scottish institute for research in economics

SIRE DISCUSSION PAPER

SIRE-DP-2013-28

Stochastic Choice and Consideration Sets

Paola Manzini
Marco Mariotti

University of St. Andrews

www.sire.ac.uk

Stochastic Choice and Consideration Sets*

Paola Manzini Marco Mariotti†

This version: March 2013

Abstract

We model a boundedly rational agent who suffers from limited attention. The agent considers each feasible alternative with a given (unobservable) probability, the *attention parameter*, and then chooses the alternative that maximises a preference relation within the set of considered alternatives. We show that this random choice rule is the only one for which the impact of removing an alternative on the choice probability of any other alternative is asymmetric and menu independent. Both the preference relation and the attention parameters are identified uniquely by stochastic choice data.

J.E.L. codes: D0.

Keywords: Discrete choice, Random utility, Logit model, Luce model, Consideration sets, bounded rationality, revealed preferences.

*We are grateful to three exceptionally thorough referees and the Co-Editor Wolfgang Pesendorfer for many insightful suggestions and for correcting errors. We also thank the seminar audiences at the UCL "Bounded Rationality, Revealed Preferences and Consumer Demand Workshop" and the Econometric Society NASM 2012, as well as to Chris Flynn, Jacob Goeree, Sean Horan, Rosa Matzkin, Mauro Papi, Ivan Soraperra, Rani Spiegler and Alex Tetenov for helpful comments on a previous related paper, "A salience theory of choice errors" (IZA Discussion Paper No. 5006, 2010) in which the present paper originates. Financial support from ESRC (grant ES/J012513/1) is gratefully acknowledged.

†Both authors at School of Economics and Finance, University of St. Andrews, Castlecliffe, The Scores, St. Andrews KY16 9AL, Scotland, U.K. (e-mail Manzini: paola.manzini@st-andrews.ac.uk; e-mail Mariotti: marco.mariotti@st-andrews.ac.uk).

1 Introduction

We consider a boundedly rational agent who maximises a preference relation but makes random choice errors due to imperfect attention. We extend the classical revealed preference method to this case of bounded rationality, and show how an observer of choice frequencies can (1) test by means of simple axioms whether the data can have been generated by the model, and (2) if the answer to (1) is in the affirmative, infer uniquely both preferences and attention.

Most models of economic choice assume deterministic behaviour. The choice responses are a function c that indicates the selection $c(A)$ the agent makes from menu A. This holds true both for the classical 'rational' model of preference maximisation (Samuelson [35], Richter [29]) and for more recent models of boundedly rational choice.[1] Yet there is a gap between such theories and real data, which are noisy: individual choice responses typically exhibit variability, in both experimental and market settings (McFadden [27]). The choice responses in our model are given by a probability distribution p that indicates the probability $p(a, A)$ that alternative a is selected from menu A, as in the pioneering work of Luce [20], Block and Marschak's [4] and Marschak's [22], and more recently Gul, Natenzon and Pesendorfer [16] (henceforth, GNP).[2]

Imperfect attention is what generates randomness and choice errors in our model. Attention is a central element in human cognition (e.g. Anderson [3]) and was recognized in economics as early as in the work of Simon [37]. For example, a consumer buying a new PC is *not aware* of all the latest models and specifications and ends up making a selection he later regrets;[3] a doctor short of time for formulating a diagnosis *overlooks* the relevant disease for the given set of symptoms; an ideological voter *deliberately ignores*

[1] The works on deterministic choice mentioned in the paper constitute examples for this assertion.

[2] Stochastic choice has also been used recently as a device in the literature of choice *over* menus. E.g. Koida [19] studies how a decision maker's (probabilistic) mental states drive the choice of an alternative from each menu, in turn determining the agent's preference for commitment in his choice over menus. Ahn and Sarver [2] instead use the Gul and Pesendorfer's [15] random expected utility model in the second period of a menu choice model, and show how preference for flexibility yields a unique identification of subjective state probabilities. In this paper we focus on choice *from* menus.

[3] Goeree [13] quantifies this phenomenon with empirical data.

some candidates independently of their policies.[4] In these examples the decision-maker is able to *evaluate* the alternatives he considers (unlike, for example, a consumer who is uncertain about the quality of a product). Yet, for various reasons the agent misses some relevant options through unawareness, overlooking, or deliberate avoidance. In these examples, an agent does not rationally evaluate all objectively available alternatives in A, but only pays attention to a (possibly strict) subset of them, $C(A)$, which we call the *consideration set* following the extensive marketing literature[5] on brands and some recent economics literature discussed below. Once a $C(A)$ 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).

This two-step conceptualisation of the act of choice is rooted and well-accepted in psychology and marketing science, and it has recently gained prominence in economics through the works of Masatlioglu, Nakajima and Ozbay [23] - (henceforth, MNO) - and Eliaz and Spiegler [10], [11]. The core development in our model with respect to earlier works is that the composition of the consideration set $C(A)$ is stochastic. Each alternative a is considered with a probability $\gamma(a)$, the *attention parameter* relative to alternative a. For example, $\gamma(a)$ may indirectly measure the degree of brand awareness for a product, or the (complement of) the willingness of an agent to seriously evaluate a political candidate. Such partial degrees of awareness or willingness to consider are assumed to be representable by a probability.[6]

We view the amount of attention paid to an alternative as a fixed characteristic of the relationship between agent and alternative. The assumption that the attention parameter is menu independent is undoubtedly a substantive one. It does have, however, empirical

[4] See Wilson [44] for a consideration set approach to political competition. It is reported there that African Americans tend to ignore Republican candidates in spite of the overlap between their policy preferences and the stance of the Republicans, and even if they are dissatisfied with the Democratic candidate.

[5] Originating in Wright and Barbour [45]. See also Shocker, Ben-Akiva, Boccara and Nedungadi [36], Roberts and Lattin ([31],[32]) and Roberts and Nedungadi [33].

[6] In some cases one could also adopt a multi-person interpretation, where $\gamma(a)$ measures the proportion in the population who pays attention to brand or political candidate a.

support.[7] And, at the theoretical level, the hypothesis of independent attention parameters is a natural starting point, as on the one hand we show that unrestricted menu dependence yields a model with no observable restrictions (Theorem 2), and on the other hand it is not clear a priori what partial restrictions should be imposed on stochastically menu dependent parameters.

The work by MNO [23] is especially relevant for this paper as it is the first to study how attention and preferences can be retrieved from choice data in a consideration set model of choice. However, the choice responses in their model are deterministic, and like in many other two-stage deterministic models of choice,[8] it is not possible to pin down the primitives entirely by observing the choice data that it generates, even after imposing some structure on the first-stage selection.[9] An attractive feature of our model is that it affords a unique identification of the primitives (preferences and attention parameters) by means of stochastic choice data. The key observation for preference revelation in our model generalizes a feature of classical revealed preference analysis. If an alternative a is preferred to an alternative b, the probability with which a is chosen (in the deterministic case, *whether* a is chosen) from a menu cannot depend on the presence of b, whereas the probability with which b is chosen is affected by the presence of a. As explained in section 3) it is also easy to identify the attention parameters in several choice domains.

The main formal result of the paper is a characterization of the model by means of two axioms (Theorem 1) that make simple assumptions on the effect of the removal of an alternative b from a menu A on the choice probability of another alternative a, measured by the ratio $\frac{p(a,A\setminus\{b\})}{p(a,A)}$, which we call in short the *impact* of b on a. Our random choice rule is the only one for which the impact of b on a is asymmetric (if the presence of b affects the probability of choosing a from A, then the reverse cannot hold) and menu-independent (it does not depend on which alternatives are in A beside a and b).

Our model can be viewed, as we detail in section 7.1, as a special type of Random

[7] For example van Nierop *et al* [28] estimate an unrestricted probabilistic model of consideration set membership for product brands, and find that the covariance matrix of the stochastic disturbances to the consideration set membership function can be taken to be diagonal.

[8] E.g. our own "shortlisting" method [21].

[9] See example in section 6. Tyson [42] clarifies the general structure of two-stage models of choice.

Utility Maximisation, and rationalises some plausible types of choice mistakes that cannot be captured by the Luce [20] rule (the leading type of restriction of Random Utility Maximisation), in which

$$p(a, A) = u(a) / \sum_{b \in A} u(b),$$

for some strictly positive utility function u. The Luce rule is equivalent to the multinomial logit model (McFadden [26]) popular in econometric studies, which assumes the maximisation of a random utility with additive and Gumbel-distributed errors. This is a very specific error model and it is plausible to conjecture that an agent may make different types of mistakes. The Luce rule is incompatible, for example, with *choice frequency reversals* of the form $p(a, \{a, b, c\}) > p(b, \{a, b, c\})$ and $p(b, \{a, b\}) > p(a, \{a, b\})$, which can instead be accommodated by our model in spite of the asymmetry of the preference relation (see Example 1). Choice frequency reversals of various nature have been observed experimentally and they are natural when attention influences choice: for example, a superior but unbranded cereal a may be chosen less frequently than a mediocre but branded cereal b, simply because a is not noticed. But if a third intermediate cereal c becomes available, then b will be chosen less often (it will not be chosen whenever c is noticed), while a will be chosen with the same frequency as before, so that a reversal may occur. Similarly, in spite of the transitivity of the underlying preference, the random consideration set model is compatible with widely observed forms of *stochastic intransitivity* that are instead excluded by Luce (section 4.2).

In section 4 we dwell on the well-known blue bus/red bus type of menu effect (Debreu [8]), in which an agent's odds of choosing a bus over a train increase following the removal of another bus that differs only in colour. Our model, like some extensions of the Luce rule (such as the recent one by GNP [16]), provides a story to explain this effect. But it also shows that a different type of menu effect, not considered by Debreu and the subsequent commentators, and not captured by a Luce-type model, can be plausible in the example: the removal of 'duplicate alternatives' (e.g. a blue bus when a red bus is available) may well *decrease* the odds of choosing the remaining duplicate alternatives (by reducing the attention paid to them) over a third alternative instead of increasing it.

