Stochastic Choice and Consideration Sets*

Paola Manzini Marco Mariotti†

This version: August 2012

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. Both the preference and the attention parameters are identified uniquely by stochastic choice data. The model is the only one for which the impact of removing any alternative a on the choice probability of any other alternative b is non-negative, asymmetric (either a impacts b or vice-versa), menu independent, neutral (the same on any alternative in the menu), and consistent with the impacts on a and b by a common third alternative.

J.E.L. codes: D0.

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

---

*We are grateful to three insightful referees and the Co-Editor Wolfgang Pesendorfer for many suggestions and for correcting an error. We also thank the seminar audiences at the UCL "Bounded Rationality, Revealed Preferences and Consumer Demand Workshop" and the Ecometric Society NASM 2012, as well as to Chris Flynn, Jacob Goeree, 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.

†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 who 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 [39], Richter [33]) 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 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 [19], Block and Marschak's [3] and Marschak's [21], and more recently Gul, Natenzon and Pesendorfer [15] (henceforth, GNP).

Imperfect attention is what generates randomness and choice errors in our model. Attention is a central element in human cognition (e.g. Anderson [2]) and was recognized in economics as early as in the work of Simon [41]. 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;[2] a doctor short of time for formulating a diagnosis *overlooks* the relevant disease for the given set of symptoms; an ideological voter *deliberately ignores* some candidates independently of their policies.[3] In these examples the decision-maker is able to *evaluate* the alternatives he considers (unlike, for example, a consumer who

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

[2] Goeree [12] quantifies this phenomenon with empirical data.

[3] See Wilson [47] 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.

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[4] 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 [8], [9]. 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.

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 support.[5] 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.

[4]Originating in Wright and Barbour [48]. See also Shocker, Ben-Akiva, Boccara and Nedungadi [40], Roberts and Lattin ([35],[36]) and Roberts and Nedungadi [37].

[5]For example van Nierop *et al* [32] 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.

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,[6] 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.[7] 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 can be 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 five axioms (Theorem 1) that make simple regularity and independence assumptions on the effect of the removal of an alternative a from a menu on the choice probability of another alternative b, which we call in short the *impact* of a on b. Our random choice rule is the only one for which the impact of a on b is: non-negative, asymmetric (exactly one of a and b has a strictly positive impact on the other), menu-independent (it does not depend on other alternatives in the menu), neutral (it is the same, when positive, on any other alternative in the menu), and consistent with the impacts on a and b by a third alternative c (if c impacts b more than a, then a impacts b more than b impacts a).

Our model can be viewed, as explained in section 7.1, as a special type of Random Utility Maximisation, and rationalises plausible types of choice mistakes that cannot be captured by the Luce [19] rule (another special type of Random Utility Maximisation), in which

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

[6] E.g. our own "shortlisting" method [20].

[7] An example is provided in section 6. Tyson [45] clarifies the general structure of two-stage models of choice.

for some strictly positive utility function v.[8] The Luce rule cannot, for example, explain any choice pattern such that $p(a,\{a,b\}) = p(a,\{a,b,c\})$ (which as noted defines a preference for a over c in our setting), since this would imply $p(a,\{a,b\}) = \frac{u(a)}{u(a)+u(b)}$ and $p(a,\{a,b,c\}) = \frac{u(a)}{u(a)+u(b)+u(c)}$, and therefore the contradiction $u(c) = 0$. Yet this type of pattern appears natural when stochastic behavior is governed by a preference ordering and preserves a core feature of rational deterministic choice, namely that inferior alternatives have no bearing on the choice of better ones. Put another way, the Luce rule forces in this respect a strong departure (a form of menu-dependence) from rational deterministic choice. In section 4 we expand on other behavioural phenomena that are compatible with a random consideration set model but not with other models (such as the recent extension of the Luce rule by GNP [15]). Our model reproduces in particular the well-known blue bus/red bus type of menu effect (Debreu [6]), in which an agent's odds of choosing a bus over a train are sensitive to the addition of buses that differ only in colour. 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 choice probability of the remaining duplicate alternatives (by reducing the attention paid to them) instead of increasing it. Finally, we highlight how 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.

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

[8] The Luce model is equivalent to the multinomial logit model (McFadden [26]) popular in econometric studies.

voting, exceeding the time limit for a move in a game of chess).[9] 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, $p(a^*, \varnothing) = 1$.

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 chooser 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 $\gamma(a)$ of being in $C(A)$. Formally:

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*[10]

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

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 (see GNP [15]) to declare

[9] For a recent work on allowing 'not choosing' in the deterministic case, see Gerasimou [11]. Earlier work is Clark [5].

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

$a \succ b$ iff $p(a, A) > p(b, A)$ for some menu A. 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 of a over b has (among others) two observable features: (1) when a is chosen it remains chosen when b is removed;[11] and (2) b can turn from rejected to chosen when a is removed. Either feature reveals unambiguously that $a \succ b$ and has an analog in our random consideration set framework. In case (1), if $a \succ b$, then removing b won't affect the probability that a is chosen, since the latter only depends on the probability that itself and *higher ranked* alternatives are considered. Thus, $p(a, A) = p(a, A \backslash \{b\})$. And conversely, if removing b leaves a's choice probability unchanged, then it must have been the case that b was ranked below a, for otherwise there would have been a positive probability event of it being considered and preventing a from being chosen. Similarly in case (2), if removing a increases the choice probability of b, 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 preventing b from being chosen, so that $p(b, A) > p(b, A \backslash \{a\})$. Together, $p(a, A) = p(a, A \backslash \{b\})$ and $p(b, A) > p(b, A \backslash \{a\})$ constitute the revealed preference relation $a \succ b$ of our model. We will show that this relation is revealed uniquely.

Next, given a preference $\succ$, clearly 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. However, provided that there are at least three alternatives and that binary menus are included in the domain, $\gamma(a)$ is identified even without the need of knowing the preference

[11] This feature is called Chernoff's axiom.

or indeed the choice probability of a from any menu, 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

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

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

As explained in the introduction, all our axioms constrain the 'impacts' $\frac{p(a,A \setminus \{b\})}{p(a,A)}$ that an alternative b has on another alternative a. The following properties are intended for all $A, B \in \mathcal{D}$:

A1 (Non-negativity). For all $a \in A^*$ and $b \in A$: $\frac{p(a,A \setminus \{b\})}{p(a,A)} \geq 1$.

