Dual Random Utility Maximisation

Paola Manzini*

University of Sussex and IZA

Marco Mariotti†

Queen Mary University of London

This version: May 2018

Abstract

Many prominent regularities of stochastic choice, such as the attraction, similarity and compromise effects, are incompatible with Random Utility Maximisation (RUM) as they violate Monotonicity. We argue that these regularities can be conveniently represented by a variation of RUM in which utility depends on only two states and state probabilities are allowed to depend on the menu. We call this model Dual Random Utility Maximisation (dRUM). dRUM is a parsimonious model that admits violations of Monotonicity. We characterise dRUM in terms of three transparent expansion/contraction conditions. We also characterise the important special case in which state probabilities are constant across menus.

J.E.L. codes: D03, D01

Keywords: Stochastic Choice, Attraction Effect, Similarity Effect.

*School of Business, Management and Economics, University of Sussex, Sussex House, Falmer, Brighton, BN1 9RH, United Kingdom, email:p.manzini@sussex.ac.uk

†Corresponding author. School of Economics and Finance, Queen Mary University of London, Mile End Road, London E1 4NS, UK.

1 Introduction

In Random Utility Maximisation (RUM) choices are determined by the maximisation of a utility that depends on stochastic states. RUM satisfies the property of Monotonicity (the addition of a new alternative cannot increase the probabilities of choice of the existing alternatives). As a consequence, in spite of its popularity both in theory and applications, this model suffers from the weakness that it cannot handle a host of prominent regularities related to violations of Monotonicity.

With a view to expanding the explanatory power of the theory in a tractable way, in this paper we modify RUM in two ways. First, we allow state probabilities to depend on the menu. This move immediately makes the model consistent with the regularities. However, completely unrestricted RUM with menu dependent states is a very permissive model. Our second assumption, which substantially disciplines the theory while maintaining the crucial relaxation of Monotonicity, is that the number of states is exactly two. This particular restriction is motivated -as we detail below- by the 'binariness' intrinsic both in the regularities and in several other cognitive mechanisms or population assumptions that have been examined in the literature. We call this model dual RUM (or dRUM in short).

Our main contribution is a characterisation of the behaviour of a dual random utility maximiser with a small set of conditions. These conditions are transparent restrictions on behaviour, and they are structurally similar to standard expansion/contraction axioms in revealed preference theory - a form of characterisation that at present the RUM model lacks. As an intermediate step, we also characterise the special case of a menu-independent dual random utility maximiser. We argue that, while limited as a model of individual decision making due to the reasons just explained, this special case of RUM may be of independent interest for certain types of population data.

Several types of behaviour that have been perceived as 'anomalies' for the utility maximisation model are binary in nature. For example, the 'similarity' and 'attraction' effects have received much attention in the psychology and behavioural economics literature. These effects display a dual form of behaviour, with the switch from one to the other being triggered by the presence or absence of a 'decoy'. We can model such effects naturally as dual RUM. Analogous considerations apply for the 'compromise effect', according to which the frequency of choice of an alternative jumps upwards when its position switches from extreme to intermediate in the space of characteristics. We discuss these anomalies further in section 4.

While it perhaps involves a different cognitive mechanism, a formally similar dual structure features in the well-rehearsed 'frog legs' thought experiment (Luce and Raiffa's [27]). Here the logic is that the presence of a specific item a^* (frog legs in the example) in a menu triggers the maximisation of a different preference order because a^* conveys information on the nature of the available alternatives. Dual RUM generalises this idea to a probabilistic context, avoiding the extreme assumption that one item is chosen for sure depending on the availability of a^*.

An attractive aspect of dual RUM is that it captures many other scenarios of interest. Here are a few examples:

- *Dual-self processes.* The decision maker may be either in a 'cool' state, in which a long-run utility is maximised, or in a 'hot' state, in which a myopic self, subject to short-run impulses (such as temptation), takes control. Indeed, the menu-independent version of dRUM appears in the contracting model by Eliaz and Spiegler [13], and the menu-dependent version describes the implicit second-stage choices of the dual-self model characterised by Chatterjee and Krishna [5] using preferences *over* menus. Correspondingly, our results can be seen as a direct characterisation of this model in terms of choice *from* menus.[1]

- *Household decisions.* Consider decision units composed of two agents, notably a household, for which the observer is uncertain about the exact identity of the decision maker. Household purchases can be observed through standard consumption data, but typically it is not known which of the two partners made the actual purchase decision on any given occasion. Dual RUM constitutes the basis for a 'random

[1]Eliaz and Spiegler [13] do not focus on characterisation but rather explore the rich implications of this probabilistic notion of naïveté in a contract-theoretic framework, where a principal chooses the optimal menu of contracs to offer to partially naive agents.

dictatorship' model of household decisions that could complement the 'collective' model (Chiappori [8]; Cherchye, De Rock and Vermeulen [6]).[2]

- *Normativity vs selfishness.* Many situations of choice present a conflict between a 'normative' mode and a 'selfish' mode of decision. This conflict is both introspectively intuitive and consistent with experimental observations, for example in dictator games (Frolich, Oppenheimer and Moore [18]). This dichotomy can refer both to a typology of individuals and to a single conflicted individual.

In all of these examples, the possible menu dependence of the ranking probabilities is a compelling feature. In the dual-self interpretation, if the duality of the self is due to temptation, then the presence of tempting alternatives may increase the probability that the short-term self is in control - and possibly the more so the more numerous the tempting alternatives. In the household interpretation, husband and wife may have different 'spheres of control', so that menus containing certain items are more likely to be under the control of a specific partner. Finally, in the individual interpretation of the normative versus selfish example, the probability of triggering a normative mode may hinge on the discrepancy between the selfishly optimal choice and the best normative choice that is available, along the lines of the 'warm glow' theory of Cherepanov, Feddersen and Sandroni [7]. On the other hand, if the normative/selfish distinction referred to types in a population, menu-independent state probabilities would be an appropriate assumption, as we detail in section 3.

When menu dependence is allowed in RUM, the information contained in stochastic choice data pertains only to the certainty, possibility or impossibility of an alternative being chosen. In this sense the properties we consider to characterise dRUM all have a 'modal' nature. One new property we introduce is Modal Impact Consistency. It says that if the addition of an alternative b to a menu A has no modal impact on any other alternative a (i.e. it neither changes the choice probability of a from one to less than

[2] de Clippel and Eliaz [9] propose a dual-self model that can be thought of as a model of intra-household bargaining. Interestingly, this model exhibits the attraction and compromise effects that we study later in the paper.

one, nor from greater than zero to zero) then b has zero probability of being chosen from $A \cup \{b\}$. Observe that the non-modal version of this property is satisfied by standard RUM: if the addition of an alternative b to a menu A does not affect the choice probability of any other alternative a, then b has zero probability of being chosen from $A \cup \{b\}$. The other new property we introduce is Modal Contraction Consistency. It simply says that if b has a modal impact in a menu A then b has a modal impact in any sub-menu of A. Again, the non-modal version of this property is satisfied by RUM. Together with the modal version of Monotonicity, the two properties characterise dual RUM. In sum, we obtain a characterisation of dual RUM, simply by taking the modal versions of three properties that are satisfied by any (menu independent) RUM.