At the axiomatic level, such a difference in behavioural implications between our model

and Luce-type models well illustrates the crucial importance of the exact specification of what is meant by 'menu independence'. The Luce rule requires menu independence of the ratios $\frac{p(a,A)}{p(b,A)}$. Applying instead menu independence to the impacts $\frac{p(a,A\setminus\{b\})}{p(a,A)}$ seems equally natural, and yields (together with asymmetry) an entirely different model. If one thinks that preference should be menu independent, then the a priori appeal of one or the other axiom hinges on a hypothesis about what pattern reveals preference in the data. And, in turn, this rests on a hypothesis on the cognitive process underlying choice.

2 Random choice rules

There is a nonempty finite set of alternatives X, and a *domain* $\mathcal{D}$ of subsets (the *menus*) of X. We allow the agent to not 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).[10] Denote $X^* = X \cup \{a^*\}$ and $A^* = A \cup \{a^*\}$ for all $A \in \mathcal{D}$.

Definition 1 *A random choice rule is a map* $p : X^* \times \mathcal{D} \to [0,1]$ *such that* $\sum_{a \in A^*} p(a,A) = 1$ *for all* $A \in \mathcal{D}$ *and* $p(a,A) = 0$ *for all* $a \notin A^*$.

The interpretation is that $p(a,A)$ denotes the probability that the alternative $a \in A^*$ 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 $p(a^*, \varnothing) = 1$.

We define a new type of random choice rule by assuming that the agent has a strict preference ordering $\succ$ on A. The preference $\succ$ is applied only to a *consideration set* $C(A) \subseteq A$ of alternatives (the set of alternatives the decision maker pays attention to). We allow for $C(A)$ to be empty, in which case the agent picks the default option a^*, so that $p(a^*, A)$ is the probability that $C(A)$ is empty. The membership of $C(A)$ for the alternatives in A is probabilistic. For all $A \in \mathcal{D}$, each alternative a has a probability $\gamma(a) \in (0,1)$ of being in $C(A)$. Formally:

[10] For a recent work on allowing 'not choosing' in the deterministic case, see Gerasimou [12]. Earlier work is Clark [7].

Definition 2 *A random consideration set rule is a random choice rule $p_{\succ,\gamma}$ for which there exists a pair $(\succ, \gamma)$, where $\succ$ is a strict total order on X and γ is a map $\gamma : X \to (0,1)$, such that*[11]

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

3 Characterisation

3.1 Revealed preference and revealed attention

Suppose the choice data are generated by a random consideration set rule. Can we infer the preference ordering from the choice data? One way to extend the revealed preference ordering of rational deterministic choice to stochastic choices is to declare $a \succ b$ iff $p(a, A) > p(b, A)$ for some menu A (see GNP [16]). However, depending on the underlying choice procedure, a higher choice frequency for a might not be due to a genuine preference for a over b, and indeed this is *not* the way preferences are revealed in the random consideration set model. The discrepancy is due to the fact that an alternative may be chosen more frequently than another in virtue of the attention paid to it as well as of its ranking. We consider a different natural extension of the deterministic revealed preference that accounts for this feature while retaining the same flavour as the standard non stochastic environment.

In the deterministic case the preference for a over b has (among others) the observable feature that b can turn from rejected to chosen when a is removed. This feature reveals unambiguously that a is preferred to b, and has an analog in our random consideration set framework. When a is ranked below b, there is no event in which the presence of a in the consideration set matters for the choice of b; therefore if removing a increases the choice probability of b, it means that a must be better ranked than b. And conversely if $a \succ b$ then excising a from A removes the event in which a is considered (in which case b is not chosen), so that the probability of choosing b increases. Thus $p(b, A) > p(b, A \setminus \{a\})$ defines the revealed preference relation $a \succ b$ of our model. We will show that this relation

[11] We use the convention that the product over the empty set is equal to one.

is revealed uniquely.[12]

Next, given a preference $\succ$, the attention paid to an alternative a is revealed directly by the probability of choice in any menu in which a is the best alternative. For example in Theorem 1 we admit all singleton menus, so that $\gamma(a) = p(a, \{a\}) = 1 - p(a^*, \{a\})$. However $\gamma(a)$ may be identified even when the choice probabilities from some menus (singletons in particular) cannot be observed. Provided that there are at least three alternatives and that binary menus are included in the domain, identification occurs via the formula

$$\gamma(a) = 1 - \sqrt{\frac{p(a^*, \{a, b\})\, p(a^*, \{a, c\})}{p(a^*, \{b, c\})}}$$

which must hold since under the model $p(a^*, \{b, c\}) = (1 - \gamma(b))(1 - \gamma(c))$ and therefore

$$\begin{aligned}
(1 - \gamma(a))^2 p(a^*, \{b, c\}) &= (1 - \gamma(a))^2 [(1 - \gamma(b))(1 - \gamma(c))] \\
&= [(1 - \gamma(a))(1 - \gamma(b))][(1 - \gamma(a))(1 - \gamma(c))] \\
&= p(a^*, \{a, b\})\, p(a^*, \{a, c\}).
\end{aligned}$$

This identification strategy can be further generalised using any disjoint menus B and C instead of the alternatives b and c in the formula.[13]

These considerations suggest that the restrictions on observable choice data that characterize the model are those ensuring that, firstly, the revealed preference relation $\succ$ indicated above is well-behaved, i.e. it is a strict total order on the alternatives; and, secondly, that the observed choice probabilities are consistent with this $\succ$ being maximised on the consideration sets that are stochastically generated by the revealed attention parameters.

3.2 Axioms and characterisation theorem

Let us denote

$$aAb = \frac{p(a, A \setminus \{b\})}{p(a, A)}$$

the *impact* that an alternative $b \in A \in \mathcal{D}$ has, in menu A, on another alternative $a \in A^*$ with $a \neq b$. If $aAb > 1$ we say that b *boosts* a and if $aAb = 1$ that b is *neutral* for a.

[12] It is easy to see that $p(a, A) = p(a, A \setminus \{b\})$ also reveals the preference for a over b in our model (again in analogy to rational deterministic choice).

[13] We thank two referees for suggesting these points.

Our axioms constrain the impacts. The axioms are intended for all $A, B \in \mathcal{D}$ and for all $a, b \in A \cap B$, $a \neq b$.

i-Asymmetry. $aAb \neq 1 \Rightarrow bAa = 1$.

i-Independence. $aAb = aBb$ and $a^*Ab = a^*Bb$.

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. It is a simple form of menu-independence, alternative to Luce's IIA (Luce [20]):

Luce's IIA. $\frac{p(a,A)}{p(b,A)} = \frac{p(a,B)}{p(b,B)}$.

i-Independence is structurally similar to Luce's IIA except that it relates to the impacts aAb instead of the odd ratios $\frac{p(a,A)}{p(b,A)}$. We discuss further in the next section the relationship between these two properties, which appear a priori equally plausible ways to capture a notion of menu independence.

A first interesting implication of the axioms (valid on any domain including all pairs and their subsets) is instructive on how they act and will be used in the proof of the main result:

i-Regularity. $aAb \geq 1$ and $a^*Ab \geq 1$.

i-Regularity yields by iteration the standard axiom of Regularity (or Monotonicity)[14], and says that if an alternative is not neutral for another alternative then it must boost it. While it is often assumed directly, this is not a completely innocuous property: it excludes for example the phenomenon of 'asymmetric dominance', whereby *adding* an alternative that is clearly dominated by a but not by b increases the probability that a is chosen.

[14] Regularity: $A \subset B \Rightarrow p(a, A) \geq p(a, B)$.

Lemma 1 *Let $A \in \mathcal{D}$ for all $A \subseteq \{a, b\}$, for all distinct $a, b \in X$. Let p be a random choice rule such that $p(a, A) \in (0, 1)$ for all $a \in A^*$, for all $A \in \mathcal{D} \backslash \varnothing$. If p satisfies i-Asymmetry and i-Independence then p also satisfies i-Regularity.*

Proof. Let p satisfy the assumptions in the statement. By i-Independence it is sufficient to show that $a\{a, b\}b \geq 1$ and $a^*\{a\}a \geq 1$ for all $a, b \in X$. The latter inequality is immediately seen to be satisfied since, by the definition of a random choice rule and of a^*,

$$\frac{p(a^*, \varnothing)}{p(a^*, \{a\})} = \frac{1}{1 - p(a, \{a\})} > 1$$

in view of $p(a, \{a\}) \in (0, 1)$. Next, suppose by contradiction that there exist $a, b \in X$ such that $a\{a, b\}b < 1$. By i-Independence we have $a^*\{a, b\}b = a^*\{b\}b$, that is

$$\begin{aligned}
\frac{p(a^*, \{a\})}{p(a^*, \{a, b\})} &= \frac{p(a^*, \varnothing)}{p(a^*, \{b\})} \Leftrightarrow \\
p(a^*, \{a\})\, p(a^*, \{b\}) &= p(a^*, \{a, b\}) \Leftrightarrow \qquad (1) \\
(1 - p(a, \{a\}))(1 - p(b, \{b\})) &= 1 - p(a, \{a, b\}) - p(b, \{a, b\}) \Leftrightarrow \\
p(a, \{a\}) + p(b, \{b\}) - p(a, \{a\})\, p(b, \{b\}) &= p(a, \{a, b\}) + p(b, \{a, b\}) \qquad (2)
\end{aligned}$$

Moreover the assumption $a\{a, b\}b < 1$ implies by i-Asymmetry that $b\{a, b\}a = 1$, that is $p(b, \{b\}) = p(b, \{a, b\})$. Therefore formula 2 simplifies to

$$p(a, \{a\})(1 - p(b, \{b\})) = p(a, \{a, b\})$$

implying (since $(1 - p(b, \{b\})) < 1$) $p(a, \{a\}) > p(a, \{a, b\})$, which contradicts the assumption $a\{a, b\}b < 1$. ■

A useful additional observation is that formula (2) rules out $b\{a, b\}a = 1 = a\{a, b\}b$, for otherwise the contradiction $p(a, \{a\})\, p(b, \{b\}) = 0$ would follow. Therefore, in the presence of i-Independence, i-Asymmetry is in fact equivalent to the stronger version

i-Asymmetry*. $aAb \neq 1 \Leftrightarrow bAa = 1$.