A2 (Asymmetry). For all $a, b \in A$: $\frac{p(a,A \setminus \{b\})}{p(a,A)} \neq 1 \Leftrightarrow \frac{p(b,A \setminus \{a\})}{p(b,A)} = 1$.

A3 (Menu independence). For all $a \in A^* \cap B^*$ and $b \in A \cap B$: $\frac{p(a,A \setminus \{b\})}{p(a,A)} = \frac{p(a,B \setminus \{b\})}{p(a,B)}$.

A4 (Neutrality). For all $a, b \in A^*$ and $c \in A$: $\frac{p(a,A \setminus \{c\})}{p(a,A)}, \frac{p(b,A \setminus \{c\})}{p(b,A)} > 1 \Rightarrow \frac{p(a,A \setminus \{c\})}{p(a,A)} = \frac{p(b,A \setminus \{c\})}{p(b,A)}$.

A5 (Third-alternative consistency). For all $a, b, c \in A$: $\frac{p(a,A \setminus \{c\})}{p(a,A)} < \frac{p(b,A \setminus \{c\})}{p(b,A)} \Rightarrow \frac{p(b,A \setminus \{a\})}{p(b,A)} > \frac{p(a,A \setminus \{b\})}{p(a,A)}$

The iteration of A1 leads to the standard axiom of Regularity (or Monotonicity)[12]. A2 says that (as in rational deterministic choice) exactly one of any two alternatives

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

has a strictly positive impact on the other. A3 states that the effect of removing an alternative on the choice probability of another alternative should not depend on which other alternatives are present in the menu. It is a simple form of menu-independence, alternative to Luce's IIA. Instead A4, stating that the positive impact of an alternative is the same on any other alternative in the menu, is implied by the Luce axiom. In fact, A4 also states that $\frac{p(a,A\backslash\{c\})}{p(b,A\backslash\{c\})} = \frac{p(a,A)}{p(b,A)}$ under the restriction in the premise, while the Luce axiom asserts the same form of menu independence (and more) unconditionally. We elaborate on why this restriction of Luce's IIA is sensible in section 4. Finally, A5 concludes that a is more impacted by b than vice-versa from the observation of the stronger impact exerted on a by a common third alternative c. In this sense the relative strengths of two alternatives are revealed by the behaviour of these alternatives against the common third alternative (similarly to the way in which, in rational deterministic choice, if c does not affect whether b is chosen but it does affect whether a is chosen, then b must be better than a).

Theorem 1 *Let $\mathcal{D} = 2^X$. A random choice rule p such that $p(a, A) \in (0, 1)$ for all $A \in \mathcal{D}$ and for all $a \in A$, satisfies A1-A5 if and only if it is a random consideration set rule $p_{\succ,\gamma}$. Moreover, both the quality ranking $\succ$ and the attention profile γ are unique, that is, for any random choice rule $p_{\succ',\gamma'}$ such that $p_{\succ',\gamma'} = p$ we have $(\succ', \gamma') = (\succ, \gamma)$.*

The logic behind the sufficiency part of the proof is simple. Under A1-A3 the revealed preference described in section 3.1 is total and asymmetric, and adding A5 makes it 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 A3-A4 are shown to imply the following property: whenever b has a positive impact on 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 impact of b on a must depend only on the 'strength' of b in singleton choice.[13] Finally, the iterated application of this formula shows that the preference and the attention parameters defined above retrieve the given choice probabilities via the assumed procedure.

[13] A similar stochastic path independence property appears as an axiom in Yildiz [50].

4 Explaining Menu effects and Stochastic Intransitivity

4.1 Menu effects

Luce [19] states a menu-independence property that characterizes the Luce choice rule:

Luce's IIA: For all $A, B \in \mathcal{D}$ and $a, b \in A \cap B$: $\frac{p(a,A)}{p(b,A)} = \frac{p(a,B)}{p(b,B)}$.

Our model suggests that a natural reason why this property 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 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 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 is the general property that the presence of a high-attention alternative (c in the example) strongly affects the choice probabilities of the alternatives worse than c while leaving unaffected the choice probabilities of those better than c.

However, as our characterization makes clear, a random consideration set rule does satisfy other forms of menu-independence and consistency that are arguably as natural as Luce's IIA.

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.[14] The *blue bus/red bus example* (Debreu [6]) is the standard illustration, in which menu effects appear 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 pairwise 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.[15]

GNP [15] 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.[16]

Our model (once straightforwardly adapted to account for preference ties), highlights

[14] By adding a nested structure to 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.

[15] For the sake of precision, 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.

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

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 equal 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 $p_{\succ,\gamma}(t, \{b, r, t\}) = y$. Furthermore, we have (assuming that in case both buses are considered then each of them is chosen with probability $\frac{1}{2}$):

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

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

and therefore

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

so that, independently of the attention profile γ,

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

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:

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

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

and therefore

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

so that independently of the attention profile γ

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

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, we argue that the blue bus/red bus example is 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, and in this case it is based in particular on the hypothesis that a Luce-like model continues to hold (as the GNP [15] modification of the Luce rule well illustrates). 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 not necessarily a reliable 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 [42], 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[17]

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

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

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 attention ordering 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. In this section we show

[17] In their survey on choice anomalies Rieskamp, Busemeyer and Meller [34] write: "Does human choice behavior obey the principle of strong stochastic transitivity? An overwhelming number of studies suggest otherwise" (p. 646).

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 rule $p_{\succ,\delta}$ for which there exists a pair $(\succ, \delta)$, where $\succ$ is a linear order on X and δ is a map $\delta : X \times \mathcal{D} \rightarrow (0,1)$, such that

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

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 $A \in \mathcal{D}$ and $a \in A$, 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 are appropriate, especially as 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.* [32]).

6 Related literature

The most related economic papers to ours conceptually are MNO [23] and Eliaz and Spiegler ([8], [9]). 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 ([43], [44]) 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 [15] 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.