This clean axiomatisation contrasts with the situation for RUM. The behaviour corresponding to RUM is not very well-understood: RUM is equivalent to the satisfaction of a fairly complex set of conditions that illuminate the representation rather than providing a behavioural interpretation (Block and Marschak [3]; Falmagne [16]; Barberá and Pattanaik [2]).[3] For this reason other scholars have focussed on interesting special cases where additional structure is added, leading to clear behavioural characterisations. Beside the classical Luce [26] model and its modern variations (e.g. Gul, Natenzon and Pesendorfer [21]; Echenique and Saito [12]), among the recent contributions we recall in particular Gul and Pesendorfer [20], who assume alternatives to be lotteries and preferences to be von Neumann Morgenstern; Apesteguia, Ballester and Lu [1], who examine, in an abstract context, a family of utility functions that satisfies a single-crossing condition; and Lu and Saito [25], who characterise RUM in a intertemporal choice context.

[3] To elaborate, the Block-Marschak-Falmagne (BMF) conditions require $\sum_{B:A\subseteq B\subseteq X} (-1)^{|B\setminus A|}\, p\,(a,B) \geq 0$ for all menus A and alternative $a \in A$, where $p\,(a,B)$ is the probability of choosing a from menu B. These conditions can be easily interpreted only in terms of the representation itself, i.e. as indicating certain features of the probabilities of the rankings. As noted by Fiorini [17], they come from a re-statement of the RUM model as $p\,(a,A) = \sum_{B:A\subseteq B\subseteq X} q\,(a,B)$, where q is the probability that a is ranked top in B and not in any of the supersets of B. Solving this equation for q in terms of p via Moebius inversion yields the terms of the BMF conditions, which then simply assert the non-negativity of the ranking probabilities q.

2 Preliminaries

Let X be a finite set of $n \geq 2$ alternatives. The nonempty subsets of X are called *menus*. Let $\mathcal{D}$ be the set of all nonempty menus.

A *stochastic 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 $A \in \mathcal{D}$; $p(a, A) = 0$ for all $a \notin A$; and $p(a, A) \in [0,1]$ for all $a \in A$, for all $A \in \mathcal{D}$. The value $p(a, A)$ may be interpreted either as the probability with which an individual agent (on whose behaviour we focus) chooses a from the menu A, or as the fraction of a population choosing a from A.

In Random Utility Maximisation (RUM) a stochastic choice rule is built by assuming that there is probabilistic state space and a state dependent utility which is maximised in the realised state. Provided that the event that two alternatives have the same realised utility has probability zero, RUM can be conveniently described in terms of random rankings (see e.g. Block and Marschak [3]), which we henceforth do. A (strict) *ranking* of X is a bijection $r : \{1, ..., n\} \to X$. Let $\mathcal{R}$ be the set of all rankings. The image of $r \in \mathcal{R}$ is interpreted as describing alternatives in decreasing order of preference: $r(1)$ is the most preferred alternative, $r(2)$ the second most preferred, and so on.

Let $\mathcal{R}(a, A)$ denote the set of rankings for which a is the top alternative in A, that is,

$$\mathcal{R}(a, A) = \left\{ r \in \mathcal{R} : r^{-1}(b) < r^{-1}(a) \Rightarrow b \notin A \right\}.$$

Let μ be a probability distribution on $\mathcal{R}$. A *RUM* is a stochastic choice rule p such that

$$p(a, A) = \mu(\mathcal{R}(a, A))$$

A *dual RUM (dRUM)* is a RUM that uses only two rankings and such that, in addition, the probabilities of the rankings may depend on the menu to which they are applied. That is, a dRUM is a stochastic choice rule p for which there exists a triple $(r_1, r_2, \tilde{\alpha})$, where r_1 and r_2 are rankings on X and $\tilde{\alpha} : 2^X \backslash \varnothing \to (0,1)$ is a function that for each menu A assigns probabilities $\tilde{\alpha}(A)$ to r_1 and $1 - \tilde{\alpha}(A)$ to r_2, such that for all A:

$$p(a, A) = \tilde{\alpha}(A) \mathbf{1}_{\mathcal{R}(a,A)}(r_1) + (1 - \tilde{\alpha}(A)) \mathbf{1}_{\mathcal{R}(a,A)}(r_2)$$

where $\mathbf{1}_K$ denotes the indicator function for a set K.

A *Menu-independent dRUM (idRUM)* is a dRUM for which there exists an α such that $\tilde{\alpha}(A) = \alpha$ for all menus A. An idRUM p is thus identified by a triple (r_1, r_2, α) of two rankings and a number $\alpha \in (0,1)$.

In the previous definition say that $(r_1, r_2, \tilde{\alpha})$ and (r_1, r_2, α) *generate* p. To ease notation, from now on given rankings r_i, $i = 1,2$, we shall write $a \succ_i b$ instead of $r_i^{-1}(a) < r_i^{-1}(b)$.

3 Menu-independent dRUM

It is convenient to study first the menu independent version of the model, as we develop here the key tools for the characterisation of the general case. As we shall see, the latter is tightly linked to the $\alpha = \frac{1}{2}$ case of idRUM. Although it must satisfy Monotonicity, idRUM is nevertheless of some interest since menu independence is likely to hold in many situations where stochastic choices describe random draws from a population rather than individual (or household) level decisions. idRUM in this case represents the variability of choice due to a binary form of *hidden population heterogeneity*. The states correspond to two types of individuals in the population. We have already argued that menu-independence is an appropriate assumption in the population interpretation of the normativeness/selfishness example of the Introduction. Several other interesting examples of this kind of binary heterogeneity have been considered in the literature: the population may be split, in unknown proportions, into high and low cognitive ability, strategic and non-strategic, 'fast and slow' or 'instinctive and contemplative' types (e.g. Caplin and Martin [4]; Gill and Prowse [19]; Proto, Rustichini and Sofianos [34]; Rubinstein [35],[36]).[4]

Say that b *impacts* a in A whenever $p(a, A) > p(a, A \cup \{b\})$ (Manzini and Mariotti [28]). The following four conditions are observable implications of idRUM:

Monotonicity: If $a \in A \subset B$ then $p(a, A) \geq p(a, B)$.

Contraction Consistency: If b impacts a in A and $B \subset A$, then b impacts a in B.

[4]Caplin and Martin [4] are distinctive in that they consider agents who can *choose* whether to make automatic (fast) or considered (slow) choices, depending on an attentional cost.

Impact Consistency: If, for all $a \in A$, b does not impact a in A then $p(b, A \cup \{b\}) = 0$.

Negative Expansion: If $a \in A$ and $p(a, A) < p(a, B) < 1$ then $p(a, A \cup B) = 0$.

The first three conditions are satisfied by any RUM. Monotonicity is standard.

Contraction Consistency is a new condition and says that impact is inherited from menus to sub-menus. It is a stochastic generalisation of a property that holds for deterministic choice functions that maximise a partial ordering. In that case, if b impacts a (i.e. turns a from chosen to unchosen) in a menu A then (1) b is higher than a in the ordering and (2) a is a maximal alternative in A (otherwise it would not be chosen from A and it could not be impacted by b). It thus follows that a continues to be chosen in a subset B of A, and that b continues to impact a in B.