Our main result is:

Theorem 1 *Let $\{a,b,c\} \in \mathcal{D}$ for all distinct $a,b,c \in X$, and let $A \in \mathcal{D}$ whenever $B \in \mathcal{D}$ and $A \subseteq B$. Let p be a random choice rule such that $p(a,A) \in (0,1)$ for all $a \in A^*$, for all $A \in \mathcal{D}\backslash\varnothing$. Then p 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 $p_{\succ',\gamma'} = p$ we have $(\succ',\gamma') = (\succ,\gamma)$.*

All remaining proofs are in the Appendix. However, the logic behind the sufficiency part of the proof is simple. Under the axioms the revealed preference relation described in section 3.1 can be shown to be total, asymmetric, and transitive, so that it is taken as our preference ranking $\succ$. Given our domain, the attention value $\gamma(a)$ can be defined from the probabilities $p(a,\{a\})$. Then the axioms are shown to imply the following property: whenever b boosts a,

$$p(a, A\backslash\{b\}) = \frac{p(a,A)}{1-p(b,\{b\})}.$$

This is a weak property of 'stochastic path independence' that may be of interest in itself: it asserts that the boost of b on a must depend only on the 'strength' of b in singleton choice.[15] Finally, the iterated application of this formula shows that the preference and the attention parameters defined above retrieve in any menu the given choice probabilities via the assumed procedure.

4 Explaining Menu effects and Stochastic Intransitivity

4.1 Menu effects

Our model suggests that a reason why Luce's IIA might not hold is that a third alternative may be in different positions (in the preference ranking) relative to a and b and thus may arguably impact on their choice probabilities in different ways. For a random consideration set rule, Luce's IIA is only satisfied for sets A and B that differ exclusively for alternatives each of which is either better or worse than *both* a and b, but otherwise menu-effects can

[15] A similar stochastic path independence property appears as an axiom in Yildiz [47].

arise. So if $a \succ c \succ b$ and $a, b, c \in A$

$$\begin{aligned}\frac{p_{\succ,\gamma}(a,A)}{p_{\succ,\gamma}(b,A)} &= \frac{\gamma(a)}{\gamma(b)\prod\limits_{d\in A:a\succ d\succ b}(1-\gamma(d))} \\ &> \frac{\gamma(a)}{\gamma(b)\prod\limits_{d\in A\setminus\{c\}:a\succ d\succ b}(1-\gamma(d))} = \frac{p_{\succ,\gamma}(a,A\setminus\{c\})}{p_{\succ,\gamma}(b,A\setminus\{c\})}\end{aligned}$$

violating Luce's IIA. In fact, for certain configurations of the attention parameters, the addition or elimination of other alternatives can even reverse the *ranking* between the choice frequencies of two alternatives a and b:

Example 1 ***(Choice frequency reversal)*** *Let $a \succ c \succ b$ and $\gamma(b) > \frac{\gamma(a)}{1-\gamma(a)} > \gamma(b)(1-\gamma(c))$. Then*

$$p_{\succ,\gamma}(a,\{a,b,c\}) = \gamma(a) > \gamma(b)(1-\gamma(a))(1-\gamma(c)) = p(b,\{a,b,c\})$$

and

$$p_{\succ,\gamma}(a,\{a,b\}) = \gamma(a) < \gamma(b)(1-\gamma(a)) = p_{\succ,\gamma}(b,\{a,b\})$$

The basis for the choice frequency reversal in our model is that while a better alternative a may be chosen with lower probability than an inferior alternative b in pairwise contests due to low attention for a, the presence of an alternative c that is better than b but worse than a will reduce the probability that b is noticed without affecting the probability that a is noticed, and possibly, if c attracts sufficiently high-attention, to the point that the initial choice probability ranking between a and b is reversed.[16]

However, a random consideration set rule does satisfy other forms of menu-independence and consistency that look a priori as natural as Luce's IIA. In addition to i-Independence, it also satisfies, for all $A \in \mathcal{D}$, $a \in A^*$ and $b, c \in A$:

i-Neutrality. $aAc > 1, bAc > 1 \Rightarrow aAc = bAc$.

i-Neutrality states that an alternative has the *same* impact on any alternative in the menu which it boosts. While an interesting property in itself, as it simplifies dramatically

[16] Choice frequency reversals of various nature have been observed experimentally. See e.g. Tsetsos, Usher and Chater [38].

the structure of impacts by forcing them to take on only a single value in addition to 1, this is also a weakening of Luce's IIA. In fact, i-Neutrality also states that $\frac{p(a,A\setminus\{c\})}{p(b,A\setminus\{c\})} = \frac{p(a,A)}{p(b,A)}$ under the boosting restriction in the premise (guaranteeing, in our interpretation, that c is ranked above both a and b), while Luce's IIA asserts the same form of menu independence (and more) unconditionally. Our previous discussion explains why this restriction of Luce's IIA may be sensible.

The dependence of the choice odds on the other available alternatives is often a realistic feature, which applied economist have sought to incorporate, for example, in the multinomial logit model.[17] The *blue bus/red bus example* (Debreu [8]) is the standard illustration, in which menu effects occur because of an extreme 'functional' similarity between two alternatives (a red and a blue bus). Suppose the agent chooses with equal probabilities a train (t), a red bus (r) or a blue bus (b) as a means of transport in every binary set, so that the choice probability ratios in pairwise choices for any two alternatives are equal to 1. Then, on the premise that the agent does not care about the colour of the bus and so is indifferent between the buses, it is argued that adding one of the buses to a pairwise choice set including t will increase the odds of choosing t over either bus, thus violating IIA.[18]

GNP [16] suggest to deal with this form of menu-dependence by proposing that 'duplicate' alternatives (such as a red and a blue bus) should be identified observationally, by means of choice data, and by assuming that duplicate alternatives are (in a specific sense) 'irrelevant' for choice. In the example each bus is an observational duplicate of the other because replacing one with the other does not alter the probability of choosing t in a pairwise contest. The assumption of *duplicate elimination* says in this example that the probability of choosing t should not depend on whether a duplicate bus is added to either choice problem that includes the train.[19]

[17] By adding a nested structure to the choice process (nested logit) or by allowing heteroscedasticity of the choice errors (see e.g. Greene [14] or Agresti [1]). A probit model also allows for menu effects.

[18] To be pedantic, Debreu's original example used as 'duplicate' alternatives two recordings of Beethoven's eighth symphony played by the same orchestra but with two different directors. As preferences for directors can be very strong, we use instead McFadden's [26] version of the example.

[19] The general duplicate elimination assumption is more involved but follows the same philosophy.

Our model (once straightforwardly adapted to account for preference ties), highlights however that a new type of menu effect may be plausibly caused by the elimination of duplicate alternatives. In general, it is immediate to see that two indifferent alternatives in a random consideration set rule are always observational duplicates whenever they are paid equal attention, but their elimination can have very different effects depending on their ranking with respect to the other alternatives. We illustrate this in the blue bus/red bus example. The preference relation is now a weak order $\succsim$. We assume that all alternatives in the consideration set that tie for best are chosen with a given probability, and otherwise the model is unchanged. Let $\gamma(t) = y$ and $\gamma(b) = \gamma(r) = x$. Assume first that

$$t \succ b \sim r$$

In this case r and b are duplicates according to GNP's definition because $p_{\succ,\gamma}(t, \{b, t\}) = p(t, \{r, t\}) = y$. The duplicate elimination assumption holds because $p_{\succ,\gamma}(t, \{b, r, t\}) = y$. Let $\beta \in (0, 1)$ be the probability that the blue bus is chosen when both buses are considered. We have:

$$\frac{p_{\succ,\gamma}(b, \{b, r, t\})}{p_{\succ,\gamma}(t, \{b, r, t\})} = \frac{(1-y)(\beta x^2 + x(1-x))}{y}$$

$$\frac{p_{\succ,\gamma}(b, \{b, t\})}{p_{\succ,\gamma}(t, \{b, t\})} = \frac{(1-y)x}{y}$$

and therefore

$$\frac{\frac{p_{\succ,\gamma}(b,\{b,r,t\})}{p_{\succ,\gamma}(t,\{b,r,t\})}}{\frac{p_{\succ,\gamma}(b,\{b,t\})}{p_{\succ,\gamma}(t,\{b,t\})}} = 1 - (1-\beta)x$$

so that, independently of the attention profile γ and of $\beta \in (0, 1)$,

$$\frac{p_{\succ,\gamma}(b, \{b, r, t\})}{p_{\succ,\gamma}(t, \{b, r, t\})} < \frac{p_{\succ,\gamma}(b, \{b, t\})}{p_{\succ,\gamma}(t, \{b, t\})}$$

That is, the odds that the blue bus is chosen over the train necessarily increase when the red bus is made unavailable, which accords (observationally) with the Debreu story.