Stochastic choice is also the focus of the anticipated choice model characterised in Koida [18]. In this model, though, the emphasis is on how a decision maker's (probabilistic) mental states drive the choice of an alternative from each menu, in turn determining in non-obvious ways the agent's preferences over menus. In our setup we instead concentrate on mistakes *before* choice is made.

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 [46]). 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. In one way 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 the latter assumes 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 more recent works by Matějka and McKay [24] and Cheremukhin, Popova and A. Tutino [4] in which it is assumed that an agent faces the costly option of *studying* (rather than implementing) the outcome and chooses the level of attention optimally.

Recently, Rubinstein and Salant [38] 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. However, their analysis takes a very different direction from ours in that it eschews any stochastic element.

Finally, we note that the natural 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. William James [17], Daniel Dennett [7], Martin Heisenberg [16]) as a fundamental feature of human choices, and as a solution of the general problem of free will.

7 Concluding remarks[18]

7.1 Random consideration sets and RUM

A Random Utility Maximization (RUM) rule [3] is defined by a probability distribution π on the possible rankings of the elements of X and the assumption that the agent picks the top element of the ranking extracted according to π. Block and Marschak [3], McFadden [26] and Yellot [49] 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. The example given in the introduction of the pattern $p(a, \{a, b\}) = p(a, \{a, b, c\})$ shows that the choice probabilities of a random consideration set rule can never be generated by the particular RUM rule constituted by a Luce model. However a random consideration set rule $(\succ, \gamma)$ is itself a special type of RUM rule. In this rule, π assigns independently for each alternative a probability $(1 - \gamma(a))$ to all rankings in which a is below the default alternative a^*, and if π assigns positive probability to a ranking in which b is above a but $a \succ b$, then a is also ranked below a^*. An interpretation of this RUM is that the agent is 'in the mood' for each alternative a with probability $\gamma(a)$ (and otherwise he prefers the default alternative), and he 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 wasn't in the mood for b.

7.2 Comparative attention

A natural definition of comparative attention based on observed choice probabilities is as follows: $(\succ_1, \gamma_1)$ is *more attentive than* $(\succ_2, \gamma_2)$, denoted $(\succ_1, \gamma_1) \, \alpha \, (\succ_2, \gamma_2)$, 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

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

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.

For general domains, the previous characterisation of the relation α 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 if 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 $p(b, \{a,b\}) = \beta$.[19] 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 $\frac{\gamma(b)}{1-\gamma(b)} = \frac{\beta}{\alpha}\gamma(a)$.

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

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

References

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

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

[3] 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.

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

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

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

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

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

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

[10] Falmagne, Jean-Calude (1978) "A Representation Theorem for Finite Random Scale Systems", *Journal of Mathematical Psychology* 18: 52-72.

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

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

[13] Goeree, J. K., C. A. Holt and T. R. Palfrey (2008) “Quantal response equilibria”, in S. N. Durlauf and L. E. Blume (eds.),The New Palgrave Dictionary of Economics, 2nd Edition.

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

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

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

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

[18] Koida, N. (2010) “Anticipated Stochastic Choice”, mimeo, Iwate Prefectural University.

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

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

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

[22] Masatlioglu, Y. and D. Nakajima (2008) “Choice by Constrained Elimination” mimeo, University of Michigan.

[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] McFadden, D. and K. Richter (1991) “Stochastic Rationality and Revealed Stochastic Preference”, in J. Chipman, D. McFadden, K. Richter (eds) Preferences, Uncertainty, and Rationality, Westview Press, 161-186.

[29] McKelvey, R. D. and T. R. Palfrey R. (1995) “Quantal Response Equilibrium for Normal Form Games,” *Games and Economic Behavior*, 10, 6-38.

[30] McKelvey, R. D. and T. R. Palfrey R. (1998) “Quantal Response Equilibria for Extensive Form Games,” *Experimental Economics* 1: 9-41.

[31] Nedungadi, P. (1990) “Recall and consumer consideration sets: Influencing choice without altering brand evaluations", *Journal of Consumer Research*, 17(3): 263–276.

[32] 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.

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

[34] Rieskamp, Jörg, Jerome R. Busemeyer and Barbara A. Mellers (2006) "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice" Journal of Economic Literature 44: 631-661.

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

[36] 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.

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

[38] 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.

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

[40] 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.

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

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

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

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

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

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

[47] 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.

[48] Wright, Peter, and Fredrick 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.

[49] 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.

[50] 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. For sufficiency, let p be a random choice rule that satisfies A1-A5. Define a binary relation R on X by aRb iff there exists $A \in \mathcal{D}$ with $a, b \in A$ such that $p(b, A \setminus \{a\}) > p(b, A)$. We show that R is total, asymmetric and transitive. Totality follows immediately from A1-A2 and asymmetry from A2-A3. For transitivity, suppose $aRbRc$. Then $p(b, A \setminus \{a\}) > p(b, A \setminus \{a\})$ and $p(c, B \setminus \{c\}) > p(c, B)$ for some A and B with $a, b \in A$ and $b, c \in C$. By A3 the first inequality implies

$$p(b, \{b, c\}) > p(b, \{a, b, c\})$$

and similarly $p(c, \{a, c\}) > p(c, \{a, b, c\})$. If by contradiction *not* (aRc), then since R is total cRa, that is $p(a, \{a, b\}) > p(a, \{a, b, c\})$ and thus by A2

$$p(c, \{a, b, c\}) = p(c, \{b, c\})$$

By the two displayed formulae and A5 we have $\frac{p(b,\{a,b\})}{p(b,\{a,b,c\})} > \frac{p(c,\{a,c\})}{p(a,\{a,b,c\})}$, a contradiction with $\frac{p(c,\{a,c\})}{p(a,\{a,b,c\})} > 1$ (A1-A2 imply that exactly one of the two ratios is greater than one). Finally, observe that (using A2 and A3) the following three statements are equivalent:

$$\begin{array}{rcl} & aRb & \\ p(b, A \setminus \{a\}) & > & p(b, A) \text{ for all } A \in \mathcal{D} \text{ with } a, b \in A \\ p(a, A \setminus \{b\}) & = & p(a, A) \text{ for all } A \in \mathcal{D} \text{ with } a, b \in A \end{array} \quad (1)$$