Impact Consistency holds generally for random utility maximisation because if b does not impact any other alternative in a menu, it cannot be a predecessor, in any ranking, of any alternative that is chosen with positive probability (and which therefore is at the top of some ranking). Thus b must be chosen with zero probability when added to the menu. Impact Consistency is redundant for the characterisation below (it is implied by Monotonicity), but we will need to impose explicitly a version of it in the characterisation of dRUM.

Finally, Negative Expansion does not necessarily hold in any RUM, and is a strong form of menu-independence condition. Under the model, different frequencies of choice for a in two menus reveal (given that they are less than unity) that in one menu a was not supported by one group, and in the other menu a was not supported by the other group. Therefore, if the reasons for support are menu-independent, no group supports a when the menus are merged.

The interplay of Monotonicity and Negative Expansion yields the binary structure of the model. If they are satisfied, in any menu there are at most two alternatives receiving positive choice probability, and furthermore the possible probability values do not change across menus.

Lemma 1. *Let p be a stochastic choice rule that satisfies Monotonicity and Negative Expansion. Then for any menu A, if $p(a, A) \in (0, 1)$ for some $a \in A$ there exists $b \in A$ for which $p(b, A) =$*

$1 - p(a, A)$.

Proof: Suppose by contradiction that for some menu A there exist $b_1, ..., b_n \in A$ such that $n > 2$ and $p(b_i, A) > 0$ for all i. Then, for all $i < n$, $p(b_i, \{b_1, ..., b_{n-1}\}) \geq p(b_i, A)$ by Monotonicity, with strict inequality for some i since $p(b_n, A) > 0$. Fixing such an i and noting that $p(b_i, \{b_1, ..., b_{n-1}\}) < 1$ and $p(b_i, A) < 1$, Negative Expansion yields the contradiction $p(b_i, A) = 0$. $\square$

Lemma 2. *Let p be a stochastic choice rule that satisfies Monotonicity and Negative Expansion. Then there exists $\alpha \in (0, 1)$ such that, for any menu A and all $a \in A$, $p(a, A) \in \{0, \alpha, 1 - \alpha, 1\}$.*

Proof: The case $p(a, A) = 1$ for all A for some a is trivial. Then take A for which $p(a, A) = \alpha \in (0, 1)$ for some a. By Lemma 1 there exists exactly one alternative $b \in A$ for which $p(b, A) = 1 - \alpha$. Note that by Monotonicity $p(a, \{a, c\}) \geq \alpha$ for all $c \in A \setminus \{a\}$ and $p(b, \{b, c\}) \geq 1 - \alpha$ for all $c \in A \setminus \{b\}$.

Now suppose by contradiction that there exists a menu B and a $c \in B$ for which $p(c, B) = \beta \notin \{0, \alpha, 1 - \alpha, 1\}$. By Lemma 1 there exists exactly one alternative $d \in B$ for which $p(d, B) = 1 - \beta$.

Consider the menu $A \cup B$. By Monotonicity $p(e, A \cup B) = 0$ for all $e \in (A \cup B) \setminus \{a, b, c, d\}$. Moreover, by Lemma 1 $p(e, A \cup B) = 0$ for some $e \in \{a, b, c, d\}$, and w.l.o.g., let $p(a, A \cup B) = 0$.

If $p(b, A \cup B) > 0$ then by Monotonicity and Negative Expansion $p(b, A \cup B) = 1 - \alpha$ (if it were $p(b, A \cup B) < 1 - \alpha$, then $p(b, A \cup B) \neq p(b, A)$ with $p(b, A \cup B), p(b, A) < 1$, so that Negative Expansion would imply the contradiction $p(b, A \cup B) = 0$). It follows from Lemma 1 that either $p(c, A \cup B) = \alpha$ or $p(d, A \cup B) = \alpha$. In the former case, since $\alpha > 0$ a reasoning analogous to the one followed so far implies that $p(c, A \cup B) = p(c, B) = \beta$, contradicting $\alpha \neq \beta$; and in the latter case we obtain $p(d, A \cup B) = p(d, B) = 1 - \beta$, contradicting $\alpha \neq 1 - \beta$.

The proof is concluded by reaching analogous contradictions starting from $p(c, A \cup B) > 0$ and $p(d, A \cup B) > 0$. $\square$

The main result in this section says that three of the necessary properties we have considered are also sufficient for the representation.

Theorem 1. *A stochastic choice rule is an idRUM if and only if it satisfies Monotonicity, Negative Expansion and Contraction Consistency.*

We have derived from more fundamental properties the fact that in a menu-independent dRUM alternatives can be chosen from any menu with only two fixed non-degenerate choice probabilities α and $1-\alpha$. One may wonder whether a characterisation can be obtained by simply postulating directly that there exists $\alpha \in (0,1)$, such that for all $A \subseteq X$ and $a \in A$, $p(a,A) \in \{0, \alpha, 1-\alpha, 1\}$. This "range" property would certainly not be very satisfactory as an axiom at the *theoretical* level, since it would be very close to the representation, with the same binary flavour, without providing any insights in the choice procedure. Nevertheless, it could serve as a step for a *practical* test of the theory, since its validity should be easily detectable from the data. As it turns out, the range property and Monotonicity are anyway not sufficient to ensure that the data are generated by an idRUM: they must still be disciplined by further across-menu consistency properties. Table 1 presents a stochastic choice rule p that satisfies the range property and Monotonicity (and even Constant Expansion) but is not an idRUM.

	$\{a,b,c,d\}$	$\{a,b,c\}$	$\{a,b,d\}$	$\{a,c,d\}$	$\{b,c,d\}$	$\{a,b\}$	$\{a,c\}$	$\{a,d\}$	$\{b,c\}$	$\{b,d\}$	$\{c,d\}$
a	0	$\frac{2}{3}$	$\frac{1}{3}$	0	–	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	–	–	–
b	0	0	0	–	0	$\frac{1}{3}$	–	–	$\frac{2}{3}$	$\frac{1}{3}$	–
c	$\frac{1}{3}$	$\frac{1}{3}$	–	$\frac{1}{3}$	$\frac{1}{3}$	–	$\frac{1}{3}$	–	$\frac{1}{3}$	–	$\frac{1}{3}$
d	$\frac{2}{3}$	–	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	–	–	$\frac{2}{3}$	–	$\frac{2}{3}$	$\frac{2}{3}$

Table 1 A Stochastic Choice Rule that fails Negative Expansion

That p is not an idRUM follows from Theorem 1 and the fact that p fails Negative Expansion (e.g. $p(a,\{a,b\}) = \frac{2}{3} \neq \frac{1}{3} = p(a,\{a,d\})$ while $p(a,\{a,b,d\}) > 0$).

