Stochastic Choice and Consideration Sets*

Paola Manzini        Marco Mariotti†

This version: February 2012

Abstract

An 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 characterized by three axioms: Regularity, the acyclicity of the revealed preference relation, and a stochastic form of binariness. The model explains menu effects and stochastic intransitivity. The errors committed by the agent (deviations from utility maximization) can never be generated by the Luce model, logit, probit or in general Random Utility Maximisation.

J.E.L. codes: D0.

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

---

*We are grateful to the seminar audience at the UCL “Bounded Rationality, Revealed Preferences and Consumer Demand Workshop” for stimulating discussion, and 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” [23] in which the present paper originates. Any errors are of course our own.

†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 whose behaviour is induced by a preference relation but who, due to imperfect attention, makes random choice errors. We extend the classical revealed preference method to this case of bounded rationality, and study 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.

The vast majority of theoretical models of economic choice assume that behaviour is deterministic. 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 [42], Richter [36]) 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 [30]). 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 [21], Block and Marschak’s [5] and Marschak’s [24], and more recently Gul, Natenzon and Pesendorfer [17] (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 [44]. 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

¹ All the works in this area mentioned in the paper constitute examples for this assertion.
² Goeree [14] quantifies this phenomenon with empirical data.
³ See Wilson [51] 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.
decision-maker has no problem in evaluating 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\(^4\) 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 in this paper 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 [26] - (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 measure the degree of brand awareness for a product, or the (complement of) the willingness of an agent to seriously evaluate a political candidate.

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 4), 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 [52]. See also Shocker, Ben-Akiva, Boccara and Nedungadi [43], Roberts and Lattin ([38],[39]) and Roberts and Nedungadi [40].

\(^5\)For example van Nierop et al [35] 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 [26] is especially relevant for this paper as it is the first to study the problem of 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 other two-stage deterministic models of choice (e.g. our own "shortlisting" method [22]), 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.\(^6\) 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 \(a\). Once the preference is thus retrieved, it is easy to identify the attention parameters (as explained in section 3).

The main result of the paper is a characterization of the model by means of three axioms (Theorem 1). One is a standard monotonicity requirement on the probability of choices in the size of the menu (Regularity). The other two axioms generalize properties of rational deterministic choice. Acyclicity is an analog of SARP and simply demands that the revealed preference relation should have no cycles. Finally, Stochastic binariness requires that the effect of removing an alternative \(b\) from a menu on the probability of choice of an alternative \(a\) should be measured only by the strength of \(b\) versus \(a\) in pairwise contests (i.e. the probability of \(b\) being chosen over \(a\)).

The appeal of this characterization is its simplicity. Contrast it, for example, with the case of one of the leading models of random choice, the Random Utility maximisation (RUM) model. RUM assumes that the agent maximises without error a utility picked at random from a set. While popular in its econometrics versions such as logit and probit, the hypothesis of RUM is difficult to test, because the Block-Marschak-Falmagne conditions ([5], [12]) (or equivalently the Axiom of stochastic revealed preference, McFadden and Richter [31]) that characterize it are complex recursive condi-

\(^6\)Tyson [49] has clarified the general structure of two-stage models of choice.
Our model explains plausible types of choice mistakes that cannot be captured by other classical models of random choice such as RUM or the Luce model (or, equivalently the multinomial logit model, McFadden [29]). In Luce’s [21] model

\[ p(a, A) = \frac{v(a)}{\sum_{b \in A} v(b)}, \]

for some strictly positive utility function \( v \). Theorems 2 and 3 show that any data generated by our model cannot be generated by the Luce or the RUM model (subject to a small restriction). This means in particular (through the results by McFadden [29] and Yellot [53]) that the type of errors that our Homo Considerans makes are not of the types normally studied in econometric works, such as logit, in which an alternative \( a \) generates a ‘random utility stimulus’ \( u(a) + \varepsilon(a) \), where \( \varepsilon(a) \) is an error term, and is chosen over alternative \( b \) if \( u(a) - u(b) > \varepsilon(b) - \varepsilon(a) \). Different error specifications generate different models: as shown by McFadden [29] and Yellot [53], the multinomial logit (or Luce) model is identified by the \( \varepsilon \) being assumed i.i.d. Gumbel (or extreme value type I) distributed random variables.\(^7\)

We make the point that, while computationally convenient, it is hard to provide an intuitive economic or psychological justification of this (or any) specific additive error structure.\(^8\) In these models deviations from utility maximisation follow from the black box of additive exogenous random errors grafted on the utility term.\(^9\) We take instead a specific view the cognitive process of the agent.

\(^7\)I.e. with distribution function \( F \) given by the double exponential

\[ F(x) = \exp \left( -\exp \frac{x}{\sigma} \right) \]

\(^8\)A probit model would follow instead by assuming normal distributions. In general, the basic constraint is that larger errors are made with smaller probabilities. Closely related ideas have also found their way in modelling strategic behaviour, for the first time with McKelvey and Palfrey’s ([32], [33]) notion of Quantal Response Equilibrium (QRE) - see Goeree, Holt and Palfrey [15] for an overview.

\(^9\)We recall that the Gumbel distribution function can be seen as the limit distribution function of (a suitable transformation of) the maximum value statistics for a sample of \( N \) i.i.d. random variables, as \( N \) tends to infinity. The statistics needs to be appropriately transformed when taking the limit since, obviously, letting \( G \) be the common distribution function of the random variables and \( G^M \) the distrib-
In the latter part of the paper (section 5) we study at some length other behavioural phenomena warranted by a random consideration set model as opposed to other models (such as the recent extension of the Luce model by GNP [17]). We focus first on menu effects intended as violations of Luce’s IIA axiom (Luce [21]). This axiom characterizes Luce’s model and states that the ratio between the choice probabilities of two alternatives, \( p(a, A) / p(b, A) \), is the same across all \( A \)s that contain both alternatives. Our model reproduces in particular 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 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, can be plausible in the example.

Our model also explains phenomena of (stochastic) intransitivity that are empirically relevant. In general, we make the point that it is hard to infer the a priori plausibility of a particular model by the axioms that characterize it: for example our model does satisfy a menu independence condition that appears to be as appealing as Luce’s IIA, but is violated by Luce’s model. We advocate the use of axioms mainly as bridges to choice data, while the a priori plausibility of the model should stem from the psychological mechanism that it describes.

2 Random choice rules

There is a finite set of alternatives \( X \) with at least three elements, and a domain \( D \) of nonempty subsets (the menus) of \( X \). We assume that \( |A| \geq 2 \) for all \( A \in D \). We wish to allow the agent to not pick any element from a menu, so we also assume the existence of a default alternative \( a^* \) (e.g. walking away from the shop, abstaining from voting, abstaining from voting,

\[ \lim_{N \to \infty} G(x) = 0 \] for any \( x \) unless \( G(x) = 1 \).