Assume instead that

$$b \sim r \succ t$$

In this case too b and r are duplicates because $p_{\succ,\gamma}(t, \{b, t\}) = p_{\succ,\gamma}(t, \{r, t\}) = y(1-x)$. But now the duplicate elimination assumption fails since $p_{\succ,\gamma}(t, \{b, r, t\}) = y(1-x)^2 \neq p_{\succ,\gamma}(t, \{b, t\})$. In addition, we have:

$$\frac{p_{\succ,\gamma}(b,\{b,r,t\})}{p_{\succ,\gamma}(t,\{b,r,t\})} = \frac{\beta x^2 + x(1-x)}{y(1-x)^2}$$

$$\frac{p_{\succ,\gamma}(b,\{b,t\})}{p_{\succ,\gamma}(t,\{b,t\})} = \frac{x}{y(1-x)}$$

and therefore

$$\frac{\frac{p_{\succ,\gamma}(b,\{b,r,t\})}{p_{\succ,\gamma}(t,\{b,r,t\})}}{\frac{p_{\succ,\gamma}(b,\{b,t\})}{p_{\succ,\gamma}(t,\{b,t\})}} = \frac{\beta x + (1-x)}{(1-x)} = 1 + \frac{\beta x}{1-x}$$

so that independently of the attention profile γ and of $\beta \in (0,1)$

$$\frac{p_{\succ,\gamma}(b,\{b,r,t\})}{p_{\succ,\gamma}(t,\{b,r,t\})} > \frac{p_{\succ,\gamma}(b,\{b,t\})}{p_{\succ,\gamma}(t,\{b,t\})}$$

Therefore the odds that the blue bus is chosen over the train in this case necessarily *decrease* when the red bus is made unavailable, for all possible levels of attention paid to buses and train, which is the reverse of the Debreu story.

In conclusion, the blue bus/red bus example may be slightly misleading in one respect. All commentators accept Debreu's conclusion that once a red bus is added to the pair {blue bus, train}, the odds of choosing the train over the blue bus should increase. But this conclusion is not evident in itself: it must depend on some conjecture about the cognitive process that generates the choice data. A Luce-like model that captures this type of process is studied by GNP [16]. The analysis above suggests that menu effects of a different type may occur. A consumer faced with multiple bus options may well be more inclined to choose one of them at the expense of the train option. In short, the random consideration set rule shows that crude choice probabilities are an insufficient guide to uncovering the underlying preferences: once this is recognised, some menu effects cease to appear paradoxical.

4.2 Stochastic Intransitivity

Several psychologists, starting from Tversky [39], have noted how choices may well fail to be transitive. When choice is stochastic there are many ways to define analogues of transitive behaviour in deterministic models. A weak such analogue often observed to be violated in experiments is the following:

Weak stochastic transitivity: For all $a, b, c \in X$, $p(a, \{a, b\}) \geq \frac{1}{2}$, $p(b, \{b, c\}) \geq \frac{1}{2} \Rightarrow p(a, \{a, c\}) \geq \frac{1}{2}$.

It is easy to see that a random consideration set rule can account for violations of Weak stochastic transitivity, and thus of the stronger version[20]

Strong stochastic transitivity: For all $a, b, c \in X$, $p(a, \{a, b\}) \geq \frac{1}{2}$, $p(b, \{b, c\}) \geq \frac{1}{2} \Rightarrow p(a, \{a, c\}) \geq \max\{p(a, \{a, b\}), p(b, \{b, c\})\}$.

Consider the following

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.

The key to obtaining the violation in the example is that the ordering of the attention parameters is exactly opposite to the quality ordering of the alternatives. It is easy to check that if the attention ordering weakly agrees with the quality ordering, choices are stochastically transitive.

Thus, the random consideration set rule reconciles a fundamentally transitive motivation (the deterministic preference $\succ$) with stochastic violations of transitivity in the data. In contrast, the Luce rule must necessarily satisfy Weak stochastic transitivity.

5 Menu-dependent attention parameters

In some circumstances it may be plausible to assume that the attention parameter of an alternative depends on which other alternatives are feasible. For example, a brightly

[20] In their survey on choice anomalies Rieskamp, Busemeyer and Meller [30] write: "Does human choice behavior obey the principle of strong stochastic transitivity? An overwhelming number of studies suggest otherwise" (p. 646).

coloured object will stand out more in a menu whose other elements are all grey than in a menu that only contains brightly coloured objects. In this section we show however that a less restricted version of our model that allows for the menu dependence of attention parameters is too permissive. A *menu dependent random consideration set rule* is a random choice rule $p_{\succ,\delta}$ for which there exists a pair $(\succ, \delta)$, where $\succ$ is a strict total order on X and δ is a map $\delta : X \times \mathcal{D}\backslash\varnothing \to (0,1)$, such that

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

Theorem 2 *For every strict total order $\succ$ on X and for every random choice rule p for which $p(a, A) \in (0,1)$ for all $a \in A^*$, all $A \in \mathcal{D}\backslash\varnothing$, there exists a menu dependent random consideration rule $p_{\succ,\delta}$ such that $p = p_{\succ,\delta}$.*

So, once we allow the attention parameters to be menu dependent, not only does the model fail to place any observable restriction on choice data, but the preference relation is also entirely unidentified. Strong assumptions on the function δ must be made to make the model with menu dependent attention useful, but we find it difficult to determine a priori what assumptions would be appropriate. The available empirical evidence on brands seems to suggest at best weak correlations between the probabilities of memberships of the consideration set, and therefore weak menu effects (van Nierop *et al.* [28]).

6 Related literature

The economics papers that are most related to ours conceptually are MNO [23] and Eliaz and Spiegler ([10], [11]). Exactly as in their models, an agent in our model who chooses from menu A maximises a preference relation on a consideration set $C(A)$. The difference lies in the mechanism with which $C(A)$ is formed (note that in the deterministic case, without any restriction, this model is empirically vacuous, as one can simply declare the observed choice from A to be equal to $C(A)$). While Eliaz and Spiegler focus on market competition and the strategic use of consideration sets, MNO focus on the direct testable implications of the model and on the identification of the parameters. Our work is thus more closely related to that of MNO. When the consideration set formation and

the choice data are deterministic as in MNO, consider a choice function c for which $c(\{a,b\}) = a = c(\{a,b,c\})$, $c(\{b,c\}) = b$, $c(\{a,c\}) = c$. Then (as noted by MNO), we cannot infer whether (i) $a \succ c$ (in which case c is chosen over a in a pairwise contest because a is not paid attention to) or (ii) $c \succ a$ (in which case c is never paid attention to). The random consideration set model shows how richer data can help break this type of indeterminacy. In case (i), the data would show that the choice frequency of a is the same in $\{a,b,c\}$ as in $\{a,b\}$. In case (ii), the data would show that the choice frequency of a would be higher in $\{a,b\}$ than in $\{a,b,c\}$.

We next focus on the relationship with models of stochastic choice.

Tversky's ([40], [41]) classical Elimination by Aspects (EBA) rule p_ε, which satisfies Regularity, is such that there exists a real valued function $U : 2^X \to \Re_+$ such that for all $A \in \mathcal{D}$, $a \in A$:

$$p_\varepsilon(a, A) = \frac{\sum_{B \subseteq X : B \cap A \neq A} U(B)\, p_\varepsilon(a, B \cap A)}{\sum_{B \subseteq X : B \cap A \neq \varnothing} U(B)}$$

There are random consideration set rules that are not EBA rules. Tversky showed that for any three alternatives a, b, c EBA requires that if $p_\varepsilon(a, \{a,b\}) \geq \frac{1}{2}$ and $p_\varepsilon(b, \{b,c\}) \geq \frac{1}{2}$, then $p_\varepsilon(a, \{a,c\}) \geq \min\{p_\varepsilon(a, \{a,b\}), p_\varepsilon(b, \{b,c\})\}$ (Moderate stochastic transitivity). Example 2 shows that this requirement is not always met by a random consideration set rule.

Recently, GNP [16] have shown that, in a domain which is 'rich' in a certain technical sense, the Luce model is equivalent to the following Independence property (which is an ordinal version of Luce's IIA): $p(a, A \cup C) \geq p(b, B \cup C)$ implies $p(a, A \cup D) \geq p(b, B \cup D)$ for all sets A, B, C and D such that $(A \cup B) \cap (C \cup D) = \varnothing$. They also generalise the Luce rule to the Attribute Rule in such a way as to accommodate red bus/blue bus type of violations of Luce's IIA (see section 4). We have seen that a random consideration set rule violates one of the key axioms (duplicate elimination) for an Attribute Rule. And the choice frequency reversal Example 1 violates the Independence property above.

Mattsson and Weibull [25] obtain an elegant foundation for (and generalisation of) the Luce rule. In their model the agent (optimally) pays a cost to get close to implementing any desired outcome (see also Voorneveld [43]). More precisely, the agent has to exert more effort the more distant the desired probability distribution from a given default

distribution. When the agent makes an optimal trade-off between the expected payoff and the cost of decision control, the resulting choice probabilities are a 'distortion' of the logit model, in which the degree of distortion is governed by the default distribution. Our paper shares with this work the broad methodology to focus on a detailed model to explain choice errors. However, it is also very different in that Mattsson and Weibull assume a (sophisticated form of) rational behaviour on the part of the agent. One may then wonder whether 'utility-maximisation errors' might not occur at the stage of making optimal tradeoffs between utility and control costs, raising the need to model those errors. A second major difference stems from the fact that our model uses purely *ordinal* preference information. Similar considerations apply to the recent wave of works on rational inattention, such as Matějka and McKay [24], Cheremukhin, Popova and Tutino [6], and Caplin and Dean [5], in which it is assumed that an agent solves the problem of allocating attention optimally.

Recently, Rubinstein and Salant [34] have proposed a general framework to describe an agent who expresses different preferences under different *frames* of choice. The link with this paper is that the set of such preferences is interpreted as a set of deviations from a true (welfare relevant) preference, so this is a model of 'mistakes'. However, the deviations are not analysed as stochastic events.