There is a subtle relation between the case $\alpha = \frac{1}{2}$ of the menu-independent version of our theory and the deterministic "top-and-the-top" (TAT) model studied in Eliaz, Richter and Rubinstein [14]. A TAT is a choice procedure in which the agent uses two ordering and deterministically picks all the alternatives in a menu that are top in at least one of the two orderings. While we are dealing with a stochastic procedure, when the orderings have the same probabilities ($\alpha = \frac{1}{2}$) the resulting probabilities of choice are uninformative in distinguishing the orderings: the only information they give concerns their support (whereas when $\alpha \neq \frac{1}{2}$ the orderings can be told apart as an α-ordering and an $(1-\alpha)$-ordering). So, in this special case the stochastic choice function contains in fact the same information on the rankings as a TAT. Note, however, the important point that of course our axioms do not only characterise the special case. The generic situation they characterise is for $\alpha \neq \frac{1}{2}$. This case has no equivalent in the deterministic procedure, as it contains strictly more information.[5]

4 General dRUM

Compared to a menu independent dRUM, in a general dRUM the specific magnitudes of choice probabilities, which depend on the menu in an unrestricted way, do not carry any information beyond the *possibility,* the *impossibility* and the *certainty* of the event in which an alternative is chosen. In this sense they are 'modal'.[6] In particular dRUM fails Monotonicity. But it is precisely this feature that gives the model its additional descriptive power.

Say that b *modally impacts* a in A if

$$p(a, A) > 0 \text{ and } p(a, A \cup \{b\}) = 0$$
$$\text{or } p(a, A) = 1 \text{ and } p(a, A \cup \{b\}) \in (0, 1)$$

That is, b modally impacts a if adding b transforms the choice of a from possible (including certain) to impossible or from certain to merely possible.

The following properties are the modal versions of the properties seen before, and are clearly necessary for dRUM:

[5]The key property used by Eliaz, Richter and Rubinstein [14] to characterise TAT says that if an a is chosen from two menus A and B and also from $A \cap B$, and if the choice from $A \cap B$ consists of exactly two elements, then a is chosen from $A \cup B$. In addition their axiomatisation also includes a direct assumption on the number of chosen alternatives.

[6]'Mode' and 'modal' are meant here and elsewhere in the logical and not statistical meaning.

Modal Monotonicity: Let $a \in B \subset A$. (i) If $p(a, B) = 0$ then $p(a, A) = 0$. (ii) If $p(a, A) = 1$ then $p(a, B) = 1$.

Modal Contraction Consistency: If b modally impacts a in A and $B \subset A$, then b modally impacts a in B.

Modal Impact Consistency: If b does not modally impact a in A for all $a \in A$ then $p(b, A \cup \{b\}) = 0$.

Modal Negative Expansion: If $a \in A$, $p(a, A) = 0$ and $p(a, B) \in (0, 1)$ then $p(a, A \cup B) = 0$.

Modal Monotonicity says that, after alternatives are removed from a menu, any alternative that was possibly chosen and is still feasible remains possibly chosen (including certainly chosen), and any alternative that was certainly chosen and is still feasible remains certainly chosen. Thus, Modal Monotonicity only excludes certain types of 'extreme' menu dependence, whereby the composition of the menu does not only affect the numerical values of the probabilities of choice, but the very possibility of choice.

Modal Contraction Consistency and Modal Negative Expansion are straightforward modal analogs of the corresponding axioms seen previously. The latter property has very little bite in this context and is implied by the other axioms.

As previously observed, the analog of Modal Impact Consistency holds in the menu-independent model, and indeed in any RUM. Observe that (unlike for the other axioms) the modal version is stronger, as it states its conclusion from a weaker premise.

We will show that the first, third and fourth properties are sufficient as well as necessary. To do in a simple way, we exploit a relationship that exists between dRUMs and the $\alpha = \frac{1}{2}$ case of menu independent dRUMs. One can associate with each p another stochastic choice rule $\hat{p}$ that is an idRUM if and only if p is a dRUM. This trick allows us to rely on the previous results, by 'translating' the properties for that case to the modal context.

As in the previous theorem, we avoid any assumption explicitly relating to the cardinality of the support of p. So we first prove a lemma analogous to Lemma 1. Say that a stochastic choice rule p is *binary* whenever it assigns positive probability to at most two elements in any menu (i.e. for any menu A, if $p(a, A) \in (0, 1)$ for some $a \in A$ then there exists $b \in A$ for which $p(b, A) = 1 - p(a, A)$). The interplay of Modal Monotonicity and Modal Impact Consistency yields binariness.

Lemma 3. *Let p be a stochastic choice rule that satisfies Modal Monotonicity and Modal Impact Consistency. Then p is binary.*

Proof: Suppose by contradiction that for some menu A there exist $b_1, ..., b_n \in A$ such that $n > 2$, $p(b_i, A) > 0$ for all $i = 1, ...n$ and $\sum_{i=1}^{n} p(b_i) = 1$. By Modal Monotonicity we have both $p(b_i, \{b_1, ..., b_{n-1}\}) > 0$ for all $i = 1, ..., n-1$ and $p(b_i, \{b_1, ..., b_n\}) > 0$ for all $i = 1, ..., n$. Then b_n does not modally impact any $b_i \in \{b_1, ..., b_{n-1}\}$, contradicting Modal Impact Consistency. $\square$

Theorem 2. *A stochastic choice rule is a dRUM if and only if it satisfies Modal Monotonicity, Modal Contraction Consistency and Modal Impact Consistency.*

Proof. Necessity is obvious. For sufficiency, associate with any stochastic choice rule p another stochastic choice rule $\hat{p}$ defined as follows. For all menus A and $a \in A$:

$$\hat{p}(a, A) = 1 \Leftrightarrow p(a, A) = 1$$
$$\hat{p}(a, A) = \frac{1}{2} \Leftrightarrow 1 > p(a, A) > 0$$
$$\hat{p}(a, A) = 0 \Leftrightarrow p(a, A) = 0$$

If p satisfies Modal Monotonicity and Modal Impact Consistency, by Lemma 3 it is binary and so $\hat{p}$ always exists and is defined uniquely, but it is not necessarily an idRUM. Whenever $\hat{p}$ is an idRUM, say generated by some $\left(r_1, r_2, \frac{1}{2}\right)$, then p is a dRUM generated by $(r_1, r_2, \tilde{\alpha})$, where $\tilde{\alpha}$ is defined so that for any A such that $p(a, A) > 0$ and $p(b, A) > 0$ for distinct a and b, $\tilde{\alpha}(A) = p(a, A)$ and $1 - \tilde{\alpha}(A) = p(b, B)$ (up to relabeling). Therefore, in view of Theorem 1, for p to be a dRUM it is sufficient that it is binary and that $\hat{p}$ satisfies Monotonicity, Negative Expansion and Contraction Consistency. We now verify that $\hat{p}$ satisfies these axioms whenever p satisfies the axioms in the statement. Since Lemma 3 has shown that the axioms imply binariness, this will prove the theorem.

Step 1. $\hat{p}$ satisfies Monotonicity.