The exclusion of singleton menus is not necessary for the characterisation, but it seems important to show that a theory based on attention does not depend on choices from singletons.
exceeding the time limit for a move in a game of chess). Denote \( X^* = X \cup \{ a^* \} \) and \( A^* = A \cup \{ a^* \} \) for all \( A \in D \).

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

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 \( A \).

In our model, a random choice rule is defined 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 \( a^* \). The membership of \( C(A) \) for the alternatives in \( A \) is probabilistic. For all \( A \in D \), each alternative \( a \) has a probability \( \gamma(a) \) of being in \( C(A) \). Formally:

**Definition 2** A random consideration set rule is an r.c.r. \( p_{\succ, \gamma} \) for which there exists a pair \( (\succ, \gamma) \), where \( \succ \) is a strict linear order on \( X \) and \( \gamma \) is a map \( \gamma : X \to (0, 1) \), such that

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

Note that, since the product only runs on the number of alternatives better than a given alternative, which is finite, the formula for a random consideration set rule remains well-defined also on countably infinite choice sets without any additional parameter restriction.

### 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

\[^{11}\text{For a recent work on allowing 'not choosing' in the deterministic case, see Gerasimou [13]. Earlier work is Clark [7].}\]

7
preference ordering of rational deterministic choice to stochastic choices is (see GNP [17]) to declare \( a \succ b \) 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 quality. 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,\(^{12}\) 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, \( p(a, A) = p(a, A \setminus \{b\}) \). 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 \( p(b, A) > p(b, A \setminus \{a\}) \). Together, \( p(a, A) = p(a, A \setminus \{b\}) \) and \( p(b, A) > p(b, A \setminus \{a\}) \) constitute the revealed preference relation \( \succ \) of our model. We will show that this relation is revealed uniquely.

Next, observe that, once the preference \( \succ \) is known, the attention paid to an alternative \( a \) is revealed directly in any menu \( A \) in which \( a \) is the \( \succ \)–best alternative, provided that \( a \) is not the overall \( \succ \)–worst alternative in \( X \). In fact, if the model is true, the attention paid to an alternative must evidently coincide with its choice probability in such an \( A \). And if there is no menu in which \( a \) is the \( \succ \)–best alternative,\(^{12}\) this feature is called Chernoff’s axiom.
the attention paid to it must coincide with the probability of choosing \( a \) from a menu conditional on not paying attention to any other alternative in the menu. Therefore the attention profile is also fully revealed by choice probabilities, and uniquely so.

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 linear order of the alternatives; and, secondly, that the observed choice probabilities are compatible with \( \succ \) being maximised on consideration set that is stochastically generated by the revealed attention parameters.

### 3.2 Axioms and characterisation theorem

The first axiom is a standard monotonicity property satisfied by virtually all main families of random choice rules.

**Regularity.** For all \( A, B \in \mathcal{D} \) such that \( a, b \in B \subseteq A, p(a, A) \leq p(a, B) \).

For the next axiom we introduce a relation \( R \) following the discussion in the previous section. Say that \( aRb \) if and only if there exists \( A \in \mathcal{D} \) such that \( p(a, A) = p(a, A \setminus \{b\}) \) or \( p(b, A \setminus \{a\}) > p(b, A) \).

**Acyclicity.** \( R \) on \( X \) is acyclic, i.e. there exist no \( a^1, ..., a^n \in X \) such that \( a^iRa^{i+1}, i = 1, ..., n - 1, \) and \( a^nRa^1 \).\(^\text{13}\)

Once \( R \) is interpreted as a revealed preference relation, Acyclicity is exactly analogous to SARP in deterministic revealed preference theory.

Our final axiom is:

**Stochastic binariness.** For all \( A \in \mathcal{D} \) and \( a, b \in A \), if \( p(a, A \setminus \{b\}) > p(a, A) \) then

\[
\frac{p(a, A)}{p(a, A \setminus \{b\})} = 1 - p(b, \{a, b\})
\]

\(^{13}\)The case \( n = 2 \) is included so acyclicity implies asymmetry.
To understand the axiom, note first that it generalises a property that holds for rational deterministic choices (zero-one probabilities): in that case, \( p(a, A \setminus \{b\}) > p(a, A) \) means that \( a \) is chosen from \( A \setminus \{b\} \) but is rejected from \( A \), that is the greater probability is one and the smaller probability is zero. And \( 1 - p(b, \{a, b\}) \) is zero or one according to whether or not \( b \) is chosen over \( a \) in pairwise comparison. The axiom in that case says that \( a \) can become chosen when \( b \) is removed from \( A \), and is otherwise rejected from \( A \), only if \( b \) is picked over \( a \) in a pairwise contest. In the stochastic case, one continues to interpret the fact that removing alternative \( b \) increases the probability that alternative \( a \) is chosen in set \( A \) as revealing that \( b \) is ‘preferred’ to \( a \). The axiom in essence says two things: first, that the effect on the choice probability of \( a \) when removing an alternative \( b \) is independent of the other alternatives in the set; and second, that this effect is entirely measured by the ‘strength’ \( p(b, \{a, b\}) \) that \( b \) has in pairwise comparisons against \( a \).

**Theorem 1** Let \( D = \{A \subseteq X : 2 \leq |A|\} \). A random choice rule \( p \) such that \( p(a, A) \in (0, 1) \) for all \( A \in D \) and for all \( a \in A \), satisfies Regularity, Acyclicity and Stochastic binariness if and only if it is a consideration set rule \( p_\succ, \gamma \). Moreover, both the quality ranking \( \succ \) and the attention profile \( \gamma \) are unique, that is, for any random choice rule \( p_{\succ', \gamma'} \) such that \( p_{\succ', \gamma'} = p \) we have \((\succ', \gamma') = (\succ, \gamma)\).

Although the proof is lengthy, thus relegated to the appendix, the logic behind the sufficiency part is easy. Under Regularity the revealed preference \( R \) defined above is complete (if \( p(a, A) = p(a, A \setminus \{b\}) \), then \( aRb \), while if \( p(a, A) \neq p(a, A \setminus \{b\}) \), then by regularity it can only be \( p(a, A) < p(a, A \setminus \{b\}) \), so that \( bRa \)). Adding Acyclicity makes of \( R \) a linear order, that we take as our preference ranking \( \succ \). Once we have the preference ranking, we construct the attention parameter for each alternative by setting it equal to the (observed) choice probability of the alternative in the largest set in which it is \( \succ \) – maximal, as in such a set the alternative is chosen for sure conditional on having been considered. The rest of the proof goes to show that this construction retrieves the observed random choice \( p \), and does so uniquely. All the three properties that characterise the random consideration set choice rule are needed for the result, as the following examples show.
Example 1 (Fails only Acyclicity). This example is discussed in Tversky [48]. The universal set of alternatives consists of a week-end break, which can be either in Rome, or in Paris, and either with or without the addition of 1 dollar (denoted respectively by \( a, a^+ \), \( b \) and \( b^+ \)), that is \( X = \{ a, a^+, b, b^+ \} \). Then letting \( i, j \in \{ a, b \} \) with \( i \neq j \) it is reasonable to expect that the presence of an extra dollar won’t tip the choice when comparing two different locations, that is

\[
p(i^+, X) = p(j^+, X) = \frac{1}{2} \\
p(i^+, \{i^+, j^+\}) = p(j^+, \{i^+, j^+\}) = p(j, \{i^+, j\}) = p(i, \{i, j\}) = \frac{1}{2} \\
p(i^+, \{i^+, i\}) = 1 \\
p(i^+, \{i, i^+, j^+\}) = p(j, \{i, i^+, j\}) = p(j^+, \{i, i^+, j^+\}) = \frac{1}{2}
\]