Next, we show that for all $A \in \mathcal{D}$,

$$p(a, A\backslash\{b\}) > p(a, A) \Rightarrow p(a, A\backslash\{b\})(1 - p(b, \{b\})) = p(a, A) \tag{2}$$

Since $p(a^*, \varnothing) = 1$ and $p(a^*, \{b\}) + p(b, \{b\}) = 1$ we have $\frac{p(a^*, \varnothing)}{p(a^*, \{b\})} = \frac{1}{1-p(b,\{b\})}$, so that by A3 $\frac{p(a^*, A\backslash\{b\})}{p(a^*, \{A\})} = \frac{1}{1-p(b,\{b\})} > 1$ for all $A \in \mathcal{D}$ with $b \in A$ and then by A4 $\frac{p(a, A\backslash\{b\})}{p(a, \{A\})} = \frac{1}{1-p(b,\{b\})}$ for all $a \in A$ for which $\frac{p(a, A\backslash\{b\})}{p(a, \{A\})} > 1$, establishing the claim.

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 (2) 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 the equality in the second to last line follows from the properties of R displayed in (1).

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 exists $a, b \in X$ such that $a \succ b$ and $b \succ' a$. Take $A = \{a\} \cup \{c \in X : a \succ c\}$, for which $b \in A$ 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

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

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 linear order of the alternatives. Define δ by setting, for $A \in \mathcal{D}$ and $a \in A$:

$$\delta(a, A) = \frac{p(a, A)}{1 - \sum_{b \in A : b \succ a} p(b, A)} \tag{3}$$

Observe that $\delta(a, A) > 0$ whenever $p(a, A) > 0$ and $\delta(a, A) < 1$ since $1 > p(a, A) + \sum_{b \in A: b \succ a} p(b, A)$.

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 derive

$$1 - \delta(b, A) = \frac{1 - \sum_{c \in A: c \succ b} p(c, A) - p(b, A)}{1 - \sum_{c \in A: c \succ b} p(c, A)} \tag{4}$$

so that

$$\prod_{b \in A: b \succ a} (1 - \delta(b, A)) = \prod_{b \in A: b \succ a} \frac{1 - \sum_{c \in A: c \succ b} p(c, A) - p(b, A)}{1 - \sum_{c \in A: c \succ b} p(c, A)} \tag{5}$$

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 (4) we have that

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

As the numerator of the expression for $1 - \delta(b^+, A)$ is equal to the denominator of the expression for $1 - \delta(b, A)$, more in general the product in (5) 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

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

as desired (where the first term in the second line follows from (3)). ■

Possible online appendices

A 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.

A1 (Non-negativity). For all $a \in A^*$ and $b \in A$: $\frac{p(a,A\setminus\{b\})}{p(a,A)} \geq 1$.

A2 (Asymmetry). For all $a, b \in A$: $\frac{p(a,A\setminus\{b\})}{p(a,A)} > 1 \Leftrightarrow \frac{p(b,A\setminus\{a\})}{p(b,A)} \neq 1$.

A3 (Menu independence). For all $a \in A^* \cap B^*$ and $b \in A \cap B$: $\frac{p(a,A\setminus\{b\})}{p(a,A)} = \frac{p(a,B\setminus\{b\})}{p(a,B)}$.

A4 (Neutrality). For all $a, b \in A^*$ and $c \in A$: $\frac{p(a,A\setminus\{c\})}{p(a,A)}, \frac{p(b,A\setminus\{c\})}{p(b,A)} > 1 \Rightarrow \frac{p(a,A\setminus\{c\})}{p(a,A)} = \frac{p(b,A\setminus\{c\})}{p(b,A)}$.

A5 (Third-alternative consistency). For all $a, b, c \in A$: $\frac{p(a,A\setminus\{c\})}{p(a,A)} < \frac{p(b,A\setminus\{c\})}{p(b,A)} \Rightarrow \frac{p(b,A\setminus\{a\})}{p(b,A)} > \frac{p(a,A\setminus\{b\})}{p(a,A)}$

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

where $\delta, x, y, z \in (0, 1)$, and x, y, z and δ are further restricted so that $p(a^*, S_i) \in 0, 1$ for all $i = 1, ..7$. This random choice rule fails Non-negativity, since e.g. $S_2 \subset S_1$ while $p(b, S_2) < p(b, S_1)$. It does satisfy the other axioms. To see this, first of all observe that, for all S_i, $\frac{p(a,S_i\setminus\{b\})}{p(a,S_i)} = 1 = \frac{p(a,S_i\setminus\{c\})}{p(a,S_i)} = \frac{p(c,S_i\setminus\{b\})}{p(c,S_i)}$ and $\frac{p(b,S_i\setminus\{a\})}{p(b,S_i)} = \delta = \frac{p(b,S_i\setminus\{c\})}{p(b,S_i)} = \frac{p(c,S_i\setminus\{a\})}{p(c,S_i)}$, so that Menu Independence holds. From these it is easy to see that the only case where the premise of Third-alternative consistency is true is when removing c from S_1, so that

$$\frac{p(a, \{a, b\})}{p(a, \{a, b, c\})} = \frac{p(a, S_1 \setminus \{c\})}{p(a, S_1)} = 1 > \delta = \frac{p(b, S_1 \setminus \{c\})}{p(b, S_i)} = \frac{p(b, \{a, b\})}{p(b, \{a, b, c\})}$$

Since however $\frac{p(a,S_1\setminus\{b\})}{p(a,S_1)} = 1 > \delta = \frac{p(b,S_1\setminus\{a\})}{p(b,S_1)}$, Third-alternative consistency holds. For the remaining two axioms it is immediate to verify that both Neutrality (since its premise

		$p(a,S_i)$	$p(b,S_i)$	$p(c,S_i)$	$p(a^*,S_i)$
S_1	$\{abc\}$	x	y	z	$1-x-y-z$
S_2	$\{ab\}$	x	δy	-	$1-x-\delta y$
S_3	$\{ac\}$	x	-	z	$1-x-z$
S_4	$\{bc\}$	-	δy	δz	$1-\delta(y+z)$
S_5	$\{a\}$	x	-	-	$1-x$
S_6	$\{b\}$	-	$\delta^2 y$	-	$1-\delta^2 y$
S_7	$\{c\}$	-	-	δz	$1-\delta z$