Finally, we note that the appeal of a two stage structure with a stochastic first stage extends beyond economics, from psychology to consumer science. In philosophy in particular, it has been taken by some (e.g. James [18], Dennett [9], Heisenberg [17]) as a fundamental feature of human choices, and as a solution of the general problem of free will.

7 Concluding remarks[21]

7.1 Random consideration sets and RUM

A Random Utility Maximization (RUM) rule [4] is defined by a probability distribution π on the possible rankings R of the alternatives and the assumption that the agent picks the

[21] We thank the referees for suggesting most of the insights in this section.

top element of the R extracted according to π. Block and Marschak [4], McFadden [26] and Yellot [46] have shown that the Luce model is a particular case of a RUM rule, in which a systematic utility is subject to additive random shocks that are Gumbel distributed. A random consideration set rule $(\succ, \gamma)$ is a different special type of RUM rule, in which π is restricted as follows:

- $\pi(R) = 0$ for any ranking R for which there are alternatives a, b with $a \succ b$, bRa, aRa^* and bRa^* (that is, if R contradicts $\succ$ on some pair of alternatives, then at least one of these alternatives must be $R-$inferior to a^*);
- for any alternative a, $\pi(\{R : aRa^*\}) = \gamma(a)$ (that is, the probability of the set of all rankings for which a is ranked above a^* coincides with the probability that a is noticed);
- for any two alternative a and b, $\pi(\{R : aRa^* \text{ and } bRa^*\}) = \gamma(a)\gamma(b)$ (that is, the events of any two alternatives being ranked above a^* are independent).

For example, a random consideration set rule with two alternatives (beside the default) such that $\gamma(a) = \frac{1}{2}$, $\gamma(b) = \frac{1}{3}$ and $a \succ b$ could be represented by the following RUM rule:[22] $\pi(aba^*) = \frac{1}{6}$, $\pi(aa^*b) = \frac{1}{3}$,$\pi(ba^*a) = \frac{1}{6}$, $\pi(a^*ab) = \pi(a^*ba) = \frac{1}{6}$, $\pi(baa^*) = 0$.

An appealing interpretation of this type of RUM is that the agent is 'in the mood' for an alternative a with probability $\gamma(a)$ (and otherwise prefers the default alternative), and picks the preferred one among all alternatives for which he is in the mood. While indistinguishable in terms of pure choice data, the RUM interpretation and the consideration set interpretation imply different attitudes of the agent to 'implementation errors': if a is chosen but $b \succ a$ is implemented by mistake (e.g. a dish different from the one ordered is served in a restaurant), the agent will have a positive reaction if he failed to pay attention to b, but he will have a negative reaction if he was not in the mood for b.

7.2 Comparative attention

The model suggests a definition of comparative attention based on observed choice probabilities. Say that $(\succ_1, \gamma_1)$ is *more attentive than* $(\succ_2, \gamma_2)$, denoted $(\succ_1, \gamma_1)\,\alpha\,(\succ_2, \gamma_2)$,

[22] Where a ranking is denoted by listing the alternatives from top to bottom.

iff $p_{\succ_1,\gamma_1}(a^*, A) < p_{\succ_2,\gamma_2}(a^*, A)$ for all $A \in \mathcal{D}$. With the domain of theorem 1, we have that $(\succ_1, \gamma_1)\, \alpha\, (\succ_2, \gamma_2)$ iff $\gamma_1(a) > \gamma_2(a)$ for all $a \in X$ (the 'if' direction follows immediately from the formula $p_{\succ,\gamma}(a^*, A) = \prod_{a \in A}(1 - \gamma(a))$, while the other direction follows from $p_{\succ,\gamma}(a^*, \{a\}) = (1 - \gamma(a))$ applied to each $\{a\} \in \mathcal{D}$). Observe that for two agents with the same preferences, $(\succ, \gamma_1)$ is more attentive than $(\succ, \gamma_2)$ iff agent 1 makes 'better choices' from each menu in the sense of first order stochastic dominance, that is $p_{\succ,\gamma_1}(a \succ b, A) > p_{\succ,\gamma_2}(a \succ b, A)$ for all $b \in A$ with $b \neq \max(A, \succ)$, where $p_{\succ,\gamma}(a \succ b, A)$ denotes the probability of choosing an alternative in A better than b.

On general domains (without the assumption that all singleton menus are included in the domain) the implication $(\succ_1, \gamma_1)\, \alpha\, (\succ_2, \gamma_2) \Rightarrow \gamma_1(a) > \gamma_2(a)$ for all $a \in X$ does not necessarily hold. However, in a one-parameter version of the model in which all alternatives receive the same attention $g \in (0, 1)$, it follows from the formula $p_{\succ,g}(a^*, A) = (1 - g)^{|A|}$ that $(\succ_1, g_1)\, \alpha\, (\succ_2, g_2)$ iff $g_1 > g_2$.

7.3 A model without default

A natural companion of our model that does not postulate a default alternative is one in which whenever the agent misses all alternatives he is given the option to 'reconsider', repeating the process until he notices some alternative. This leads to choice probabilities of the form:

$$p_{\succ,\gamma}(a, A) = \frac{\gamma(a) \prod_{b \in A : b \succ a} (1 - \gamma(b))}{1 - \prod_{b \in A} (1 - \gamma(b))}$$

This model does not have the same identifiability properties as ours. For example, take the case $X = \{a, b\}$, with $p(a, \{a, b\}) = \alpha$ and $p(b, \{a, b\}) = \beta$.[23] These observations (which fully identify the parameters in our model) are compatible with both the following continua of possibilities:

- $a \succ b$ and any γ such that $\frac{\gamma(a)}{1-\gamma(a)} = \frac{\alpha}{\beta}\gamma(b)$;
- $b \succ a$ and any γ such that $\frac{\gamma(b)}{1-\gamma(b)} = \frac{\beta}{\alpha}\gamma(a)$.

[23] Obviously in this model $p(a, \{a\}) = 1$ for all $a \in X$.

Nevertheless, the model is interesting and it would be desirable to have an axiomatic characterisation of it. We leave this as an open question.

References

[1] Agresti, A. (2002) Categorical Data Analysis, John Wiley and Sons, Hoboken, NJ.

[2] Ahn D. and T. Sarver (2013) "Preference for Flexibility and Random Choice", *Econometrica* 81: 341–361.

[3] Anderson, J. R. (2005) *Cognitive Psychology and Its Implications*, Worth, New York.

[4] Block, H.D. and J. Marschak (1960) "Random Orderings and Stochastic Theories of Responses", in Olkin, I., S. G. Gurye, W. Hoeffding, W. G. Madow, and H. B. Mann (eds.) *Contributions to Probability and Statistics*, Stanford University Press, Stanford, CA.

[5] Caplin, A. and M. Dean (2013) "Rational Inattention and State Dependent Stochastic Choice", mimeo, Brown University.

[6] Cheremukhin, A., A. Popova and A. Tutino (2011) "Experimental Evidence of Rational Inattention", Working Paper 1112, Federal Reserve Bank of Dallas.

[7] Clark, S.A. (1995) "Indecisive choice theory", *Mathematical Social Sciences* 30: 155-170.

[8] Debreu, G., (1960) "Review of 'Individual choice behaviour' by R. D. Luce" *American Economic Review* 50: 186-88.

[9] Dennett, D. (1978) "On Giving Libertarians What They Say They Want", chapter 15 in *Brainstorms: Philosophical Essays on Mind and Psychology,* Bradford Books.

[10] Eliaz, K. and R. Spiegler (2011) "Consideration Sets and Competitive Marketing", *Review of Economic Studies* 78: 235-262.

[11] Eliaz, K. and R. Spiegler (2011) "On the strategic use of attention grabbers", *Theoretical Economics* 6: 127–155.

[12] Gerasimou, G. (2010) "Incomplete Preferences and Rational Choice Avoidance", mimeo, University of St Andrews.

[13] Goeree, M. S. (2008) "Limited Information and Advertising in the US Personal Computer Industry", *Econometrica* 76: 1017-1074.

[14] Greene, W. H. (2003) "Econometric Analysis", Prentice Hall, Upper Saddle River, NJ.

[15] Gul, F. and W. Pesendorfer (2006) "Random Expected Utility", Econometrica 74: 121–146.

[16] Gul, F., P. Natenzon and W. Pesendorfer (2010) "Random Choice as Behavioral Optimization", mimeo, Princeton University.

[17] Heisenberg, M. (2009) "Is Free Will an Illusion?", *Nature* (459): 164-165.

[18] James, W. (1956) "The Dilemma of Determinism", in *The Will to Believe and Other Essays in Popular Philosophy* (New York, Dover), p. 145-183.

[19] Koida, N. (2010) "Anticipated Stochastic Choice", mimeo, Iwate Prefectural University.

[20] Luce, R. D. (1959) Individual Choice Behavior; a theoretical analysis. Wiley: New York.

[21] Manzini, P. and M. Mariotti (2007) "Sequentially Rationalizable Choice", *American Economic Review*, 97 (5): 1824-1839.

[22] Marschak, J. (1960) "Binary choice constraints and random utility indicators", Cowles Foundation Paper 155.

[23] Masatlioglu, Y., D. Nakajima and E. Ozbay (2009): "Revealed Attention", forthcoming, *American Economic Review.*

[24] Matějka, F. and A. McKay (2011) "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model", working paper 442 CERGE-EI.

[25] Mattsson, L.-G. and J. W. Weibull (2002) "Probabilistic choice and procedurally bounded rationality", *Games and Economic Behavior*, 41(1): 61-78.

[26] McFadden, D.L. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka (ed.), Frontiers in Econometrics, 105-142, Academic Press: New York.

[27] McFadden, D.L. (2000) "Economic choices", *American Economic Review*, 91 (3): 351-378.

[28] van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel & P.H. Franses (2010) "Retrieving Unobserved Consideration Sets from Household Panel Data", *Journal of Marketing Research* XLVII: 63–74.