From the first and last lines we have both \( a^+ Rb^+ \) and \( b^+ Ra^+ \), violating Acyclicity. Regularity is clearly satisfied. Stochastic binariness only pertains to the case where the smaller set is of the form \( \{i^+, i\} \). We have \( p(i^+, \{i^+, j\}) = \frac{1}{2} = (1 - p(j, \{i^+, j\})) \) and \( \frac{p(i^+, \{i, i^+, j^+\})}{p(i^+, \{i^+, i\})} = \frac{1}{2} = (1 - p(j^+, \{i^+, j^+\})) \) so that Stochastic binariness holds.

Example 2 (Fails only Stochastic binariness) Let \( X = \{ a, b, c \} \) and

\[
p(a, \{a, b, c\}) = p(a, \{a, b\}) = x < p(a, \{a, c\}) = y \\
p(b, \{a, b, c\}) = p(b, \{b, c\}) = p(b, \{a, b\}) = 1 - x \\
p(c, \{a, b, c\}) = 0 < p(c, \{a, c\}) = 1 - y \\
p(c, \{b, c\}) = x
\]

with \( x, y \in (0, 1) \). Here \( bRa \) and not \( (aRb) \) (since \( p(b, \{a, b, c\}) = p(b, \{b, c\}) \) and \( p(a, \{a, c\}) > p(a, \{a, b, c\}) \)), \( aRc \) and not \( (cRa) \) (since \( p(a, \{a, b, c\}) = p(a, \{a, b\}) \) and \( p(c, \{b, c\}) > p(c, \{a, b, c\}) \)) and \( bRc \) while not \( (cRb) \) (since \( p(b, \{a, b, c\}) = p(b, \{a, b\}) \) and \( p(c, \{a, c\}) > p(c, \{a, b, c\}) \)) so that Acyclicity holds. Regularity also clearly holds. However Stochastic binariness fails, since e.g. \( p(c, \{a, b, c\}) < p(c, \{a, c\}) \) but \( \frac{p(c, \{a, b, c\})}{p(c, \{a, c\})} = 0 \neq x = (1 - p(b, \{b, c\})) \).

Example 3 (Fails only Regularity). Let \( x, y \in (0, 1) \) be such that \( x > y(1 - y) \) and
\(x + y < 1\), and define \(\varepsilon = \frac{(1-x-y)y}{1-y}\). Let \(X = \{a, b, c\}\) and

\[
\begin{align*}
p(a, \{a, b, c\}) &= x \\
p(b, A) &= y \text{ for all } A \subseteq X \text{ such that } b \in A \\
p(c, \{a, b, c\}) &= 1 - x - y = p(c, \{b, c\}) \\
p(a, \{a, b\}) &= p(a, \{a, c\}) = x - \varepsilon \\
p(c, \{a, c\}) &= 1 - x - y + \varepsilon
\end{align*}
\]

Note that that \(p(a^*, \{b, c\}) = x, p(a^*, \{a, b\}) = 1 - x - y + \varepsilon = \frac{1-x-y}{1-y}\) and \(p(a^*, \{a, c\}) = y\). The restrictions on \(x\) and \(y\) guarantee that all the above choice probabilities are between 0 and 1.\(^{14}\) Regularity fails, since \(p(a, \{a, b, c\}) = x > p(a, \{a, b\})\). However, \(R\) is acyclic (since the fact that the choice probability of \(b\) is constant across sets implies \(bRa\) and \(bRc\), while \(p(c, \{a, b, c\}) = p(c, \{b, c\})\), so that \(cRa\), and \(p(c, \{a, b, c\}) > p(c, \{a, c\})\), so that not \((cRb)\). Finally, since the probability of choosing a decreases when removing any of the other alternatives from \(X\), we have not \((aRb)\) and not \((aRc)\). Since \(p(c, X) < p(c, X \setminus \{b\})\), Stochastic binariness prescribes \(\frac{p(c, X)}{p(c, X \setminus \{b\})} = 1 - p(b, \{bc\})\); substitution in the terms on the left hand side yield \(\frac{1-x-y}{1-x-y+\varepsilon}\); substituting the expression for \(\varepsilon\) we obtain

\[
\frac{1-x-y}{1-x-y+\frac{1-x-y}{1-y}} = 1 - y = 1 - p(b, \{b, c\})
\]

as desired.

4 Luce and RUM rules

In this section we study the relationship of our model with that of the two classical models of stochastic choice. Luce [21] introduced an r.c.r. for which there exists strictly positive utility numbers associated with the alternative such that the choice probability of an alternative \(a\) is proportional to the utility of \(a\). Formally,

**Definition 3** A Luce choice rule is an r.c.r. \(p_u\) for which there exists a function \(u : X^* \rightarrow \mathbb{R}_{++}\)

\(^{14}\)To verify the claim, consider the following: \(\varepsilon > 0\) if and only if \(x + y < 1\), while \(\varepsilon < 1\) given our restrictions on \(x\) and \(y\) (observe that \(\frac{(1-x-y)y}{1-y} < 1\) if and only if \(-xy < (1-y)^2\), which is always true given that \(x\) and \(y\) are positive ). Moreover, \(x - \varepsilon > 0\) if and only if \(\frac{(1-x-y)y}{1-y} < x \Leftrightarrow y(1-y) < x\), while obviously \(x - \varepsilon < 1\) since \(x < 1\). Finally, \(x - \varepsilon > 0\) implies \(x + y - \varepsilon > 0\) which is equivalent to \(1 - x - y + \varepsilon < 1\). Finally, observe \(1 - x - y + \varepsilon = \frac{1-x-y}{1-y} > 0\).
such that

\[ p_u(a, A) = \frac{u(a)}{\sum_{b \in A^*} u(b)} \text{ for all } A \in \mathcal{D} \text{ and } a \in X \]

Another classical model is the RUM model of Block and Marschak [5], in which a utility function is extracted at random from a set, and the probability that an alternative \(a\) is chosen from \(A\) is the probability with which a utility maximised by \(a\) in \(A\) is extracted. Because the utility functions in a RUM are unique up to monotonic increasing transformations and (as is easy to check\(^\text{15}\)) the probability of a utility tie (given single valuedness) is zero, it is easier to define the RUM model directly in terms of random rankings. Let \(L\) be the set of all linear orderings \(P\) on \(X\), which we identify with the set of all bijections \(r : X \rightarrow \{1, 2, \ldots\}\). For each \(r \in L\) and \(a \in X\), \(r(a) = i\) means that alternative \(a\) is the \(i^{th}\) in the ranking, that is \(\{b \in X : bPa\} = i - 1\). Denote \(\max (r, A) = \arg \min_{a \in A} r(a)\).

