, y) = \Phi((V_x - V_y)/\sigma_{xy})$, with $\sigma_{xy}^2 = \mathrm{Var}(\varepsilon_x - \varepsilon_y)$.
- **Nested logit**: options in **nests** (car; buses). Choose the nest, then the option within it.
- **Mixed logit**: logit with **random coefficients** (heterogeneous tastes). It approximates **any** RUM (McFadden and Train, 2000).

Buses with probit

If the two buses have perfectly correlated errors (ties broken at random), the commuter compares car vs bus: $\frac{1}{2}$, then $\frac{1}{4}$ for each bus.

Beyond logit: allowing for similarity

Three random utility models

- **Probit**: $\varepsilon \sim \mathcal{N}(0, \Sigma)$, errors can be **correlated**. Binary choice: $p(x, y) = \Phi((V_x - V_y)/\sigma_{xy})$, with $\sigma_{xy}^2 = \mathrm{Var}(\varepsilon_x - \varepsilon_y)$.
- **Nested logit**: options in **nests** (car; buses). Choose the nest, then the option within it.
- **Mixed logit**: logit with **random coefficients** (heterogeneous tastes). It approximates **any** RUM (McFadden and Train, 2000).

Buses with probit

If the two buses have perfectly correlated errors (ties broken at random), the commuter compares car vs bus: $\frac{1}{2}$, then $\frac{1}{4}$ for each bus.

But

These are all random utility models: they **all** satisfy regularity. None can explain a choice that **rises** when an option is added: the decoy effect (Lecture 4).

Two views of noise in binary choice

Fechner model

$p(x, y) = F(u(x) - u(y))$ with F strictly increasing, $F(-d) = 1 - F(d)$.

- Mistakes depend on the **utility difference**. Binary logit and probit are of this form.
- Implies **strong stochastic transitivity**: if $p(x, y) \geq \frac{1}{2}$ and $p(y, z) \geq \frac{1}{2}$, then $p(x, z) \geq \max\{p(x, y), p(y, z)\}$ (Exercise 4).

Tremble model

With probability $1 - \varepsilon$ choose the best option; with probability ε choose at random.

- Mistakes do **not** depend on how close the options are.

Two views of noise in binary choice

Fechner model

$p(x, y) = F(u(x) - u(y))$ with F strictly increasing, $F(-d) = 1 - F(d)$.

- Mistakes depend on the **utility difference**. Binary logit and probit are of this form.
- Implies **strong stochastic transitivity**: if $p(x, y) \geq \frac{1}{2}$ and $p(y, z) \geq \frac{1}{2}$, then $p(x, z) \geq \max\{p(x, y), p(y, z)\}$ (Exercise 4).

Tremble model

With probability $1 - \varepsilon$ choose the best option; with probability ε choose at random.

- Mistakes do **not** depend on how close the options are.

The evidence sides with Fechner: people err more when options are close (Mosteller and Nogee, 1951). This is one reason why logit and probit dominate applied work.

4

Wrapping up

Recap

Theory

- Choice data are **probabilistic**: we model $p(x \mid A)$.
- **Random utility**: random preference, perfect maximization. It implies **regularity** (and, with three options, nothing else).
- A RUM can violate **stochastic transitivity**: rational individuals, intransitive population.
- **Logit** = RUM + IIA. Tractable, but blind to **similarity**; probit, nested and mixed logit fix this.

Evidence and next steps

- People switch when the same question is asked again, more so when options are close.
- Next lecture: **decoy effects** violate regularity, so no RUM explains them.
- Lecture 6: randomness in **attention**, not in tastes.

Recap

Theory

- Choice data are **probabilistic**: we model $p(x \mid A)$.
- **Random utility**: random preference, perfect maximization. It implies **regularity** (and, with three options, nothing else).
- A RUM can violate **stochastic transitivity**: rational individuals, intransitive population.
- **Logit** = RUM + IIA. Tractable, but blind to **similarity**; probit, nested and mixed logit fix this.

Evidence and next steps

- People switch when the same question is asked again, more so when options are close.
- Next lecture: **decoy effects** violate regularity, so no RUM explains them.
- Lecture 6: randomness in **attention**, not in tastes.

Vote again

A: €14 for sure, or B: €30 with probability $\frac{1}{2}$? Did you switch?

Exercises (1/2)

1. (Easy)

Logit with $u(x) = 2$, $u(y) = 1$, $u(z) = 0$ and $\lambda = 1$. (a) Compute $p(\cdot \mid \{x, y, z\})$ and $p(\cdot \mid \{x, y\})$. (b) Check IIA for x and y. (c) For which λ is $p(x, y) = 0.9$?

2. (Intermediate)

Prove Luce's theorem: if all probabilities are positive and IIA holds, then $p(x \mid A) = w(x) / \sum_{y \in A} w(y)$ for some $w > 0$. *Hint: try* $w(x) = p(x \mid X)$.

Exercises (2/2)

3. (Intermediate)

A population: 50% $x \succ y \succ z$, 30% $y \succ z \succ x$, 20% $z \succ x \succ y$. (a) Compute all binary probabilities and $p(\cdot \mid \{x, y, z\})$. (b) Check regularity. (c) Does weak stochastic transitivity hold? Strong stochastic transitivity?

4. (Advanced)

(a) Prove that every Fechner model satisfies strong stochastic transitivity. (b) Use Exercise 3 to show that some RUMs are **not** Fechner models. (c) Is logit both?

Questions for discussion

Discuss

- Is random choice a **mistake** or an expression of preference? Does the answer matter for welfare?
- If choices are random, what does a choice "reveal" about preferences? Can choice probabilities measure the **strength** of a preference?
- With market-share data only, can we tell random tastes within a person from heterogeneity across people?
- Why might someone **want** to randomize, as in Agranov and Ortoleva (2017)?

Final take-aways

1. The same person facing the same menu does not always choose the same option: we model **choice probabilities**.
2. **Random utility** (random preference, perfect maximization) implies **regularity**, but not stochastic transitivity.
3. **Logit** follows from IIA: simple and tractable, but it ignores **similarity** between options.
4. Regularity is a **testable** prediction. Next lecture: the **decoy effect** violates it, so no random utility model can explain it.

References I

Agranov, M. and Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1):40–68.

Barberà, S. and Pattanaik, P. K. (1986). Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica*, 54(3):707–715.

Block, H. D. and Marschak, J. (1960). Random orderings and stochastic theories of responses. In Olkin, I. et al., editors, *Contributions to Probability and Statistics*, pages 97–132. Stanford University Press.

Debreu, G. (1960). Review of R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis*. *American Economic Review*, 50(1):186–188.

Falmagne, J.-C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18(1):52–72.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6):1291–1326.

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

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

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

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

References II

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.

Mosteller, F. and Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, 59(5):371–404.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1):31–48.