[29] Richter, M. K. (1966) "Revealed Preference Theory", *Econometrica*, 34: 635-645.

[30] Rieskamp, J., J.R. Busemeyer and B.A. Mellers (2006) "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice" *Journal of Economic Literature* 44: 631-661.

[31] Roberts, J. and Lattin (1991) "Development and testing of a model of consideration set composition", *Journal of Marketing Research*, 28(4): 429–440.

[32] Roberts, J. H. and J. M. Lattin (1997) "Consideration: Review of research and prospects for future insights" *Journal of Marketing Research*, 34(3): 406–410.

[33] Roberts, J. and P. Nedungadi, P. (1995) "Studying consideration in the consumer decision process", *International Journal of Research in Marketing*, 12: 3–7.

[34] Rubinstein, A. and Y. Salant (2009) "Eliciting Welfare Preferences from Behavioral Datasets", mimeo, University of Tel Aviv Cafes, Tel Aviv university, New York University and Northwestern University.

[35] Samuelson, P. A. (1938) "A Note on the Pure Theory of Consumer's Behaviour", *Economica*, 5 (1): 61-71.

[36] Shocker, A. D., M. Ben-Akiva, B. Boccara and P. Nedungadi (1991) "Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions" *Marketing Letters*, 2(3): 181–197.

[37] Simon, H. A. (1957) Models of man: social and rational: mathematical essays on rational human behavior in a social setting, Wiley.

[38] Tsetsos, K., M. Usher and N. Chater (2010) "Preference Reversal in Multiattribute Choice", Psychological Review 117: 1275–1293.

[39] Tversky, A. (1969) "Intransitivity of Preferences" *Psychological Review*, 76: 31-48.

[40] Tversky, A. (1972) "Choice by Elimination", *Journal of Mathematical Psychology* 9: 341-67.

[41] Tversky, A. (1972) "Elimination by Aspects: A Theory of Choice", *Psychological Review* 79: 281-99.

[42] Tyson, C. J. (2011) "Behavioral Implications of Shortlisting Procedures", mimeo, Queen Mary U. of London.

[43] Voorneveld, M. (2006) "Probabilistic choice in games: properties of Rosenthal's t-solutions", *International Journal of Game Theory* 34: 105-121.

[44] Wilson, C. J. (2008) "Consideration Sets and Political Choices: A Heterogeneous Model of Vote Choice and Sub-national Party Strength", *Political Behavior* 30: 161–183.

[45] Wright, P., and F. Barbour (1977) "Phased Decision Strategies: Sequels to an Initial Screening." In: Studies in Management Sciences, Multiple Criteria Decision Making, ed. Martin K. Starr, and Milan Zeleny, 91-109, Amsterdam: North-Holland.

[46] Yellot, J.I., JR. (1977) "The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution", *Journal of Mathematical Psychology*, 15: 109-144.

[47] Yildiz, K. (2012) "List rationalizable choice", mimeo, New York University.

8 Appendix: Proofs

Proof of Theorem 1

The necessity part of the statement is immediately verified by checking the formula and thus omitted here (see the online appendix). For sufficiency, let p be a random choice rule that satisfies i-Asymmetry and i-Independence. By Lemma 1 p also satisfies i-Regularity, and by the observation after the proof of Lemma 1 it satisfies i-Asymmetry* (below we will highlight where this stronger version of i-Asymmetry is needed). Define a binary relation R on X by aRb iff $aAb > 1$ for some $A \in \mathcal{D}$, $a, b \in A$, i.e. iff $p(b, A \setminus \{a\}) > p(b, A)$. We show that R is total, asymmetric and transitive. For totality, given $a, b \in X$, suppose $aAb \leq 1$ for some $A \in \mathcal{D}$ (by the domain assumption there exists an $A \in \mathcal{D}$ such that aAb is well-defined); then by i-Regularity $aAb = 1$ and by i-Asymmetry* $bAa > 1$. For asymmetry, suppose $aAb > 1$ for some $A \in \mathcal{D}$; then by i-Asymmetry $bAa = 1$ and by i-Independence $bBa = 1$ for all $B \ni a, b$, $B \in \mathcal{D}$.

For transitivity, let $aRbRc$, so that $bAa > 1$ and $cBb > 1$ for some $A, B \in \mathcal{D}$. Therefore by i-Independence

$$\begin{aligned} b\{a,b,c\}a &> 1 \\ c\{a,b,c\}b &> 1 \\ b\{a,b\}a &> 1 \\ c\{b,c\}b &> 1 \end{aligned}$$

(these impacts are well-defined by the domain assumption) and then by i-Asymmetry

$$\begin{aligned} a\{a,b,c\}b &= 1 = b\{a,b,c\}c \\ a\{a,b\}b &= 1 = b\{b,c\}c \end{aligned}$$

Suppose that, contrary to transitivity, $not\,(aRc)$, so that by totality cRa, i.e. there exists $C \in \mathcal{D}$ such that $aCc > 1$ and thus by i-Independence

$$\begin{aligned} a\{a,b,c\}c &> 1 \\ a\{a,c\}c &> 1 \end{aligned}$$

and then by i-Asymmetry

$$c\{a,b,c\}\,a = 1 = c\{a,c\}\,a$$

In short, the above displayed formulas imply the following system:

$$\begin{aligned} p(a,\{a\}) = p(a,\{a,b\}) > p(a,\{a,c\}) = p(a,\{a,b,c\}) \\ p(b,\{b\}) = p(b,\{b,c\}) > p(b,\{a,b\}) = p(b,\{a,b,c\}) \\ p(c,\{c\}) = p(c,\{a,c\}) > p(c,\{b,c\}) = p(c,\{a,b,c\}) \end{aligned} \tag{3}$$

Furthermore, i-Independence also implies:

$$\frac{p(a^*,\varnothing)}{p(a^*,\{a\})} = \frac{p(a^*,\{b\})}{p(a^*,\{a,b\})} = \frac{p(a^*,\{c\})}{p(a^*,\{a,c\})} = \frac{p(a^*,\{b,c\})}{p(a^*,\{a,b,c\})}$$

or

$$\begin{aligned} \frac{1}{1-p(a,\{a\})} &= \frac{1-p(b,\{b\})}{1-p(a,\{a,b\})-p(b,\{a,b\})} \\ \frac{1}{1-p(a,\{a\})} &= \frac{1-p(c,\{c\})}{1-p(a,\{a,c\})-p(c,\{a,c\})} \\ \frac{1}{1-p(a,\{a\})} &= \frac{1-p(b,\{b,c\})-p(c,\{b,c\})}{1-p(a,\{a,b,c\})-p(b,\{a,b,c\})-p(c,\{a,b,c\})} \end{aligned}$$

which using formulas (3) can be re-written as:

$$\frac{1}{1-p(a,\{a,b\})} = \frac{1-p(b,\{b,c\})}{1-p(a,\{a,b\})-p(b,\{a,b\})} \tag{4}$$

$$\frac{1}{1-p(a,\{a,b\})} = \frac{1-p(c,\{a,c\})}{1-p(a,\{a,c\})-p(c,\{a,c\})} \tag{5}$$

$$\frac{1}{1-p(a,\{a,b\})} = \frac{1-p(b,\{b,c\})-p(c,\{b,c\})}{1-p(a,\{a,c\})-p(b,\{a,b\})-p(c,\{b,c\})} \tag{6}$$

But the solution to these equations requires $p(c,\{a,c\}) = p(c,\{b,c\})$ (see the online appendix for the workout), contradicting the third line in (3), and we can conclude that R is transitive.

Finally, concerning R, observe that (using i-Asymmetry* and i-Independence) the following three statements are equivalent:

$$\begin{aligned} & aRb \\ p(b,A\backslash\{a\}) &> p(b,A) \text{ for all } A\in\mathcal{D} \text{ with } a,b\in A \\ p(a,A\backslash\{b\}) &= p(a,A) \text{ for all } A\in\mathcal{D} \text{ with } a,b\in A \end{aligned} \tag{7}$$

Next, we show that for all $A \in \mathcal{D}$, the following implication holds:

$$p(a, A\setminus\{b\}) > p(a, A) \Rightarrow \frac{p(a, A\setminus\{b\})}{p(a, A)} = \frac{1}{1 - p(b, \{b\})} \tag{8}$$

for all $a \in A^*$ and $b \in A$. We begin by proving (8) for $a = a^*$. Suppose first that $A = \{b\}$ for some $b \in X$. Since $p(a^*, \varnothing) = 1$ and $p(a^*, \{b\}) = 1 - p(b, \{b\})$, we have

$$\frac{p(a^*, \varnothing)}{p(a^*, \{b\})} = \frac{1}{1 - p(b, \{b\})} \tag{9}$$

so that the assertion holds for this case. Then applying i-Independence to (9) we have immediately

$$\frac{p(a^*, A\setminus\{b\})}{p(a^*, \{A\})} = \frac{1}{1 - p(b, \{b\})}$$

for all $A \in \mathcal{D}$, for all $b \in A$. Next, fix $a, b \in A$ and assume $p(a, A\setminus\{b\}) > p(a, A)$, so that by i-Asymmetry $p(b, A\setminus\{a\}) = p(b, A)$. Using this equation and i-Independence yields

$$\begin{aligned}\frac{p(a, A\setminus\{b\})}{p(a, A)} &= \frac{p(a, \{a\})}{p(a, \{a, b\})} = \frac{1 - p(a^*, \{a\})}{1 - p(b, \{a, b\}) - p(a^*, \{a, b\})} = \\ &\frac{1 - p(a^*, \{a\})}{1 - p(b, \{b\}) - p(a^*, \{a, b\})}\end{aligned}$$

and since as shown before

$$p(a^*, \{a\}) = \frac{p(a^*, \{a, b\})}{1 - p(b, \{b\})}$$

we have

$$\frac{p(a, A\setminus\{b\})}{p(a, A)} = \frac{1 - \frac{p(a^*, \{a, b\})}{1 - p(b, \{b\})}}{1 - p(b, \{b\}) - p(a^*, \{a, b\})} = \frac{1}{1 - p(b, \{b\})}$$

This concludes the proof that formula (8) holds.