**Definition 4** A Random Utility maximisation (RUM) rule is an r.c.r. \(p_\pi\) for which there exists a probability measure \(\pi\) on \(L\) such that

\[ p_\pi(a, A) = \sum_{r \in L} \pi(r) \text{ for all } A \in \mathcal{D} \text{ and } a \in X \]

Can we reproduce the choice probabilities generated by our model with a Luce model? And, more in general, with an additive RUM model like logit and probit? We give negative answers to both questions. Since Block and Marschak [5], McFadden [29] and Yellot [53] have shown that a Luce model is a particular case of an additive RUM model, the first result below is also a consequence of their theorems and our second result below, but it is easy and more transparent to prove it directly without invoking the rather involved indirect arguments.

**Theorem 2** Let \(p_{\succ, \gamma}\) be a random consideration set rule defined on a domain \(\mathcal{D}\) that includes at least a triple and the subset pairs. There is no Luce rule \(p_u\) such that \(p_{\succ, \gamma} = p_u\).

**Proof.** Fix \(\{a, b, c\} \in \mathcal{D}\) and suppose to the contrary that there exists a Luce rule \(p_u\) such that \(p_{\succ, \gamma} = p_u\). Suppose w.l.o.g. that \(a \succ b, c\). Then from the model

\[ p_{\succ, \gamma}(a, \{a, b\}) = \gamma(a) = p_{\succ, \gamma}(a, \{a, b, c\}) \]

\(^{15}\)See Block and Marschack [5], section 3.
Since \( p_{\succ, \gamma} = p_u \),

\[
p_{\succ, \gamma}(a, \{a, b\}) = \frac{u(a)}{u(a) + u(b) + u(a^*)}
\]

\[
p_{\succ, \gamma}(a, \{a, b, c\}) = \frac{u(a)}{u(a) + u(b) + u(c) + u(a^*)}
\]

Together with the previous observation this implies that \( u(c) = 0 \), a contradiction.

For a RUM rule, we use the following shortened notation. Given alternatives \( a, b, c, ... \), \( n \in X \), let \( abc...n \) denote the set of linear orders \( r \in L \) that rank the alternatives in the order in which they are listed, that is \( r \in abc...n \) if and only if \( r(a) < r(b) < r(c) < ... < r(n) \). A RUM rule is standard if, for all \( a, b, c \in X \),

\[
\pi(abc) > 0 \& \pi(cba) > 0 \Rightarrow \pi(cab) > 0
\]

Standardness says that if a ranking of the alternatives \( a, b \) and \( c \) is possible and so is a complete reversal of it, then a partial reversal of it must also be possible. It is a way of saying that if big mistakes are possible then smaller ones are also possible. It is satisfied, for example, by any of the popular additive RUM models for which independent errors have full support on the range of utilities, such as logit and probit.

**Theorem 3** Let \( p_{\succ, \gamma} \) be a random consideration set rule defined on a domain \( D \) that includes at least a triple and all subset pairs. There is no standard RUM rule \( p_\pi \) such that \( p_{\succ, \gamma} = p_\pi \).

**Proof.** We note first a general principle:

**Fact:** for any RUM rule \( p_\pi \) (standard or otherwise) such that \( p_\pi = p_{\succ, \gamma} \), any ranking \( r \) assigned strictly positive probability by \( \pi \) must be such that an alternative can only appear either top or bottom in any set of alternatives consisting of itself and elements of its \( \succ \) -lower contour set: that is, it must be \( \pi(cab) = 0 \) whenever \( a \succ b, c \).

To see that the Fact holds, if it were \( \pi(cab) > 0 \), then \( p_\pi(a, \{a, b\}) > p_\pi(a, \{a, b, c\}) \). But given \( a \succ b, c \), if \( p_{\succ, \gamma} = p_\pi \) then also \( p_\pi(a, \{a, b\}) = \gamma(a) = p_\pi(a, \{a, b, c\}) \), a contradiction.

We now apply the Fact to a three-alternative set \( A = \{a, b, c\} \), with \( a \succ b \succ c \), designating \( d \) the default alternative. By the argument in the previous paragraph, it
is easy to check that the only sets of rankings consistent with a RUM rule for which $p_\pi = p_{\pi, \gamma}$ that may have strictly positive probability are $abcd, abdc, acdb, adbc, adcb, bcda, bdca, cbda, ddba, dcba$.

Since $p_\pi(x, A) > 0$ for all $x \in A^*$, there exists a ranking $r$ (depending on $x$) for which $\pi(r) > 0$ and $x = \max(r, A^*)$ for all $x \in A^*$. So in particular, in view of $p_\pi(a, A) > 0$, at least one of the first six sets of rankings is assigned strictly positive probability by $\pi$. If either $\pi(abcd) > 0$ or $\pi(abdc) > 0$ or $\pi(acbd) > 0$, then since either $\pi(dbca) > 0$ or $\pi(dcba) > 0$ in view of $p(d, A) > 0$, by standardness applied to $abd$ and $dba$ it follows that for at least one set of rankings $xywz$ with $y \in \{a, d\}$ or $w \in \{a, d\}$ it must be $\pi(xywz) > 0$, in contradiction with the Fact. Similarly, if either $\pi(acdb) > 0$ or $\pi(adbc) > 0$ or $\pi(adcb) > 0$, then since either $\pi(bcda) > 0$ or $\pi(bdca) > 0$ in view of $p_\pi(b, A) > 0$, once again by standardness applied to $adb$ and $bda$ we have the contradiction that $\pi(xywz) > 0$ for some set of rankings $xywz$ with $y \in \{a, b\}$ or $w \in \{a, b\}$.

The four-alternative restriction is necessary for the result since any random choice rule is a RUM rule when there are only three alternatives (Block and Marschak [5]).

These results show in particular that the type of choice mistakes stemming from consideration errors is of an entirely different nature from the choice mistakes stemming from shocks (additive or otherwise) to a ‘true’ utility function, and so the model can explain different regularities. We examine this aspect in more detail in the next section.