Table 1: A random choice rule that fails Non-negativity

is never true) and Asymmetry hold. The latter is easily checked by inspection of the following:

$$\begin{array}{c} p(a,\{a,b,c\}\setminus\{b\}) = p(a,\{a,b,c\}) \\ p(b,\{a,b,c\}\setminus\{a\}) \neq p(b,\{a,b,c\}) \end{array} ; \begin{array}{c} p(a,\{a,b,c\}\setminus\{c\}) = p(a,\{a,b,c\}) \\ p(c,\{a,b,c\}\setminus\{a\}) \neq p(c,\{a,b,c\}) \end{array}$$

$$\begin{array}{c} p(a,\{a,b\}\setminus\{b\}) = p(a,\{a,b\}) \\ p(b,\{a,b\}\setminus\{a\}) \neq p(b,\{a,b\}) \end{array} ; \begin{array}{c} p(a,\{a,c\}\setminus\{c\}) = p(a,\{a,c\}) \\ p(c,\{a,c\}\setminus\{a\}) \neq p(c,\{a,c\}) \end{array}$$

$$\begin{array}{c} p(b,\{a,b,c\}\setminus\{c\}) \neq p(b,\{a,b,c\}) \\ p(c,\{a,b,c\}\setminus\{b\}) = p(c,\{a,b,c\}) \end{array} ; \begin{array}{c} p(b,\{b,c\}\setminus\{b\}) \neq p(b,\{b,c\}) \\ p(c,\{b,c\}\setminus\{b\}) = p(c,\{b,c\}) \end{array}$$

Finally we also show directly that there is no $p_{\succ,\gamma}$ that returns the above probabilities. Since $p(a,S_i) = x$ for all S_i, in order to have $p_{\succ,\gamma} = p$ it must be $a \succ b, c$ and $\gamma(a) = x$. Then $\frac{p_{\succ,\gamma}(b,\{a,b\})}{p_{\succ,\gamma}(b,\{b\})} = (1-\gamma(a))$, so that $p_{\succ,\gamma} = p$ and $\frac{p(b,\{a,b\})}{p(b,\{b\})} = \frac{1}{\delta}$ imply the contradiction $1 > 1 - x = \frac{1}{\delta} > 1$.

Fails only Asymmetry: Let $X = \{a,b\}$, $p(a,\{a,b\}) = x = p(a,\{a\})$, $p(b,\{a,b\}) = y = p(b,\{b\})$, $x, y \in (0,1)$ and $x + y < 1$.

It is immediate to verify that the other four axioms hold (noting that the premise of the last three is false in both). If there were a $p_{\succ,\gamma}$ that returned the above probabilities, let $a \succ b$. Then $p_{\succ,\gamma}(a,\{a,b\}) = \gamma(a)(1-\gamma(b)) \neq \gamma(a) = p_{\succ,\gamma}(a,\{a\})$ since $\gamma(a), \gamma(b) > 0$. Similarly if $b \succ a$, so that the above random choice rule is not a random consideration set rule.

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

		$p(a, S_i)$	$p(b, S_i)$	$p(c, S_i)$	$p(a^*, S_i)$
S_1	$\{abc\}$	x	y	z	$1 - x - y - z$
S_2	$\{ab\}$	x	αy	-	$1 - x - \alpha y$
S_3	$\{ac\}$	x	-	z	$1 - x - z$
S_4	$\{bc\}$	-	αy	αz	$1 - \alpha(y + z)$
S_5	$\{a\}$	αx	-	-	$1 - \alpha x$
S_6	$\{b\}$	-	αy	-	$1 - \alpha y$
S_7	$\{c\}$	-	-	z	$1 - z$

Table 2: A random choice rule that fails Menu Independence

where $\alpha > 1$, $x, y, z \in (0, 1)$, and x, y, z and α are further restricted so that $p(a^*, S_i) \in 0, 1$ for all $i = 1, ..7$. This random choice rule fails Menu Independence, since across sets S_1 and S_3 we have $\frac{p(a, S_1 \setminus \{c\})}{p(a, S_1)} = 1 < \alpha = \frac{p(a, S_3 \setminus \{c\})}{p(a, S_3)}$. Verifying that Non-negativity and Asymmetry hold is immediate - for Asymmetry refer to the following:

$$\begin{array}{c} p(a, \{a, b, c\} \setminus \{b\}) = p(a, \{a, b, c\}) \\ p(b, \{a, b, c\} \setminus \{a\}) \neq p(b, \{a, b, c\}) \end{array} ; \begin{array}{c} p(a, \{a, b, c\} \setminus \{c\}) = p(a, \{a, b, c\}) \\ p(c, \{a, b, c\} \setminus \{a\}) \neq p(c, \{a, b, c\}) \end{array}$$

$$\begin{array}{c} p(a, \{a, b\} \setminus \{b\}) \neq p(a, \{a, b\}) \\ p(b, \{a, b\} \setminus \{a\}) = p(b, \{a, b\}) \end{array} ; \begin{array}{c} p(a, \{a, c\} \setminus \{c\}) \neq p(a, \{a, c\}) \\ p(c, \{a, c\} \setminus \{a\}) = p(c, \{a, c\}) \end{array}$$

$$\begin{array}{c} p(b, \{a, b, c\} \setminus \{c\}) \neq p(b, \{a, b, c\}) \\ p(c, \{a, b, c\} \setminus \{b\}) = p(c, \{a, b, c\}) \end{array} ; \begin{array}{c} p(b, \{b, c\} \setminus \{b\}) = p(b, \{b, c\}) \\ p(c, \{b, c\} \setminus \{b\}) \neq p(c, \{b, c\}) \end{array}$$

For Neutrality, observe that it can be violated only in S_1 when a is removed, as removing either b or c leaves the impact on a unchanged, so that the premise of Neutrality fails. If instead a is removed, then since $\frac{p(b, S_1 \setminus \{a\})}{p(b, S_1)}, \frac{p(c, S_1 \setminus \{a\})}{p(c, S_1)} = \alpha > 1$, Neutrality holds. Finally, for Third-alternative consistency too it suffices to consider S_1. When either a or b are removed the premise of the axiom is violated, since either $\frac{p(b, S_1 \setminus \{a\})}{p(b, S_1)} = \alpha = \frac{p(c, S_1 \setminus \{a\})}{p(c, S_1)}$ or $\frac{p(a, S_1 \setminus \{b\})}{p(a, S_1)} = 1 = \frac{p(c, S_1 \setminus \{b\})}{p(c, S_1)}$. If c is removed we have $\frac{p(b, S_1 \setminus \{c\})}{p(b, S_1)} = \alpha > 1 = \frac{p(a, S_1 \setminus \{c\})}{p(a, S_1)}$; since $\frac{p(b, S_1 \setminus \{a\})}{p(b, S_1)} = \alpha > 1 = \frac{p(a, S_1 \setminus \{b\})}{p(a, S_1)}$, Third-alternative consistency holds.