Now define $\succ = R$ and $\gamma(a) = p(a, \{a\})$ for all $a \in X$. We show that $p_{\succ,\gamma} = p$. Fix $A \in \mathcal{D}$ and number the alternatives so that $A = \{a_1, ..., a_n\}$ and $a_i \succ a_j \Leftrightarrow i < j$. For all $a \in A$ the implication in (8) and the definitions of γ and $\succ$ imply that

$$\begin{aligned}p(a_i, A) &= p(a_i, \{a_2, ..., a_n\})(1 - \gamma(a_1)) \\ &\vdots \\ &= p(a_i, \{a_i, ..., a_n\}) \prod_{j<i} (1 - \gamma(a_j)) \\ &= p(a_i, \{a_i\}) \prod_{j<i} (1 - \gamma(a_j)) \\ &= \gamma(a_i) \prod_{j<i} (1 - \gamma(a_j)) = p_{\succ,\gamma}(a_i, A)\end{aligned}$$

where $p(a_i, \{a_i\}) = p(a_i, \{a_i, ..., a_n\})$, which is used to move from the second to the third line, follows from the properties of R in (7) (note that the probabilities in the display are all well-defined by the domain assumption).

To conclude we show that $\succ$ and γ are defined uniquely. Let $p_{\succ',\gamma'}$ be another consideration set rule for which $p_{\succ',\gamma'} = p$, and suppose by contradiction that $\succ' \neq \succ$. So there exist $a, b \in X$ such that $a \succ b$ and $b \succ' a$. Take $A = \{a\} \cup \{c \in X : a \succ c\}$, so that $b \in A$ for some b with $b \succ' a$. By definition,

$$p_{\succ,\gamma}(a, A) = \gamma(a) = p_{\succ,\gamma}(a, B)$$

for all $B \subset A$ such that $a \in B$, but also

$$p_{\succ',\gamma'}(a, A) = \gamma'(a) \prod_{c \in A: c \succ' a} (1 - \gamma'(c)) < \gamma'(a) \prod_{c \in A \setminus \{b\}: c \succ' a} (1 - \gamma'(c)) = p_{\succ',\gamma'}(a, A \setminus \{b\})$$

a contradiction in view of $p_{\succ',\gamma'} = p = p_{\succ,\gamma}$. So $\succ$ is unique. The uniqueness of γ is immediate from $p(a, \{a\}) = \gamma(a)$. ■

Proof of Theorem 2. Let p be a random choice rule. Let $\succ$ be an arbitrary strict total order of the alternatives. Define δ by setting, for $A \in \mathcal{D} \setminus \varnothing$ and $a \in A$:

$$\delta(a, A) = \frac{p(a, A)}{1 - \sum_{b \in A: b \succ a} p(b, A)} \tag{10}$$

We have $\delta(a, A) > 0$ since $p(a, A) > 0$, and we have $\delta(a, A) < 1$ since $1 > p(a, A) + \sum_{b \in A: b \succ a} p(b, A)$ (given that $p(a^*, A) > 0$).

For the rest of the proof fix $a \in A$. We define

$$p_{\succ,\delta}(a, A) = \delta(a, A) \prod_{b \in A: b \succ a} (1 - \delta(b, A))$$

and show that $p_{\succ,\delta}(a, A) = p(a, A)$. Using the definition of δ, for all $b \in A$ we have

$$1 - \delta(b, A) = \frac{1 - \sum_{c \in A: c \succ b} p(c, A) - p(b, A)}{1 - \sum_{c \in A: c \succ b} p(c, A)} \tag{11}$$

so that

$$\prod_{b \in A: b \succ a} (1 - \delta(b, A)) = \prod_{b \in A: b \succ a} \frac{1 - \sum_{c \in A: c \succ b} p(c, A) - p(b, A)}{1 - \sum_{c \in A: c \succ b} p(c, A)} \tag{12}$$

Given any $b \in A$, denote by $b^+ \in A$ the unique alternative for which $b^+ \succ b$ and there is no $c \in A$ such that $b^+ \succ c \succ b$. Letting $b \in \{c \in A : c \succ a\}$, from (11) we have that

$$1 - \delta\left(b^+, A\right) = \frac{1 - \sum\limits_{c \in A: c \succ b^+} p(c, A) - p(b^+, A)}{1 - \sum\limits_{c \in A: c \succ b^+} p(c, A)} = \frac{1 - \sum\limits_{c \in A: c \succ b} p(c, A)}{1 - \sum\limits_{c \in A: c \succ b^+} p(c, A)}$$

As the numerator of the expression for $1 - \delta(b^+, A)$ is equal to the denominator of the expression for $1 - \delta(b, A)$, the product in (12) is a telescoping product (where observe that for the $\succ$ −maximal term in A the denominator is equal to 1), and we thus have:

$$\begin{aligned} \prod_{b \in A: b \succ a} (1 - \delta(b, A)) &= 1 - \sum_{b \in A: b \succ a^+} p(b, A) - p\left(a^+, A\right) \\ &= 1 - \sum_{b \in A: b \succ a} p(b, A) \end{aligned}$$

We conclude that

$$\begin{aligned} p_{\succ,\delta}(a, A) &= \delta(a, A) \prod_{b \in A: b \succ a} (1 - \delta(b, A)) \\ &= \frac{p(a, A)}{1 - \sum\limits_{b \in A: b \succ a} p(b, A)} \left(1 - \sum_{b \in A: b \succ a} p(b, A)\right) \\ &= p(a, A) \end{aligned}$$

as desired (where the first term in the second line follows from (10)). ■

Possible online appendices

A Necessity of the axioms

i-Asymmetry. $aAb \neq 1 \Rightarrow bAa = 1$.

Let $p_{\succ,\gamma}$ be a random consideration set rule. For all $A \in \mathcal{D}$, all $a, b \in A$ with $b \neq a$:

$$\frac{p_{\succ,\gamma}(a, A \setminus \{b\})}{p_{\succ,\gamma}(a, A)} = \frac{\gamma(a) \prod\limits_{c \in A \setminus \{b\}: c \succ a} (1 - \gamma(c))}{\gamma(a) \prod\limits_{c \in A: c \succ a} (1 - \gamma(c))} = \begin{cases} \frac{1}{1-\gamma(b)} & \text{if} \quad b \succ a \\ 1 & \text{if} \quad a \succ b \end{cases}$$

so that $aAb = 1 \Leftrightarrow a \succ b$. Therefore if $aAb \neq 1$ we have $b \succ a$ and thus $bAa = 1$.

i-Independence. $aAb = aBb$ and $a^*Ab = a^*Bb$.

Let $p_{\succ,\gamma}$ be a random consideration set rule. Then for all $A, B \in \mathcal{D}$ and for all $a, b \in A \cap B$, $a \neq b$:

$$\begin{aligned}\frac{p_{\succ,\gamma}(a, A\setminus\{b\})}{p_{\succ,\gamma}(a, A)} &= \frac{\gamma(a)\prod\limits_{c\in A\setminus\{b\}:c\succ a}(1-\gamma(c))}{\gamma(a)\prod\limits_{c\in A:c\succ a}(1-\gamma(c))} \\ &= \begin{cases} \frac{1}{1-\gamma(b)} & \text{if} \quad b\succ a \\ 1 & \text{if} \quad a\succ b \end{cases} = \frac{\gamma(a)\prod\limits_{c\in B\setminus\{b\}:c\succ a}(1-\gamma(c))}{\gamma(a)\prod\limits_{c\in B:c\succ a}(1-\gamma(c))} = \frac{p_{\succ,\gamma}(a, B\setminus\{b\})}{p_{\succ,\gamma}(a, B)}\end{aligned}$$

as desired. Similarly, we have

$$\frac{p_{\succ,\gamma}(a^*, A\setminus\{b\})}{p_{\succ,\gamma}(a^*, A)} = \frac{\prod\limits_{c\in A\setminus\{b\}}(1-\gamma(c))}{\prod\limits_{c\in A}(1-\gamma(c))} = \frac{1}{1-\gamma(b)} = \frac{\prod\limits_{c\in B\setminus\{b\}}(1-\gamma(c))}{\prod\limits_{c\in B}(1-\gamma(c))} = \frac{p_{\succ,\gamma}(a, B\setminus\{b\})}{p_{\succ,\gamma}(a, B)}$$

In the text we reported that the random consideration set rule satisfies i-Neutrality - below we show that i-neutrality is indeed necessary:

i-Neutrality. $aAc > 1, bAc > 1 \Rightarrow aAc = bAc$ for all $A \in \mathcal{D}$, $a \in A^*$ and $b, c \in A$.

Let $p_{\succ,\gamma}$ be a random consideration set rule. For all $A \in \mathcal{D}$, $a \in A^*$ and $b, c \in A$ with $c \neq a$:

$$aAc = \frac{p_{\succ,\gamma}(a, A\setminus\{c\})}{p_{\succ,\gamma}(a, A)} = \frac{\gamma(a)\prod\limits_{d\in A\setminus\{c\}:d\succ a}(1-\gamma(d))}{\gamma(a)\prod\limits_{d\in A:d\succ a}(1-\gamma(d))} = \begin{cases} \frac{1}{1-\gamma(c)} & \text{if} \quad c\succ a \\ 1 & \text{if} \quad a\succ c \end{cases}$$

so that $aAc > 1 \Leftrightarrow aAc = \frac{1}{1-\gamma(c)}$ for all $a \in A$ such that $a \neq a^*$, while if $a = a^*$, then

$$a^*Ab = \frac{p_{\succ,\gamma}(a^*, A\setminus\{c\})}{p_{\succ,\gamma}(a^*, A)} = \frac{\prod\limits_{d\in A\setminus\{c\}:d\succ a}(1-\gamma(d))}{\prod\limits_{d\in A:d\succ a}(1-\gamma(d))} = \frac{1}{1-\gamma(c)}$$

Therefore $aAc > 1$ and $bAc > 1$ imply $aAc = \frac{1}{1-\gamma(c)} = aAb$.