5 Explaining Menu effects and Stochastic Intransitivity

5.1 Menu effects

Luce [21] states a menu-independence property that characterizes the Luce choice rule:

**Luce’s IIA:** For all $A, B \in D$ for which $a, b \in A \cap B$: $\frac{p(a, A)}{p(b, A)} = \frac{p(a, B)}{p(b, B)}$.

Our model suggests a natural reason why this property might not hold. For a ran-
dom 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. Suppose, for example, that \( a \succ c \succ b \) and that \( a, b, c \in A \) and \( B = A \setminus \{c\} \). Therefore

\[
p_{\succ, \gamma}(a, A) = \frac{\gamma(a)}{\gamma(b) \prod_{d \in A \setminus \{a\}; a \succ d \succ b} (1 - \gamma(d))} > \frac{\gamma(a)}{\gamma(b) \prod_{d \in A \setminus \{c\}; a \succ d \succ b} (1 - \gamma(d))} = \frac{p_{\succ, \gamma}(a, B)}{p_{\succ, \gamma}(b, B)}
\]

violating Luce’s IIA. In fact, for certain configurations of the attention parameter, the addition or elimination of other alternatives can even reverse the ranking between the choice frequencies of two alternatives \( a \) and \( b \):

**Example 4 (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

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

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 it is easy to see, the random consideration set rule does satisfy a form of menu independence that is instead violated by the Luce rule:

**Elimination Independence.** For all \( A, B \in \mathcal{D} \) such that \( a, b \in A \cap B \),

\[
\frac{p(a, B)}{p(a, B \setminus \{b\})} = \frac{p(a, A) \gamma(b) \prod_{d \in A \setminus \{a\}; a \succ d \succ b} (1 - \gamma(d))}{p(a, A \setminus \{b\}) \gamma(b) \prod_{d \in A \setminus \{c\}; a \succ d \succ b} (1 - \gamma(d))}.
\]

Elimination independence states that the effect of removing an alternative on the choice probability of another alternative should not depend on which further alternatives are present in the menu. Since for a random consideration set rule the ratios in
the statement of the axiom are always equal to either $1 - p(b, \{a, b\})$ or one according to whether $b \succ a$ or $a \succ b$, the rule satisfies Elimination independence. But for the Luce rule we have

$$\frac{\sum_{c \in A^*} u(c)}{\sum_{c \in A^* \setminus \{b\}} u(c)} \neq \frac{\sum_{c \in B^*} u(c)}{\sum_{c \in B^* \setminus \{b\}} u(c)} \quad \text{for } A \neq B$$

so that Elimination independence fails. Because Elimination independence seems a priori as reasonable as Luce’s IIA, we take these points as an indication that the a priori validity of a rule ought to be based on its psychological appeal in a given context rather than on the axioms that characterize it.\textsuperscript{16}

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.\textsuperscript{17} The blue bus/red bus example (Debreu [8]) is the standard example,\textsuperscript{18} 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 1. Nevertheless, it is argued (on the premise that the agent does not care about the colour of the bus and so is indifferent between the buses) that adding one of the buses to a pairwise choice set including c will increase the odds of choosing c over either bus, thus violating IIA.

GNP [17] 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

\textsuperscript{16}A random consideration set rule also satisfies the following weakening of Luce’s IIA:

**Odds Monotonicity.** For all $A, B \in \mathcal{D}$ such that $a, b \in A \subset B$ and $aRb$, $\frac{p(a, B)}{p(b, B)} \geq \frac{p(a, A)}{p(b, A)}$.

\textsuperscript{17}By adding a nested structure to choice process (nested logit) or by allowing heteroscedasticity of the choice errors. See e.g. Greene [16] or Agresti [1] for an overview of statistical methods for categorical data. A probit model also allows for menu effects.

\textsuperscript{18}For the sake of precision, Debreu’s original example used music recordings, either two recordings of Beethoven’s eighth symphony played by the same orchestra but with two different directors, or of Debussy’s (only) string quartet. As preferences for Directors can indeed be very strong, we rely instead on McFadden’s [29] reprise of this example using modes of transport.
specific sense) ‘irrelevant’ for choice. In the example each bus is an observational du-
plicate 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.\(^{19}\)

Our model (once straightforwardly extended to account for preference ties), high-
lights however that a new type of menu effect may be plausibly caused by the elimi-
nation 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.

Let

\[
A = \{b, r, t\}
\]
\[
B = \{b, t\}
\]
\[
B' = \{r, t\}
\]

The preference relation is now a weak order \( \succeq \). 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

\[ t \not\succ b \sim r \]

In this case \( r \) and \( b \) are duplicates because \( p_{\succ \gamma}(t, B) = p(t, B') = y \). The duplicate
elimination assumption holds because \( p_{\succ \gamma}(t, A) = y \). Furthermore, we have (assum-
ing that in case both buses are considered then each of them is chosen with probability
\( \frac{1}{2} \)):

\[
\frac{p_{\succ \gamma}(b, A)}{p_{\succ \gamma}(t, A)} = \frac{(1 - y) \left( x^2 + x (1 - x) \right)}{y}
\]
\[
\frac{p_{\succ \gamma}(b, B)}{p_{\succ \gamma}(t, B)} = \frac{(1 - y) x}{y}
\]

\(^{19}\)The general duplicate elimination assumption is more involved but follows the same philosophy.
and therefore

\[ \frac{p_{\succ,\gamma}(b,A)}{p_{\succ,\gamma}(t,A)} = 1 - \frac{x}{2} \]

so that, independently of the attention profile \( \gamma \),

\[ \frac{p_{\succ,\gamma}(b,A)}{p_{\succ,\gamma}(t,A)} < \frac{p_{\succ,\gamma}(b,B)}{p_{\succ,\gamma}(t,B)} \]

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 \( b \) and \( r \) are also duplicates because \( p_{\succ,\gamma}(t,B) = p_{\succ,\gamma}(t,B') = y(1-x) \).

But now the duplicate elimination assumption fails since \( p_{\succ,\gamma}(t,A) = y(1-x)^2 \neq p_{\succ,\gamma}(t,B) \). In addition, we have:

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

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

and therefore

\[ \frac{p_{\succ,\gamma}(b,A)}{p_{\succ,\gamma}(t,A)} = \frac{1 - \frac{x}{2}}{1 - x} \]

so that independently of the attention profile \( \gamma \)

\[ \frac{p_{\succ,\gamma}(b,A)}{p_{\succ,\gamma}(t,A)} > \frac{p_{\succ,\gamma}(b,B)}{p_{\succ,\gamma}(t,B)} \]

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 [17] WAR model, a 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 choice 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.

5.2 Stochastic Intransitivity

Several psychologists, starting from Tversky [46], 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 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

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

In their survey on choice anomalies Rieskamp, Busemeyer and Meller [37] write: “Does human choice behavior obey the principle of strong stochastic transitivity? An overwhelming number of studies suggest otherwise” (p. 646).