To see that there is no $p_{\succ,\gamma}$ that returns the above probabilities, observe that $p_{\succ,\gamma}(a,\{a\}) = \gamma(a)$, so that to have $p_{\succ,\gamma}(a,\{a\}) = p(a,\{a\})$ it must be $\gamma(a) = \alpha x$. Since $p_{\succ,\gamma}(a,\{a,b\}) < \gamma(a)$, then $b \succ a$, so that $p_{\succ,\gamma}(a,\{a,b\}) = \gamma(a)(1-\gamma(b))$ and to have $p_{\succ,\gamma}(a,\{a,b\}) = p(a,\{a,b\}) = x$ it must be $1-\gamma(b) = \frac{1}{\alpha}$. Similar arguments for $\{a,c\}$ yield $c \succ a$ and $1-\gamma(c) = \frac{1}{\alpha}$. Then $p_{\succ,\gamma}(a,S_1) = \gamma(a)(1-\gamma(b))(1-\gamma(c)) = \frac{x}{\alpha} \neq p(a,S_1)$, so that the above random choice rule is not a random consideration set rule.

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

		$p(a,S_i)$	$p(b,S_i)$	$p(c,S_i)$	$p(a^*,S_i)$
S_1	$\{abc\}$	x	y	z	$1-x-y-z$
S_2	$\{ab\}$	αx	βy	-	$1-\alpha x-\beta y$
S_3	$\{ac\}$	x	-	z	$1-x-z$
S_4	$\{bc\}$	-	βy	z	$1-\beta y-z$
S_5	$\{a\}$	αx	-	-	$1-\alpha x$
S_6	$\{b\}$	-	$\beta^2 y$	-	$1-\beta^2 y$
S_7	$\{c\}$	-	-	z	$1-z$

Table 3: A random choice rule that fails Neutrality

where $\beta > \alpha > 1$, $x,y,z \in (0,1)$, and x,y,z,α and β are further restricted so that $p(a^*,S_i) \in 0,1$ for all $i = 1,..7$. This random choice rule violates Neutrality since although $\frac{p(a,A\setminus\{c\})}{p(a,A)}, \frac{p(b,A\setminus\{c\})}{p(b,A)} > 1$, we have $\frac{p(a,A\setminus\{c\})}{p(a,A)} = \alpha \neq \beta = \frac{p(b,A\setminus\{c\})}{p(b,A)}$. As for the other axioms, it is immediate to verify that this random choice rule satisfies Non-negativity and Asymmetry - for the latter refer to the following:

$$\begin{array}{c} p(a,\{a,b,c\}\setminus\{b\}) = p(a,\{a,b,c\}) \\ p(b,\{a,b,c\}\setminus\{a\}) \neq p(b,\{a,b,c\}) \end{array} ; \begin{array}{c} p(a,\{a,b,c\}\setminus\{c\}) \neq p(a,\{a,b,c\}) \\ p(c,\{a,b,c\}\setminus\{a\}) = p(c,\{a,b,c\}) \end{array}$$

$$\begin{array}{c} p(a,\{a,b\}\setminus\{b\}) = p(a,\{a,b\}) \\ p(b,\{a,b\}\setminus\{a\}) \neq p(b,\{a,b\}) \end{array} ; \begin{array}{c} p(a,\{a,c\}\setminus\{c\}) \neq p(a,\{a,c\}) \\ p(c,\{a,c\}\setminus\{a\}) = p(c,\{a,c\}) \end{array}$$

$$\begin{array}{c} p(b,\{a,b,c\}\setminus\{c\}) \neq p(b,\{a,b,c\}) \\ p(c,\{a,b,c\}\setminus\{b\}) = p(c,\{a,b,c\}) \end{array} ; \begin{array}{c} p(b,\{b,c\}\setminus\{b\}) \neq p(b,\{b,c\}) \\ p(c,\{b,c\}\setminus\{b\}) = p(c,\{b,c\}) \end{array}$$

For Menu Independence, observe that for all S_i: $\frac{p(a,S_i\setminus\{b\})}{p(a,S_i)} = 1$, $\frac{p(a,S_i\setminus\{c\})}{p(a,S_i)} = \alpha$, $\frac{p(b,S_i\setminus\{a\})}{p(b,S_i)} = \frac{p(b,S_i\setminus\{c\})}{p(b,S_i)} = \beta$ and $\frac{p(c,S_i\setminus\{a\})}{p(c,S_i)} = \frac{p(c,S_i\setminus\{b\})}{p(c,S_i)} = 1$. To check Third-alternative consistency, observe that the premise holds only when eliminating either a or c from S_1. In the first case we have $\frac{p(b,S_1\setminus\{a\})}{p(b,S_1)} = \beta > 1 = \frac{p(c,S_1\setminus\{a\})}{p(c,S_1)}$. Since however $\frac{p(b,S_1\setminus\{c\})}{p(b,S_1)} = \beta > 1 = \frac{p(c,S_1\setminus\{b\})}{p(c,S_1)}$, Third-alternative consistency holds. In the second case we have $\frac{p(b,S_1\setminus\{c\})}{p(b,S_1)} = \beta > \alpha = \frac{p(a,S_1\setminus\{c\})}{p(a,S_1)}$, but since $\frac{p(b,S_1\setminus\{a\})}{p(b,S_1)} = \beta > 1 = \frac{p(a,S_1\setminus\{b\})}{p(a,S_1)}$, Third-alternative consistency is satisfied. To see that the random choice rule in Table 3 is not a random consideration set rule, observe that since $p(c,S_i) = z$ for all i, and since $\gamma(k) \in (0,1)$ for all $k \in \{a,b,c,a^*\}$, to have $p(c,S_i) = p_{\succ,\gamma}(c,S_i)$ requires $\gamma(c) = z$ and $c \succ a,b$. Then $p_{\succ,\gamma}(a,\{a,c\}) = \gamma(a)(1-\gamma(c))$, so that $p_{\succ,\gamma}(a,\{a,c\}) = p(a,\{a,c\}) = x$ requires $\gamma(a) = \frac{x}{1-z}$. In addition, $\gamma(a) = p_{\succ,\gamma}(a,\{a\}) = p(a,\{a\}) = \alpha x$ implies the restriction $\alpha = \frac{1}{1-z}$. Similarly, $p_{\succ,\gamma}(b,\{b,c\}) = p(b,\{b,c\}) = \beta y$ and $p_{\succ,\gamma}(b,\{b,c\}) = \gamma(b)(1-\gamma(c))$ requires $\gamma(b) = \frac{\beta y}{1-z}$ and $p_{\succ,\gamma}(b,\{b\}) = p(b,\{b\}) = \beta^2 y = \gamma(b)$ requires $\beta = \frac{1}{1-z}$. Since however $\beta > \alpha$, we have the contradiction $\frac{1}{1-z} = \alpha < \beta = \frac{1}{1-z}$.