B Transitivity of R in the proof of Theorem 1

Recall the conditions from the proof:

$$\begin{aligned} p(a,\{a\}) = p(a,\{a,b\}) > p(a,\{a,c\}) = p(a,\{a,b,c\}) \\ p(b,\{b\}) = p(b,\{b,c\}) > p(b,\{a,b\}) = p(b,\{a,b,c\}) \\ p(c,\{c\}) = p(c,\{a,c\}) > p(c,\{b,c\}) = p(c,\{a,b,c\}) \end{aligned} \tag{3}$$

and

$$\frac{1}{1-p(a,\{a,b\})} = \frac{1-p(b,\{b,c\})}{1-p(a,\{a,b\})-p(b,\{a,b\})} \tag{4}$$

$$\frac{1}{1-p(a,\{a,b\})} = \frac{1-p(c,\{a,c\})}{1-p(a,\{a,c\})-p(c,\{a,c\})} \tag{5}$$

$$\frac{1}{1-p(a,\{a,b\})} = \frac{1-p(b,\{b,c\})-p(c,\{b,c\})}{1-p(a,\{a,c\})-p(b,\{a,b\})-p(c,\{b,c\})} \tag{6}$$

To see that the system above yields $p(c,\{b,c\}) = p(c,\{a,c\})$, we simplify by using a product notation for sets and dropping the curly brackets. Given our restrictions on probabilities, (4) yields

$$\begin{aligned} 1-p(a,ab)-p(b,ab) &= 1-p(b,bc)-p(a,ab)+p(a,ab)\,p(b,bc) \Leftrightarrow \\ p(b,ab) &= p(b,bc)-p(a,ab)\,p(b,bc) \end{aligned} \tag{13}$$

while (5) yields

$$\begin{aligned} 1-p(a,ac)-p(c,ac) &= 1-p(c,ac)-p(a,ab)+p(a,ab)\,p(c,ac) \Leftrightarrow \\ p(a,ac) &= p(a,ab)-p(a,ab)\,p(c,ac) \end{aligned} \tag{14}$$

and (6) yields:

$$\begin{aligned} & 1-p(a,ac)-p(b,ab)-p(c,bc) = \\ & = 1-p(b,bc)-p(c,bc)-p(a,ab)+p(a,ab)\,p(b,bc)+p(a,ab)\,p(c,bc) \\ & \Leftrightarrow p(a,ac) = p(b,bc)+p(a,ab)-p(b,ab)-p(a,ab)\,p(b,bc)-p(a,ab)\,p(c,bc) \end{aligned} \tag{15}$$

Substituting the expression for $p(a,ac)$ from (15) into (14) yields

$$\begin{aligned} & p(a,ab)-p(a,ab)\,p(c,ac) = \\ = \;& p(b,bc)+p(a,ab)-p(b,ab)-p(a,ab)\,p(b,bc)-p(a,ab)\,p(c,bc) \\ \Leftrightarrow \;& p(b,ab) = p(a,ab)\,p(c,ac)+p(b,bc)-p(a,ab)\,p(b,bc)-p(a,ab)\,p(c,bc) \end{aligned} \tag{16}$$

and substituting the expression for $p(b, ab)$ from (13) in (16) yields

$$\begin{aligned}
& p(b, bc) - p(a, ab)\, p(b, bc) \\
= \; & p(a, ab)\, p(c, ac) + p(b, bc) - p(a, ab)\, p(b, bc) - p(a, ab)\, p(c, bc) \\
\Leftrightarrow \; & p(a, ab)\, p(c, bc) = p(a, ab)\, p(c, ac) \\
\Leftrightarrow \; & p(c, bc) = p(c, ac)
\end{aligned}$$

C Independence of the axioms

First we recall the axioms - next we provide examples of random choice rules that fail to satisfy only one of the axioms, and show that they are not random consideration set rules.

i-Asymmetry. $aAb \neq 1 \Rightarrow bAa = 1$.

i-Independence. $aAb = aBb$ and $a^*Ab = a^*Bb$.

Fails only i-Asymmetry. Let $X = \{a, b\}$, and assume choice probabilities are as in the table below:

		$p(a, S_i)$	$p(b, S_i)$	$p(a^*, S_i)$
S_1	$\{a, b\}$	x	y	$1 - x - y$
S_2	$\{a\}$	$\frac{x+(1-\alpha)y}{1-\alpha y}$	-	$1 - \frac{x+(1-\alpha)y}{1-\alpha y} = \frac{1-x-y}{1-\alpha y}$
S_3	$\{b\}$	-	αy	$1 - \alpha y$
	$\varnothing$	-	-	1

Table 1: A random choice rule that fails i-Asymmetry

where $x, y > 0$, $x + y < 1$ and $\alpha \in \left(\frac{1}{1-x}, \frac{x+y}{y}\right)$. We have that $\frac{x+y}{y} < \frac{1}{y}$, ensuring that $p(a^*, \{b\}) = 1 - \alpha y > 0$, while the upper bound on α also ensures that $p(a^*, \{a\}) < 1$. This random choice rule fails i-Asymmetry, since $\frac{p(a, S_1 \setminus \{b\})}{p(a, S_1)} = \frac{x+(1-\alpha)y}{x(1-\alpha y)} < 1$ (since $\alpha > \frac{1}{1-x}$) and $\frac{p(b, S_1 \setminus \{a\})}{p(b, S_1)} = \alpha > 1$. i-Independence holds trivially for a and b. For the default alternative:

$$\frac{p(a^*, S_1 \setminus \{a\})}{p(a^*, S_1)} = \frac{p(a^*, \{b\})}{p(a^*, \{a, b\})} = \frac{1-\alpha y}{1-x-y} = \frac{p(a^*, \varnothing)}{p(a^*, \{a\})} = \frac{p(a^*, S_2 \setminus \{a\})}{p(a^*, S_2)}$$

$$\frac{p(a^*, S_1 \setminus \{b\})}{p(a^*, S_1)} = \frac{p(a^*, \{a\})}{p(a^*, \{a, b\})} = \frac{\frac{1-x-y}{1-\alpha y}}{1-x-y} = \frac{1}{1-\alpha y} = \frac{p(a^*, \varnothing)}{p(a^*, \{b\})} = \frac{p(a^*, S_3 \setminus \{b\})}{p(a^*, S_3)}$$

so that i-Independence holds. Finally we show directly that there is no $p_{\succ,\gamma}$ that returns the above probabilities. First of all, by the arguments in the proof of theorem 1 it must be that $\gamma(a) = \frac{x+(1-\alpha)y}{1-\alpha y}$ and $\gamma(b) = \alpha y$. If $a \succ b$, then

$$p_{\succ,\gamma}(b, \{a,b\}) = \gamma(b)(1-\gamma(a)) = \frac{(1-x-y)\alpha y}{1-\alpha y} \neq y = p(b, \{a,b\})$$

where the inequality follows from our assumption that $\alpha > \frac{1}{1-x}$. Next, if $b \succ a$, then

$$p_{\succ,\gamma}(a, \{a,b\}) = \gamma(a)(1-\gamma(b)) = \frac{x+(1-\alpha)y}{1-\alpha y}(1-\alpha y) = x+(1-\alpha)y \neq x = p(a, \{a,b\})$$

where the inequality follows from $\alpha > \frac{1}{1-x} > 1$ and $y > 0$.

Fails only i-Independence: Let $X = \{a,b\}$, and assume choice probabilities are as in the table below:

		$p(a,S_i)$	$p(b,S_i)$	$p(a^*,S_i)$
S_1	$\{a,b\}$	x	y	$1-x-y$
S_2	$\{a\}$	x	-	$1-x$
S_3	$\{b\}$	-	αy	$1-\alpha y$
	$\varnothing$	-	-	1

Table 2: A random choice rule that fails i-Independence

where $x, y > 0$, $x+y < 1$ and $\alpha \in \left(1, \frac{1}{y}\right)$ with $\alpha \neq \frac{1}{1-x}$. While i-Asymmetry holds (since $aS_1b = 1$ and $bS_1a = \alpha \neq 1$), i-Independence does not, since

$$\frac{p(a^*,S_1\backslash\{a\})}{p(a^*,S_1)} = \frac{p(a^*,\{b\})}{p(a^*,\{a,b\})} = \frac{1-\alpha y}{1-x-y} \neq \frac{1}{1-x} = \frac{p(a^*,\varnothing)}{p(a^*,\{a\})} = \frac{p(a^*,S_2\backslash\{a\})}{p(a^*,S_2)}$$

where the inequality follows from our assumption $\alpha \neq \frac{1}{1-x}$.

To see that there is no $p_{\succ,\gamma}$ that returns the above probabilities, observe that by the usual arguments it must be $\gamma(a) = x$ and $\gamma(b) = \alpha y$. Now if $a \succ b$ we have

$$p_{\succ,\gamma}(b, \{a,b\}) = \gamma(b)(1-\gamma(a)) = \alpha y(1-x) \neq y = p(b, \{a,b\})$$

where the inequality follows from $\alpha \neq \frac{1}{1-x}$. If instead $b \succ a$, then

$$p_{\succ,\gamma}(a, \{a,b\}) = \gamma(a)(1-\gamma(b)) = x(1-\alpha y) \neq x = p(a, \{a,b\})$$

where the inequality follows from $\alpha, y > 0$.