Consider the following

**Example 5** Set \( \gamma(a) = \frac{4}{9}, \gamma(b) = \frac{2}{3} \) and \( \gamma(c) = \frac{9}{10} \) with \( a \succ b \succ c \). An interpretation is that alternatives are political candidates. The best candidate \( a \) for a particular voter has low recognition, whereas the worst \( c \) has high recognition, with candidate \( b \) being middling both in
quality and recognition. We have:

\[
p_{c, \gamma} (c, \{a, c\}) = \frac{9}{10} \frac{5}{9} = \frac{1}{2}
\]

\[
p_{b, \gamma} (b, \{b, c\}) = \frac{1}{2}
\]

but also

\[
p_{a, \gamma} (b, \{a, b\}) = \frac{15}{29} = \frac{5}{18} < \frac{1}{2}
\]

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.

### 6 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 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 an r.c.r. \( p_{\gamma, \delta} \) for which there exists a pair \((\gamma, \delta)\), where \(\gamma\) is a linear order on \(X\) and \(\delta\) is a map \(\delta : X \times A \to (0, 1)\), such that

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

**Theorem 4** 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 \( p_{\gamma, \delta} \) such that \( p = p_{\gamma, \delta} \).

**Proof.** Let \( p \) be a random choice rule. Let \( \succ \) be an arbitrary linear order of the alternatives. Define \(\delta\) by setting, for \( A \in \mathcal{D} \) and \( a \in A\):

\[
\delta (a, A) = \frac{p (a, A)}{1 - \sum_{b \in A : b \succ a} p (b, A)}
\]
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

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

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

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

so that

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

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 $b \in \{ c \in A : c \succ a \}$, we have that $1 - \delta (b^+, A)$ is equal to

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

whereas for $1 - \delta (b, A)$ we have

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

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 is a telescoping product (where observe that for the $\succ -$maximal term in $A$, the first in the product on the right hand side of (1), the denominator is equal to 1), and we thus have:

$$\prod_{b \in A: b \succ a} (1 - \delta (b, A)) = 1 - \sum_{b \in A: b \succ a^+} p (b, A) - p (a^+, A)$$

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

22
We conclude that

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

\[
= \frac{p(a, A)}{1 - \sum_{b \in A: b \succ a} p(b, A)} \left(1 - \sum_{b \in A: b \succ a} p(b, A)\right)
\]

\[
= p(a, A)
\]

as desired.

As the proof makes clear, 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 \(\delta\) 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. [35]).

7 Related literature

As explained in the introduction, the most related economic papers conceptually are MNO [26] and Eliaz and Spiegler ([10], [11]). In our model exactly as in their models, an agent who chooses from menu \(A\) maximises a preference relation on 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 trivially vacuous at the empirical level, 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 ([47], [48]) classical Elimination by Aspects (EBA) rule $p_\varepsilon$, which satisfies Regularity, is such that there exists a real valued function $U : 2^X \to \mathbb{R}_+$ such that for all $A \in D, a \in A$:

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

There are random consideration set rules that are not EBA rules. Tversky showed that for any three alternatives $a, b, c$, 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 5 shows that this requirement is not always met by a random consideration set rule.

Recently, GNP [17] 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 $(A \cup B) \cap (C \cup D) = \emptyset$. They also generalise the Luce rule to the Weighted Attribute Rule (WAR) in such a way as to accommodate red bus/blue bus type of violations of Luce’s IIA (see section 5), as well as some well-known choice anomalies. We have seen that a random consideration set rule violates one of the key axioms (duplicate elimination) for a WAR. And the choice frequency reversal Example 4 violates the Independence property above.

Stochastic choice is also the focus of the anticipated choice model characterised in Koida [20]. 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 [28] 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 [50]). 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 [27] and Cheremukhin, Popova and A. Tutino [6] 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 [41] 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.

8 Concluding remarks

In this paper we have studied a model of bounded rationality in which the specific decision procedure, where the decision maker maximises a preference over the subset of the available alternatives that he considers, is psychologically motivated. As other two stage choice procedures, it is useful in explaining observed behaviour, in that it can account for a variety of choice ‘anomalies’, and it has the advantage that its parameters
can be identified uniquely when it holds in spite of its discrete structure.

The natural appeal of a two stage structure with a stochastic first stage extends beyond economics, from psychology to consumer science. In philosophy, it has been taken by some (e.g. William James [19], Daniel Dennett [9], Martin Heisenberg [18]) as a fundamental feature of human choices, and as a solution of the general problem of free will. While these other disciplines rely on a stochastic first stage (either in an econometric specification, as e.g. in [35], or more abstractly as e.g. in Dennet’s [9] ‘Valerian’ model of free will), the approach to accounting for random choices in economics is by and large to restrict directly a priori the error structure, either through axioms (as in Luce [21] or GNP [17]) or through the selection of a particular probability distribution in a black box ‘utility plus error’ model (as in probit and logit models). We have followed the first modeling strategy by focussing on a stochastic first stage while being explicit on the behavioural restrictions of the model in the economic tradition.

Given our characterisation result, one useful aspect of our approach is that it is possible to carry out normative analysis as in deterministic choice theory. This is because choices generated by the cognitive process we have postulated, while containing mistakes with respect to pure preference maximisation, still fully reveal a preference.20

Many questions remain open. Of these, two seem particularly salient. To begin with, we have entirely focussed on an individual decision maker - what are the implications of our version of the consideration set model for market competition, as pioneered in the mentioned works of Eliaz and Spiegler ([10], [11])? Secondly, if attention is to be made dependent on the particular set of options under consideration, how should this menu dependence be restricted? We hope to be able to address these questions in future research.

20Conversely, the general RUM model, when not interpreted as a theory of errors, leaves the observer uncertain as to what utility function to use for normative purposes. General frameworks to study welfare in a bounded rationality context are those of Sugden [45], Bernheim and Rangel [4] and Rubinstein and Salant [41]. A full discussion is in Bernheim [3] and a more limited is in Manzini and Mariotti [7].
References


9 Appendix: Proof of theorem 1

Necessity. Let \( p_{\succ, \gamma} \) be a random consideration set rule. Regularity: For any \( A, B \in \mathcal{D} \) such that \( B \subseteq A \) we have

\[
\{ b \in B : b \succ a \} \subseteq \{ b \in A : b \succ a \} \iff \\
\gamma(a) \prod_{b \in A : b \succ a} (1 - \gamma(b)) \leq \gamma(a) \prod_{b \in B : b \succ a} (1 - \gamma(b))
\]
and so \( p_{\succ,\gamma} (a, A) \leq p_{\succ,\gamma} (a, B) \). **Acyclicity:** Suppose that there exist sets \( A^k \) and alternatives \( a^k, k = 1, 2, \ldots, n \), such that, for all \( k \), \( a^k, a^{k+1} \in A^k \) either

\[
p_{\succ,\gamma} (a^k, A^k) = p_{\succ,\gamma} (a^k, A^k \setminus \{a^{k+1}\})
\]

or

\[
p_{\succ,\gamma} (a^{k+1}, A^k) > p_{\succ,\gamma} (a^{k+1}, A^k \setminus \{a^k\})
\]