Let $A \subset B$. If $\hat{p}(a, B) = 0$ then Monotonicity cannot be violated. If $0 < \hat{p}(a, B) < 1$ then $\hat{p}(a, B) = \frac{1}{2}$, so that $p(a, B) > 0$ and hence by (i) of Modal Monotonicity $p(a, A) > 0$. Then $\hat{p}(a, A) \in \left\{\frac{1}{2}, 1\right\}$, satisfying Monotonicity. Finally if $\hat{p}(a, B) = 1$ then $p(a, B) = 1$, so that by (ii) of Modal Monotonicity $p(a, A) = 1$, and then $\hat{p}(a, A) = 1$ as desired.

Step 2. $\hat{p}$ satisfies Negative Expansion.

If $\hat{p}(a, A) < \hat{p}(a, B) < 1$ it must be $\hat{p}(a, A) = 0$ and $\hat{p}(a, B) = \frac{1}{2}$, and therefore $p(a, A) = 0$. If $p(a, A \cup B) > 0$ then (i) of Modal Monotonicity is violated. Therefore $p(a, A \cup B) = 0$ and thus $\hat{p}(a, A \cup B) = 0$ as desired.

Step 3. $\hat{p}$ satisfies Contraction Consistency.

Suppose that b impacts a in A, that is $\hat{p}(a, A) > \hat{p}(a, A \cup \{b\})$. This means that either $\hat{p}(a, A) \in \left\{\frac{1}{2}, 1\right\}$ and $\hat{p}(a, A \cup \{b\}) = 0$, or $\hat{p}(a, A) = 1$ and $\hat{p}(a, A \cup \{b\}) = \frac{1}{2}$. Therefore either $p(a, A) \in (0, 1]$ and $p(a, A \cup \{b\}) = 0$, or $p(a, A) = 1$ and $p(a, A \cup \{b\}) \in (0, 1)$. Let $B \subset A$. Then by Modal Contraction Consistency either $p(a, B) > 0$ and $p(a, B \cup \{b\}) = 0$, or $p(a, B) = 1$ and $p(a, B \cup \{b\}) \in (0, 1)$. This means that either $\hat{p}(a, B) \in \left\{\frac{1}{2}, 1\right\}$ and $\hat{p}(a, B \cup \{b\}) = 0$, or $\hat{p}(a, B) = 1$ and $\hat{p}(a, B \cup \{b\}) = \frac{1}{2}$, so that in either case b impacts a in B. ■

Aside from testing purposes, our characterisation, like all revealed preference style characterisations helps understand the theory by giving a succinct but exhaustive picture of it in terms of constraints on behaviour (as opposed to on the cognitive process, encapsulated in the representation, that is assumed to underlie behaviour). These constraints are about the 'comparative statics' of merging or contraction of menus and are not immediately related to the representation.[7]

As discussed in the Introduction, being able to accommodate violations of Monotonicity is a desirable feature of a stochastic theory of choice. In particular, marketers use a number of strategies to manipulate the attractiveness or otherwise of alternatives. The

[7]An empirical test using our axioms would of course need to treat statistically the assertions that probabilities of choice are equal or different to certain values. Exactly the same need arises in any other characterisation of stochastic choice models: for example, in the case of the classical Luce/logit model [26], which is characterised by a single axiom asserting the equality between two probability ratios. See McCausland and Marley [30] and McCausland, Davis-Stober, Marley, Park and Brown [31] for a thorough approach to the Bayesian testing of stochastic choice conditions.

attraction effect (also known as the 'asymmetric dominance' effect, see Huber, Payne and Puto [22], Huber and Puto [23]) refers to the fact that the choice frequency of a target alternative t increases when a new decoy alternative d is introduced in a menu, with the property that the d is markedly worse than the target t, while incomparable to a third ('other') alternative, o. This ranking is generally induced by presenting alternatives as described in two desirable attributes/dimensions: while the ranking between t and o in one dimension is reversed in the other, d is Pareto dominated by t but Pareto incomparable to o. The *compromise effect* instead refers to the introduction of a different type of decoy, which has the highest degree of one attribute and the lowest of another in such a way that t is now 'middle ranking'.

However, subsequent research has identified various other strategies to increase t's choice probability: the decoy may be Pareto dominated by both t and o, or may Pareto dominate the target but be unavailable for choice (i.e. 'phantom' decoy), and so on. The compromise and attraction effects are just members of a larger family.

All these effects are easy to accommodate in our setup, by supposing that there are two rankings $\succ_i$ such that $t \succ_1 o \succ_1 d$ and $o \succ_2 t \succ_2 d$, and that the introduction of d increases the probability of ranking $\succ_1$. The advantage of this way of modelling the phenomenon is that the mechanism holds regardless of the type of decoy that is introduced, whether it is a phantom alternative, or one that induces compromise, or an asymmetrically dominated alternative, and so on. The structure of our model is consistent with the fact that the target alternative is made more appealing not by improvements to it, but simply by framing, without taking a stance on the exact causal mechanism at play. For example, in the compromise effect it is hard to tell whether what is *behaviourally* a compromise-seeking attitude really reflects a compromise seeking *psychology*. Indeed, Mochon and Frederick [32] find experimentally that an *order* effect could be a more plausible explanation for the 'compromise effect': the alternative presented as second seems more salient in the choice between three items, regardless of its attributes. Menu-dependent probabilities of two orderings is a catch-all concept that gathers whatever factors, psychological or of other nature, affect choice. The model nevertheless allows the researcher to derive general the comparative statics conclusions (described by the axioms).

We conclude the section by noting that dRUM is also consistent with violations of Weak Stochastic Transitivity (which says that if $p(a, \{a,b\}) \geq \frac{1}{2}$ and $p(b, \{b,c\}) \geq \frac{1}{2}$, then $p(a, \{a,c\}) = \frac{1}{3} \geq \frac{1}{2}$). For example take $p(a, \{a,b\}) = p(a, \{a,b,c\}) = p(b, \{b,c\}) = \frac{2}{3}$, $p(a, \{a,c\}) = \frac{1}{3} = p(b, \{a,b,c\})$. Then p violates Weak Stochastic Transitivity but it is a dRUM generated by $a \succ_1 c \succ_1 b$, $b \succ_2 c \succ_2 a$, $\tilde{\alpha}(\{a,b\}) = \tilde{\alpha}(\{a,b,c\}) = \frac{2}{3}$ and $\tilde{\alpha}(\{b,c\}) = \tilde{\alpha}(\{a,c\}) = \frac{1}{3}$. Recent evidence of violations of Weak Stochastic Transitivity can be found in Li and Loomes [24], who examine preferences over date-outcome pairs. In this domain, a dual type of behaviour may be viewed as natural, since the two dimensions in the objects of choice -time and reward- can attract attention with different probabilities according to the binary comparison that is being made.

5 Identification

To understand the identification issues in the models we have studied, start from the generic case of idRUM in which $\alpha \neq \frac{1}{2}$ (or equivalently, in which $p(A) \neq \frac{1}{2}$ for all A). In this case identification is straightforward. Suppose that p is generated by some (r_1, r_2, α) with $\alpha \neq 1 - \alpha$. Then it is easily checked that r_1 and r_2 must satisfy:

$$a \succ_1 b \Leftrightarrow p(a, \{a,b\}) \in \{\alpha, 1\}$$
$$a \succ_2 b \Leftrightarrow p(a, \{a,b\}) \in \{1-\alpha, 1\}$$

Thus, up to a relabelling, the rankings can be uniquely inferred from any dRUM when $\alpha \neq \frac{1}{2}$.