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

		$p(a,S_i)$	$p(b,S_i)$	$p(c,S_i)$	$p(a^*,S_i)$
S_1	$\{abc\}$	x	y	z	$1-x-y-z$
S_2	$\{ab\}$	αx	y	-	$1-\alpha x-y$
S_3	$\{ac\}$	x	-	αz	$1-x-\alpha z$
S_4	$\{bc\}$	-	αy	z	$1-\alpha y-z$
S_5	$\{a\}$	αx	-	-	$1-\alpha x$
S_6	$\{b\}$	-	αy	-	$1-\alpha y$
S_7	$\{c\}$	-	-	αz	$1-\alpha z$

Table 4: A random choice rule that fails Neutrality

where $\alpha > 1$, $x,y,z \in (0,1)$, and x,y,z and α are further restricted so that $p(a^*,S_i) \in 0,1$ for all $i = 1,..7$. This random choice rule violates Third-alternative consistency since $\frac{p(b,S_1\setminus\{c\})}{p(b,S_1)} = 1 < \alpha = \frac{p(a,S_1\setminus\{c\})}{p(a,S_1)}$ while $\frac{p(a,S_1\setminus\{b\})}{p(a,S_1)} = 1 < \alpha = \frac{p(b,S_1\setminus\{a\})}{p(b,S_1)}$. As for the other axioms, it is immediate to verify that this random choice rule satisfies Non-negativity and

Asymmetry - for Asymmetry see the following:

$$\begin{array}{c} p\left(a,\{a,b,c\}\setminus\{b\}\right)=p\left(a,\{a,b,c\}\right) \\ p\left(b,\{a,b,c\}\setminus\{a\}\right)\neq p\left(b,\{a,b,c\}\right) \end{array};\begin{array}{c} p\left(a,\{a,b,c\}\setminus\{c\}\right)\neq p\left(a,\{a,b,c\}\right) \\ p\left(c,\{a,b,c\}\setminus\{a\}\right)=p\left(c,\{a,b,c\}\right) \end{array}$$

$$\begin{array}{c} p\left(a,\{a,b\}\setminus\{b\}\right)=p\left(a,\{a,b\}\right) \\ p\left(b,\{a,b\}\setminus\{a\}\right)\neq p\left(b,\{a,b\}\right) \end{array};\begin{array}{c} p\left(a,\{a,c\}\setminus\{c\}\right)\neq p\left(a,\{a,c\}\right) \\ p\left(c,\{a,c\}\setminus\{a\}\right)=p\left(c,\{a,c\}\right) \end{array}$$

$$\begin{array}{c} p\left(b,\{a,b,c\}\setminus\{c\}\right)=p\left(b,\{a,b,c\}\right) \\ p\left(c,\{a,b,c\}\setminus\{b\}\right)\neq p\left(c,\{a,b,c\}\right) \end{array};\begin{array}{c} p\left(b,\{b,c\}\setminus\{b\}\right)=p\left(b,\{b,c\}\right) \\ p\left(c,\{b,c\}\setminus\{b\}\right)\neq p\left(c,\{b,c\}\right) \end{array}$$

For Menu Independence, observe that $\frac{p(a,S\setminus\{c\})}{p(a,S)}=\alpha=\frac{p(b,S\setminus\{a\})}{p(b,S)}=\frac{p(c,S\setminus\{b\})}{p(c,S)}$ for all pertinent S and $\frac{p(a,S\setminus\{b\})}{p(a,S)}=1=\frac{p(b,S\setminus\{c\})}{p(b,S)}=\frac{p(c,S\setminus\{a\})}{p(c,S)}$ for all pertinent S, so that Menu Independence holds. This also helps checking that Neutrality holds, as eliminating any of the alternatives from S_1 produces impact ratios for the other two alternatives only one of which is greater than 1, so that the premise of Neutrality is violated. To see that this random choice rule is not a random consideration set rule, note that

$$\frac{p_{\succ,\gamma}\left(a,\{a,b\}\right)}{p_{\succ,\gamma}\left(a,\{a,b,c\}\right)}=\begin{cases} 1 & \text{if } a\succ c \\ \frac{1}{1-\gamma(c)} & \text{otherwise} \end{cases}$$