The displayed equality is equivalent to

\[
\gamma (a^k) \prod_{a \in A^k : a \succ a^k} (1 - \gamma (a)) = \gamma (a^k) \prod_{a \in A^k \setminus \{a^{k+1}\} : a \succ a^k} (1 - \gamma (a)),
\]

implying that \( a^{k+1} \not\sim a^k \). Since \( \succ \) is a linear order, \( a^k \succ a^{k+1} \). On the other hand the displayed inequality is equivalent to

\[
\gamma (a^{k+1}) \prod_{a \in A^k : a \succ a^{k+1}} (1 - \gamma (a)) > \gamma (a^{k+1}) \prod_{a \in A^k \setminus \{a^k\} : a \succ a^{k+1}} (1 - \gamma (a))
\]

so that again \( a^k \succ a^{k+1} \). Therefore \( a^1 \succ a^2 \succ \ldots \succ a^n \) and thus \( a^1 \succ a^n \). It follows that for all \( A \in \mathcal{D} \) such that \( a^1, a^n \in A \) it must be that

\[
p_{\succ,\gamma} (a^1, A) = \gamma (a^1) \prod_{a \in A : a \succ a^1} (1 - \gamma (a)) = \gamma (a^1) \prod_{a \in A \setminus \{a^n\} : a \succ a^1} (1 - \gamma (a)) = p_{\succ,\gamma} (a^1, A \setminus \{a^n\})
\]

So by the definition of \( R \) we have \( a^1Ra^n \) and \( R \) is acyclic. **Stochastic binariness.** If \( p_{\succ,\gamma} (a, A \setminus \{b\}) > p_{\succ,\gamma} (a, A \setminus \{b\}) \) it must be \( b \succ a \). Therefore we have

\[
\frac{p_{\succ,\gamma} (a, A)}{p_{\succ,\gamma} (a, A \setminus \{b\})} = \frac{\gamma (a) \prod_{c \in A : c \succ a} (1 - \gamma (a))}{\gamma (a) \prod_{c \in A \setminus \{b\} : c \succ a} (1 - \gamma (a))} = (1 - \gamma (b)) = 1 - p_{\succ,\gamma} (b, \{a, b\})
\]

as desired.

**Sufficiency.** We prove a preliminary lemma that clarifies the nature of \( R \) in the presence of Regularity and Acyclicity:
Lemma 1 \textbf{Let } p \textbf{ satisfy Regularity and Acyclicity. Then for all } A \in \mathcal{D} \textbf{ there exists a unique } a \in A \textbf{ such that } aRb \textbf{ for all } b \in A \setminus \{a\}. This is also the unique alternative } a \in A \textbf{ for which } p(a, A) = p(a, B) \textbf{ for all } B \subseteq A \textbf{ such that } a \in B.

\textbf{Proof.} Under the assumptions of the statement, R is complete: to see this, let \textit{not} (aRb), so that, for all } A \in \mathcal{D} \textbf{ such that } a, b \in A, p(a, A) \neq p(a, A \setminus \{b\}) \textbf{ and } p(b, A) \geq p(b, A \setminus \{a\}). By Regularity these two conditions become } p(a, A) < p(a, A \setminus \{b\}) \textbf{ and } p(b, A) = p(b, A \setminus \{a\}), \textbf{ so that } bRa. \textbf{ If } R \textbf{ is complete and acyclic then it is also transitive and asymmetric. So for all } A \in \mathcal{D} \textbf{ there exists a unique } R-\text{maximal element in } A.

Regarding the second part of the statement, fix } A \in \mathcal{D} \textbf{ and the } a \textbf{ which is } R-\text{maximal in } A. \textbf{ Suppose to the contrary, and take a maximal (by set inclusion) } B \subseteq A \textbf{ for which } a \in B \textbf{ and } p(a, A) \neq p(a, B). \textbf{ Since } B \textbf{ is maximal, for all } b \in A \setminus B \textbf{ it must be } p(a, A) = p(a, B \cup \{b\}) \neq p(a, B), \textbf{ and by Regularity this means that } p(a, B \cup \{b\}) < p(a, B). \textbf{ Thus (by the definition of } R \textbf{) } bRa, \textbf{ violating Acyclicity.}

To show uniqueness, note that for any } a' \in A \textbf{ distinct from } a \textbf{ such that } p(a', A) = p(a', A \setminus \{b\}) \textbf{ for all } b \in A \setminus \{a'\}, \textbf{ we have (setting } b = a) a'Ra, \textbf{ violating Acyclicity.} \ 

Define } > \textbf{ and } \gamma \textbf{ as follows. Let } \geq = R \textbf{ (note that the lemma implies immediately that } R \textbf{ is a strict linear order on } X). \textbf{ For all } a \in X \textbf{ let } L(a) = \{b \in X : a > b\} \textbf{ and let}

\begin{align*}
\gamma(a) & = p(a, \{a\} \cup L(a)) \textbf{ if } L(a) \neq \emptyset \\
\gamma(a) & = \prod_{b \in X \setminus \{a\}} p(a, X) (1 - \gamma(b)) \textbf{ if } L(a) = \emptyset
\end{align*}

(2) (3)

(The special treatment for the case } L(a) = \emptyset \textbf{ is needed since } \{a\} \cup L(a) \textbf{ is not well defined as a menu in this case). \textbf{ The resulting random consideration set rule is, for all } A \in \mathcal{D}, \textbf{ for all } a \in X:}

\[ p_{\gamma}(a, A) = p(a, \{a\} \cup L(a)) \prod_{b \in A : b > a} (1 - p(b, \{b\} \cup L(b))) \]

\textbf{To check that this is well defined, it is immediate that } \gamma(a) \in (0, 1) \textbf{ whenever } L(a) \neq \emptyset. \textbf{ Consider now the case } L(a) = \emptyset, \textbf{ so that}

\[ \gamma(a) = \prod_{b \in X \setminus \{a\}} p(a, X) (1 - p(b, \{b\} \cup L(b))) = \prod_{b \in X \setminus \{a\}} p(a, X) (1 - p(b, \{a, b\})) \]

33
where the last equality follows from the lemma and the construction of $\succ$, given that $b \succ a$ for all $b \in X \setminus \{a\}$. By Regularity and by the fact that $L(a) = \emptyset$, $p(a, A) > p(a, A \setminus \{b\})$ for all $b \in A \setminus \{a\}$ for all $A \in D$ for which $A \setminus \{b\} \in D$ (otherwise if $p(a, A) = p(a, A \setminus \{b\})$ we have the contradiction $aRb$), so that by Stochastic binariness $p(a, A) = p(a, A \setminus \{b\}) (1 - p(b, \{a, b\}))$ for all $b \in A \setminus \{a\}$. Successive substitutions of this formula in the expression for $\gamma(a)$ yield

$$\gamma(a) = \frac{p(a, X \setminus \{b\}) (1 - p(b, \{a, b\}))}{\prod_{b' \in X \setminus \{a\}} (1 - p(b', \{a, b'\}))} = \frac{p(a, X \setminus \{b\})}{\prod_{b' \in X \setminus \{a, b\}} (1 - p(b', \{a, b'\}))}$$

for some $c \in X$. But then $\gamma(a) < 1 \Leftrightarrow p(a, \{a, c\}) + p(c, \{a, c\}) < 1$ which holds true by assumption.