This uniqueness feature is lost for idRUM when $\alpha = \frac{1}{2}$, and therefore it is also lost for dRUM, which as we have seen is equivalent -as far as the information provided by choices about rankings is concerned- to the $\alpha = \frac{1}{2}$ idRUM case. For example consider the idRUM p generated by $\left(r_1, r_2, \frac{1}{2}\right)$ with

$$\begin{aligned} &a \succ_1 b \succ_1 c \succ_1 d \\ &b \succ_2 a \succ_2 d \succ_2 c \end{aligned} \tag{1}$$

The same p could be alternatively generated by $\left(r_1', r_2', \frac{1}{2}\right)$ with

$$\begin{aligned} a \succ_1' b \succ_1' d \succ_1' c \\ b \succ_2' a \succ_2' c \succ_2' d \end{aligned} \tag{2}$$

On the other hand, if these rankings had fixed probabilities α and $1-\alpha$ with $\alpha \neq \frac{1}{2}$, one could tell the two possibilities apart by observing whether $p(d, \{c,d\}) = p(a, \{a,b\})$ or $p(d, \{c,d\}) = p(b, \{a,b\})$.

However, we are going to argue that the informational disadvantage of general dRUM compared to idRUM with $\alpha \neq \frac{1}{2}$ is less than it may appear at first sight.

Non-uniqueness of the rankings in dRUM stems from the following mechanism. Given the rankings r_1 and r_2, the set of alternatives can always be partitioned as $X = A_1 \cup ... \cup A_n$ such that if $a \in A_k$, $b \in A_l$ and $k < l$, then $a \succ_i b$ for $i = 1, 2$ (if no finer partition exists, take $n = 1$ and $A_1 = X$). Now, if the partition is not trivial, construct two new rankings r_1' and r_2' in the following manner. Fixing an A_k, for all $a, b \in A_k$ set $a \succ_i b$ if and only if $a \succ_j b$, $i \neq j$, while for all other cases r_1' and r_2' coincide with r_1 and r_2 (i.e. r_1' and r_2' swap the sub-rankings within A_k). Then it is clear that the probabilities generated by $\left(r_1', r_2', \frac{1}{2}\right)$ are the same as those generated by $\left(r_1, r_2, \frac{1}{2}\right)$. Applying the process to any subset of cells generates other equivalent representations.

It may seem that if the partition of X in $A_1 \cup ... \cup A_n$ contains many cells, this construction generates a large variety of representations consistent with the same p, yielding a serious identification problem. But in fact any two representations must be rather tightly related, as we now show.

Given a ranking of X and $A \subset X$, the restriction of r to A, denoted r^A, is the map $r^A : \{1, ..., |A|\} \to A$ such that $\left(r^A\right)^{-1}(a) < \left(r^A\right)^{-1}(b) \Leftrightarrow r^{-1}(a) < r^{-1}(b)$ for all $a, b \in A$.

Definition 1. Two pairs of rankings (r_1, r_2) and (s_1, s_2) are equivalent if

$$\left\{ \left(r_1^A\right)^{-1}(a), \left(r_2^A\right)^{-1}(a) \right\} = \left\{ \left(s_1^A\right)^{-1}(a), \left(s_2^A\right)^{-1}(a) \right\}$$

for all $A \subseteq X$ and $a \in A$.

The type of information that is relevant in this definition of equivalence is in statements of the type "In a given menu, for both pairs of rankings, a is placed second by one ranking and fifth by the other ranking", without specifying which of the two rankings, in each pair, places a first and which one places a fifth. All that matters is the *set* of positions, second and fifth.

Observe that the specification "for all A" restricts substantially the extent to which pairs of rankings can be equivalent. However two pairs of rankings might satisfy the condition on some A but not in some subset of A. The two pairs of rankings in (1) and (2) are equivalent. On the other hand take

$$a \succ_1 c \succ_1 b \succ_1 d$$
$$b \succ_2 d \succ_2 a \succ_2 c$$

and

$$a \succ_1' d \succ_1' b \succ_1' c$$
$$b \succ_2' c \succ_2' a \succ_2' d$$

In $A = \{a, b, c, d\}$ the primed and the unprimed rankings place each alternative in the same set of positions: a is first and third in both the primed and the unprimed ranking, b is first and third, and so on. However, the restrictions of the unprimed rankings to the set $B = \{a, c, d\}$ place d first and third in B, whereas the restrictions of the primed rankings place instead d second and third in B. Therefore the pairs of rankings are not equivalent. Note that the idRUMs generated by these rankings with $\alpha = \frac{1}{2}$ are different: e.g. (in obvious notation) $p(a, \{a, d\}) = \frac{1}{2}$ and $p'(a, \{a, d\}) = 1$.

Definition 2. Suppose that $\left(r_1, r_2, \frac{1}{2}\right)$ generates p. Then $\left(s_1, s_2, \frac{1}{2}\right)$ also generates p if and only if (r_1, r_2) and (s_1, s_2) are equivalent.

Proof. Suppose that (r_1, r_2) and (s_1, s_2) are equivalent. Denote p and q the idRUMs generated by $\left(r_1, r_2, \frac{1}{2}\right)$ and $\left(s_1, s_2, \frac{1}{2}\right)$, respectively. If an alternative a is ranked top in a menu A by r_i, then given the equivalence condition a must be ranked top in A by s_1 or by s_2. So if $p(a, A) > 0$, we have $p(a, A) = q(a, A)$, and the adding-up constraint on probabilities implies that also $q(a, A) = 0$ for all a for which $p(a, A) = 0$. This shows that $p = q$.

Conversely, suppose that $\left(r_1, r_2, \frac{1}{2}\right)$ and $\left(s_1, s_2, \frac{1}{2}\right)$ generate the same p. Suppose that there exist A and $a \in A$ for which $\left\{ \left(r_1^A\right)^{-1}(a), \left(r_2^A\right)^{-1}(a) \right\} \neq \left\{ \left(s_1^A\right)^{-1}(a), \left(s_2^A\right)^{-1}(a) \right\}$.

Then there must exist $b \in A$ for which $\left(r_i^A\right)^{-1}(b) < \left(r_i^A\right)^{-1}(a)$ for some i and $\left(s_j^A\right)^{-1}(b) > \left(s_i^A\right)^{-1}(b)$ for $j = 1, 2$. Therefore $\left(r_1, r_2, \frac{1}{2}\right)$ and $\left(s_1, s_2, \frac{1}{2}\right)$ cannot generate the same probabilities on $\{a, b\}$, a contradiction, and we conclude that (r_1, r_2) and (s_1, s_2) are equivalent.
■

In conclusion, the type of information on rankings contained in a dRUM (or in an idRUM with $\alpha = \frac{1}{2}$) is not so different from that contained in the apparently more precise idRUMs with $\alpha \neq \frac{1}{2}$. In contexts where it matters knowing in which specific ranking an alternative appears in a given position (e.g., in the "normative" or in the "temptation" ranking), the identification of the meaning of the two rankings must in any case come from extra-choice information. In the case $\alpha \neq \frac{1}{2}$, even if for example we could identify exactly that the two second-ranked alternatives overall are a and b in the rankings labeled r_1 and r_2, respectively, we still would not be able to tell from choices alone whether a is second in the normative rather than the temptation ranking, just as for general dRUMs. On the other hand, when the two rankings in a dRUM *can* be treated "anonymously" and they are of equal importance, all that matters is the aggregate ranking obtained from the two rankings by means of some ordinal aggregation method - such as the Borda rule - and once again the information contained in a general mdRUM is equivalent to that contained in an idRUM with $\alpha \neq \frac{1}{2}$.