Since $\frac{p(a,\{a,b\})}{p(a,\{a,b,c\})}=\alpha>1$, in order to have $p_{\succ,\gamma}=p$ it must be $\alpha=\frac{1}{1-\gamma(c)}$ and $c\succ a$, so that $\gamma(c)=\frac{\alpha-1}{\alpha}$. Then substituting into $p_{\succ,\gamma}\left(a,\{a,c\}\right)=\gamma(a)\left(1-\gamma(c)\right)$ yields $p_{\succ,\gamma}\left(a,\{a,c\}\right)=\frac{\gamma(a)}{\alpha}$. On the other hand, $p\left(a,\{a,c\}\right)=x$, so that $p\left(a,\{a,c\}\right)=p_{\succ,\gamma}\left(a,\{a,c\}\right)$ requires $\gamma(a)=\alpha x$. Since $p\left(a,\{a,b\}\right)=\alpha x=\gamma(a)$, to have $p_{\succ,\gamma}=p$ it must be $a\succ b$. Then $p_{\succ,\gamma}\left(b,\{a,b,c\}\right)=\gamma(b)\left(1-\gamma(a)\right)\left(1-\gamma(c)\right)=\gamma(b)\left(1-\alpha x\right)\frac{1}{\alpha}$, and since $p\left(b,\{a,b,c\}\right)=y$ it must be $\gamma(b)=\frac{\alpha y}{1-\alpha x}$. Since $1-\alpha x\neq 1$, we have that $p=p_{\succ,\gamma}$ is contradicted on $\{b\}$, since $p\left(b,\{b\}\right)=\alpha y\neq\frac{\alpha y}{1-\alpha x}=\gamma(b)=p_{\succ,\gamma}\left(b,\{b\}\right)$.

B Necessity of the axioms

A1 (Non-negativity). For all $a\in A^*$ and $b\in A$: $\frac{p(a,A\setminus\{b\})}{p(a,A)}\geq 1$.

Let $p_{\succ,\gamma}$ be a random consideration set rule. Then for any alternative $a\in A^*$ and $b\in A$:

$$\frac{p\left(a,A\setminus\{b\}\right)}{p\left(a,A\right)}=\frac{\gamma(a)\prod_{d\in A\setminus\{b\}:d\succ a}\left(1-\gamma(d)\right)}{\gamma(a)\prod_{d\in A:d\succ a}\left(1-\gamma(d)\right)}=\begin{cases} \frac{1}{1-\gamma(b)} & \text{if } b\succ a \\ 1 & \text{if } a\succ b \end{cases}$$

so that $\frac{p(a,A\setminus\{b\})}{p(a,A)} \geq 1$.

A2 (Asymmetry). For all $a, b \in A$: $\frac{p(a,A\setminus\{b\})}{p(a,A)} > 1 \Leftrightarrow \frac{p(b,A\setminus\{a\})}{p(b,A)} \neq 1$.

Let $p_{\succ,\gamma}$ be a random consideration set rule. Based on the arguments in the proof of the necessity of **A1** it follows that $p(a, A) = p(a, A\setminus\{b\})$ iff $a \succ b$, so that by a similar argument $p(b, A) < p(b, A\setminus\{a\})$ and $\frac{p(b,A\setminus\{a\})}{p(b,A)} \neq 1$. For the converse, $\frac{p(b,A\setminus\{a\})}{p(b,A)} \neq 1$ requires $\frac{p(b,A\setminus\{a\})}{p(b,A)} = \frac{1}{1-\gamma(a)}$, so that $a \succ b$ and $p(a, A) = p(a, A\setminus\{b\})$.

A3 (Menu independence). For all $a \in A^* \cap B^*$ and $b \in A \cap B$: $\frac{p(a,A\setminus\{b\})}{p(a,A)} = \frac{p(a,B\setminus\{b\})}{p(a,B)}$.

Let $p_{\succ,\gamma}$ be a random consideration set rule. Then $\frac{p(a,A\setminus\{b\})}{p(a,A)} = \frac{\gamma(a) \prod\limits_{d \in A\setminus\{b\}: d \succ a} (1-\gamma(d))}{\gamma(a) \prod\limits_{d \in A: d \succ a} (1-\gamma(d))} = \begin{cases} \frac{1}{1-\gamma(b)} & \text{if } b \succ a \\ 1 & \text{if } a \succ b \end{cases} = \frac{\gamma(a) \prod\limits_{d \in B\setminus\{b\}: d \succ a} (1-\gamma(d))}{\gamma(a) \prod\limits_{d \in B: d \succ a} (1-\gamma(d))} = \frac{p(a,B\setminus\{b\})}{p(a,B)}$ as desired.

A4 (Neutrality). For all $a, b \in A^*$ and $c \in A$: $\frac{p(a,A\setminus\{c\})}{p(a,A)}, \frac{p(b,A\setminus\{c\})}{p(b,A)} > 1 \Rightarrow \frac{p(a,A\setminus\{c\})}{p(a,A)} = \frac{p(b,A\setminus\{c\})}{p(b,A)}$.

Let $p_{\succ,\gamma}$ be a random consideration set rule. Then on the basis of the argument for the necessity of **A1**, and given that $\frac{p(x,A\setminus\{c\})}{p(x,A)} > 1$ for $x \in \{a, b\}$, we have

$$\frac{p(b, A\setminus\{c\})}{p(b, A)} = \frac{1}{1-\gamma(c)} = \frac{p(b, A\setminus\{c\})}{p(b, A)}$$

as desired.

A5 (Third-alternative consistency). For all $a, b, c \in A$: $\frac{p(a,A\setminus\{c\})}{p(a,A)} < \frac{p(b,A\setminus\{c\})}{p(b,A)} \Rightarrow \frac{p(b,A\setminus\{a\})}{p(b,A)} > \frac{p(a,A\setminus\{b\})}{p(a,A)}$

Let $p_{\succ,\gamma}$ be a random consideration set rule. On the basis of the argument for the necessity of **A1**, the premise of **A5** requires $\frac{p(b,A\setminus\{c\})}{p(b,A)} > 1$, so that $c \succ b$. If $p(a, A) < p(a, A\setminus\{c\})$, then we would have $c \succ a$, and the premise of **A5** would fail, since $\frac{p(a,A\setminus\{c\})}{p(a,A)} = \frac{1}{1-\gamma(a)} = \frac{p(b,A\setminus\{c\})}{p(b,A)}$. Then it must be $p(a, A) = p(a, A\setminus\{c\})$, so that $a \succ c \succ b$ and $\frac{p(b,A\setminus\{a\})}{p(b,A)} = \frac{1}{1-\gamma(a)} > 1 = \frac{p(a,A)}{p(a,A\setminus\{b\})}$ from which the conclusion follows.