Next we show that $p_{\succ, \gamma} = p$. We consider first only the $a \in X$ for which $L(a) \neq \emptyset$.

There are two possibilities:

(i) If $A \subseteq \{a\} \cup L(a)$, then $\{b \in A : b \succ a\} = \emptyset$, and the above simplifies to

$$p_{\succ, \gamma}(a, A) = \gamma(a) = p(a, \{a\} \cup L(a)) = p(a, A)$$

where the last equality follows from lemma 1.

(ii) If $A \not\subseteq \{a\} \cup L(a)$, then $\{b \in A : b \succ a\} \neq \emptyset$. Partition $A$ as $\overline{A}(a) \cup A(a) = A$, where $A(a) = \{a\} \cup (A \cap L(a))$ and $\overline{A}(a) = A \setminus A(a)$. Therefore

$$p_{\succ, \gamma}(a, A) = p(a, \{a\} \cup L(a)) \prod_{b \in A : b \succ a} (1 - p(b, \{b\} \cup L(b))) \quad (4)$$

$$= p(a, \{a\} \cup L(a)) \prod_{b \in \overline{A}(a) : b \succ a} (1 - p(b, \{b\} \cup L(b))) \quad (5)$$

$$= p(a, \overline{A}(a)) \prod_{b \in \overline{A}(a) : b \succ a} (1 - p(b, \{a, b\})) \quad (6)$$

where the equality in the third line follows from lemma 1 and the definitions of $\succ$ and of $L(a)$. We argue by induction on the cardinality of $\overline{A}(a)$. If $|\overline{A}(a)| = 1$, then the above simplifies to

$$p_{\succ, \gamma}(a, A) = p(a, \overline{A}(a)) (1 - p(b, \{a, b\}))$$
Therefore combining the previous two equations

\[ p(a, A) = p(a, \overline{A}(a)) \prod_{b \in \overline{A}(a) : b \succ a} (1 - p(a, \{a, b\})) \]

and it follows that \( p_{\succ \gamma}(a, A) = p(a, A) \).

Suppose now that \( p_{\succ \gamma}(a, A) = (1 - p(c, \{a, c\})) p(a, \overline{A}(a)) \prod_{b \in \overline{A}(a) \setminus \{c\} : b \succ a} (1 - p(a, \{a, b\})) \)

Since \( c \in \overline{A}(a) \) we have (appealing to Regularity) \( p(a, A \setminus \{c\}) > p(a, A) \). Then by Stochastic binariness

\[ p(a, A) = (1 - p(c, \{a, c\})) p(a, A \setminus \{c\}) \]

Moreover by the inductive hypothesis

\[ p(a, A \setminus \{c\}) = p(a, \overline{A}(a)) \prod_{b \in \overline{A}(a) \setminus \{c\} : b \succ a} (1 - p(a, \{a, b\})) \]

Therefore combining the previous two equations

\[ p(a, A) = (1 - p(c, \{a, c\})) p(a, \overline{A}(a)) \prod_{b \in \overline{A}(a) \setminus \{c\} : b \succ a} (1 - p(a, \{a, b\})) \]

and thus \( p_{\succ \gamma}(a, A) = p(a, A) \), as required.

We now deal with the case \( L(a) = \emptyset \). We prove \( p_{\succ \gamma}(a, A) = p(a, A) \) by induction on the cardinality of \( A \). If \( A = X \), then \( p_{\succ \gamma}(a, X) = p(a, X) \) follows immediately from the definition of \( \gamma(a) \). Assume next that \( p_{\succ \gamma}(a, A) = p(a, A) \) for all \( A \in \mathcal{D} \) such that \( a \in A \) and \( |A| = k + 1 \leq |X| \), and consider \( A \in \mathcal{D} \) such that \( a \in A \) and \( |A| = k \). Then

\[ p_{\succ \gamma}(a, A) = \gamma(a) \prod_{b \in A \setminus \{a\}} (1 - \gamma(b)) = \frac{p(a, X)}{\prod_{b \in X \setminus \{a\}} (1 - \gamma(b))} \prod_{b \in A \setminus \{a\}} (1 - \gamma(b)) \]
Take some \( c \in X \setminus A \) and let \( B = A \cup \{c\} \). We have:

\[
p_{\succ \gamma} (a, A) = \gamma (a) \prod_{b \in A \setminus \{a\}} (1 - \gamma (b)) \frac{1 - \gamma (c)}{1 - \gamma (c)} = \gamma (a) \prod_{b \in B \setminus \{a\}} (1 - \gamma (b)) \frac{1}{1 - \gamma (c)} = \frac{p_{\succ \gamma} (a, B)}{1 - \gamma (c)} = \frac{p (a, B)}{1 - \gamma (c)}
\]

where \( p_{\succ \gamma} (a, B) = p (a, B) \) follows by the inductive hypothesis. By Stochastic binariness it must be

\[
\frac{p (a, B)}{p (a, A)} = 1 - p (c, \{a, c\})
\]

while by lemma 1 and the definition of \( \gamma (c) \) for the case \( L (c) \neq \emptyset \) we have that \( p (c, \{a, c\}) = \gamma (c) \), so that

\[
p (a, A) = \frac{p (a, B)}{1 - \gamma (c)} = p_{\succ \gamma} (a, A)
\]
as desired.

To conclude we show that \( \succ \) and \( \gamma \) 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 \( A = \{a\} \cup \{c \in X : a \succ c\} \) such 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. Finally by the definition of an r.c.r. it must be \( \gamma (a) = p (a, \{a\} \cup L (a)) = \gamma' (a) \) for all \( a \in X \) for which \( L (a) \neq \emptyset \), so that \( \gamma \) is defined uniquely for all such \( a \). Similarly, \( \gamma (a) (1 - \gamma (b)) = p (a, \{a, b\}) = \gamma' (a) (1 - \gamma (b)) \) in the case \( L (a) = \emptyset \), for all \( b \in X \setminus \{a\} \), so that \( \gamma (a) \) is defined uniquely also in this case.

\[\boxend\]