6 Concluding remarks

The model we have studied, in which only two alternatives receive positive probability in each menu, are of course a theoretical idealisation. In practice, an empirical distribution roughly conforming to the theory will appear to be strongly *bimodal*, expressing polarised preferences in society or within an individual. Evidence of bimodal distributions in individual and collective choice is found in a variety of contexts.[8]

[8]E.g. Frolich, Oppenheimer and Moore [18] and Dufwenberg and Muren [11] (choices in a dictator games concentrated on giving nothing or 50/50), Sura, Shmuelib, Bosec and Dubeyc [37] (bimodal distributions in ratings, such as Amazon), Plerou, Gopikrishnan, and Stanley [33] (phase transition to bimodal demand "bulls and bears"- in financial markets), Engelmann and Normann [15] (bimodality on maximum

The following two issues, one theoretical and the other empirical, seem interesting to us for future research:

- Can we obtain appealing behavioural characterisations of random utility maximisation with at most k states, with $2 < k \leq n!$?
- As noted in the introduction, bimodal choice distributions (of which the theory in this paper is a theoretical idealisation) are relatively common empirically. What type of econometric specification is best suited to treat this type of distributions when generated by the model?[9]

References

[1] Apesteguia, Jose, Miguel Angel Ballester and Jay Lu (2016) "Single-crossing Random Utility Models" , mimeo

[2] Salvador Barberá and Prasanta K. Pattanaik (1986) "Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings", *Econometrica* 54: 707-15

[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] Caplin, Andrew, and Daniel Martin (2016), "The Dual-Process Drift Diffusion Model: Evidence from Response Times," *Economic Inquiry*, 54 (2): 1274-1282.

and minimum effort levels in minimum effort games), McClelland, Schulze and Coursey [29] (bimodal beliefs for unlikely events and willingness to insure).

[9] For example, one could graft a (unimodal) noise component on a deterministic utility u_i conditional on each state i. The model would then be a binary mixture model of random utilities $u_i + \varepsilon$, for an appropriate specification of unimodal errors ε and of the dependence of the deterministic state utilities u_i on observable covariates. See for instance Došlá [10] for conditions ensuring that the mixture $g(x) = pf_1(x) + (1-p) f_2(x)$ of two unimodal distributions f_1 and f_2 (e.g. Gumbel or log-normal) is bimodal.

[5] Chatterjee, Kalyan, and R. Vijay Krishna (2009) "A "Dual Self" Representation for Stochastic Temptation" , *American Economic Journal: Microeconomics*, 1(2):148-67.

[6] Cherchye, Laurens, Bram De Rock and Frederic Vermeulen (2007) "The collective model of household consumption: a nonparametric characterization", *Econometrica*, 75: 553-574.

[7] Cherepanov, Vadim, Tim Feddersen and Alvaro Sandroni (2013), "Revealed preferences and aspirations in warm glow theory", *Economic Theory* 54(3): 501-535.

[8] Chiappori, Pierre-Andre. (1988) "Rational household labor supply", *Econometrica*, 56: 63-89.

[9] de Clippel, Geoffroy and Kfir Eliaz (2012) "Reason-based choice: a bargaining rationale for the attraction and compromise effects", *Theoretical Economics* 7: 125–162.

[10] Došlá, Š. (2009). "Conditions for bimodality and multimodality of a mixture of two unimodal densities", *Kybernetika*, 45(2): 279-292.

[11] Dufwenberg, Martin, and Astri Muren (2006) " Generosity, anonymity, gender" , *Journal of Economic Behavior & Organization*, 61(1): 42-49.

[12] Echenique, Federico and Kota Saito (2015) "General Luce Model", Caltech HSS Working Paper 1407.

[13] Eliaz, Kfir and Ran Spiegler (2006) "Contracting with Diversely Naive Agents", *The Review of Economic Studies* 73(3): 689-714

[14] Kfir Eliaz, Michael Richter, and Ariel Rubinstein (2011) "Choosing the two finalists" , *Economic Theory*, 46(2):211–219.

[15] Engelmann, Dirk, and Hans-Theo Normann (2010) "Maximum effort in the minimum-effort game," *Experimental Economics*, 13(3): 249-259.

[16] Falmagne, Jean-Claude (1978). "A Representation Theorem for Finite Random Scale Systems" , *Journal of Mathematical Psychology*, 18 (1): 52–72.

[17] Fiorini, Samuel (2004) "A short proof of a theorem of Falmagne", *Journal of Mathematical Psychology* 48 (1): 80-82.

[18] Frolich, Norman, Joe Oppenheimer and J.Bertrand Moore (2001) "Some doubts about measuring self-interest using dictator experiments: the costs of anonymity", *Journal of Economic Behavior and Organization*, 46: 271-90.

[19] Gill, David and Victoria Prowse (2016) "Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis", forthcoming in *Journal of Political Economy*.

[20] Gul, Faruk and Wolfgang Pesendorfer (2006) "Random Expected Utility", *Econometrica*, 74: 121–146.

[21] Gul, Faruk, Paulo Natenzon and Wolfgang Pesendorfer (2014) "Random Choice as Behavioral Optimization", *Econometrica*, 82(5): 1873-1912.

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

[23] Huber, J., and C. Puto, 1983 "Market boundaries and product choice: Illustrating attraction and substitution effects" , *Journal of Consumer Research*, 10: 31–44.

[24] Li, Zhihua and Graham Loomes, (2017) "Diagnosing the Causes of Intertemporal Preference Reversals", mimeo, Warwick Business School.

[25] Lu, Jay and Kota Saito (2016) "Random Intertemporal Choice", mimeo.

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

[27] Luce, R. Duncan and Howard Raiffa (1957), Games and decisions. Wiley, New York.

[28] Manzini, Paola and Marco Mariotti (2014) "Stochastic Choice and Consideration Sets", *Econometrica*, 82(3):1153–1176.

[29] McClelland, Gary H., William D. Schulze, and Don L. Coursey (1993) “Insurance for low-probability hazards: A bimodal response to unlikely events”, *Journal of risk and uncertainty*, 7(1): 95-116.

[30] McCausland, William J. and A.A.J. Marley (2014) “Bayesian inference and model comparison for random choice structures”, *Journal of Mathematical Psychology* (62–63): 33-46.

[31] McCausland, William J., Clintin Davis-Stober, A. A. J. Marley, Sanghyuk Park and Nicholas Brown (2015) “Testing Random Utility Using Falmagne’s Conditions Directly”, http://apps.olin.wustl.edu/conf/SBIES/Files/pdf/2016/194.pdf

[32] Mochon, Daniel and Shane Frederick (2012) “Interpreting Compromise Effects”, mimeo, Tulane University and Yale University School of Management.

[33] Plerou, Vasiliki, Parameswaran Gopikrishnan and H. Eugene Stanley (2003) “Two-phase behaviour of financial markets”, *Nature*, 421: 130.

[34] Proto, Eugenio, Aldo Rustichini and Andis Sofianos (2014) “Higher Intelligence Groups Have Higher Cooperation Rates in the Repeated Prisoner’s Dilemma”, IZA DP No. 8499.

[35] Rubinstein, Ariel (2007) “Instinctive and Cognitive Reasoning: A Study of Response Times”, *Economic Journal*,117: 1243-1259.

[36] Rubinstein, Ariel (2016) “Typology of Players: Between Instinctive and Contemplative”, *Quarterly Journal of Economics*, 131 (2): 859-890.

[37] Sur, Pragya, Galit Shmuel, Smarajit Bose and Paromita Dubey (2015) “Modeling Bimodal Discrete Data Using Conway-Maxwell-Poisson Mixture Models”, *Journal of Business & Economic Statistics*, 33: 352-365.

Acknowledgments

We are grateful for valuable comments and advice to two anonymous referees, the Editor Marciano Siniscalchi, Sophie Bade, Mark Dean, Kfir Eliaz, Sean Horan, Jay Lu and Yusufcan Masatlioglu, as well as to the audiences at the BRIC workshop, the Barcelona GSE Stochastic Choice workshop, ESEM Geneva, and Royal Holloway University of London. Any errors are our own. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendices

A Independence of the axioms - Theorem 1

The stochastic choice rules in cases B.2 and B.3 are not dRUM since they take on more than two interior values.

B.1: Monotonicity and Negative Expansion, but not Contraction Consistency

See example in Table 7 below.

B.2: Monotonicity and Contraction Consistency, but not Negative Expansion

	$\{a,b,c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$
a	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	–
b	$\frac{1}{3}$	$\frac{1}{3}$	–	$\frac{2}{3}$
c	$\frac{1}{3}$	–	$\frac{1}{2}$	$\frac{1}{3}$

Table 2

B.3: Negative Expansion and Contraction Consistency, but not Monotonicity

	$\{a,b,c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$
a	1	0	$\frac{1}{2}$	–
b	0	$\frac{1}{3}$	–	$\frac{2}{3}$
c	0	–	$\frac{1}{2}$	$\frac{1}{3}$

Table 3

B Independence of the axioms - Theorem 2

C.1: Modal Monotonicity (i) and (ii), Modal Contraction Consistency but not Modal Impact Consistency

The stochastic choice rule in Table 4 cannot be dRUM since in $\{a, b, c\}$ three alternatives are chosen with positive probability, which cannot be all at the top of two rankings.

	$\{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$
a	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	—
b	$\frac{1}{3}$	$\frac{1}{2}$	—	$\frac{1}{2}$
c	$\frac{1}{3}$	—	$\frac{1}{2}$	$\frac{1}{2}$

Table 4

C.2: Modal Monotonicity (i), Modal Impact Consistency, Modal Contraction Consistency but not Modal Monotonicity (ii)

The stochastic choice rule in Table 5 cannot be dRUM since on the one hand $p(a, \{a, b, c\}) = 1$ implies that a is the top alternative in both rankings, while $p(b, \{a, b\}) > 0$ would require b to be above a in at least one ranking.

	$\{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$
a	1	$\frac{1}{2}$	1	—
b	0	$\frac{1}{2}$	—	$\frac{1}{2}$
c	0	—	0	$\frac{1}{2}$

Table 5

C.3: Modal Monotonicity (ii), Modal Impact Consistency, Modal Contraction Consistency but not Modal Monotonicity (i)

The stochastic choice rule in Table 6 cannot be a mdRUM since $p(b, \{a, b\}) = 1$ requires b to be above a in both rankings, while $p(a, \{a, b, c\}) > 0$ would require a to be above b in at least one ranking.

C.4: Modal Monotonicity (i) and (ii), Modal Impact Consistency, but not Modal Contraction Consistency

	$\{a,b,c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$
a	$\frac{1}{2}$	0	1	–
b	$\frac{1}{2}$	1	–	$\frac{1}{2}$
c	0	–	0	$\frac{1}{2}$

Table 6

	$\{a,b,c,d\}$	$\{a,b,c\}$	$\{a,b,d\}$	$\{a,c,d\}$	$\{b,c,d\}$	$\{a,b\}$	$\{a,c\}$	$\{a,d\}$	$\{b,c\}$	$\{b,d\}$	$\{c,d\}$
a	0	$\frac{1}{2}$	$\frac{1}{2}$	0	–	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	–	–	–
b	0	0	0	–	0	$\frac{1}{2}$	–	–	$\frac{1}{2}$	$\frac{1}{2}$	–
c	$\frac{1}{2}$	$\frac{1}{2}$	–	$\frac{1}{2}$	$\frac{1}{2}$	–	$\frac{1}{2}$	–	$\frac{1}{2}$	–	$\frac{1}{2}$
d	$\frac{1}{2}$	–	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	–	–	$\frac{1}{2}$	–	$\frac{1}{2}$	$\frac{1}{2}$

Table 7

Suppose that the example in Table 7 is dRUM. Then $p(a, \{a,b,c,d\}) = 0$ and $p(a, \{a,b,c\}) = \frac{1}{2}$ imply that d is an immediate predecessor of a in one of the two rankings on $\{a,b,c,d\}$. This must remain the case when the rankings are restricted to $\{a,b,d\}$, yet $p(a, \{a,b,d\}) = \frac{1}{2} = p(a, \{a,b\})$ imply that d is not an immediate predecessor of a in $\{a,b,d\}$ in either ranking.

C Proof of Theorem 1

Necessity is straightforward, so we only show sufficiency.[10]

Preliminaries and outline Let p satisfy Monotonicity, Negative Expansion and Contraction Consistency. The proof consists of several blocks. We define an algorithm to construct two rankings r_1 and r_2 explicitly (block 1). Next, we show that the algorithm is well defined (block 2), and finally we show that r_1 and r_2 so constructed retrieve p (block 3).

[10]An alternative indirect proof to the constructive one presented here shows that our conditions imply the Block Marschak Falmagne conditions for a RUM, which in conjuction with the Lemmas implies the result. Such a proof is available from the authors upon request.

1. ***Algorithm to construct the two rankings.***

Let $a, b \in X$ be such that $p(a, X), p(b, X) > 0$ (where possibly $a = b$). Enumerate the elements of the rankings r_1 and r_2 to be constructed by $r_1(i) = x_i$ and $r_2(i) = y_i$ for $i = 1, ...n$. Set $x_1 = a$ and $y_1 = b$.

The rest of the rankings r_1 and r_2 for $i \geq 2$ are defined recursively. Let

$$L_i^1 = X \setminus \bigcup_{j=1}^{i-1} x_j$$

Next, define the set S_i of alternatives that are impacted by x_{i-1} in L_i^1:

$$S_i = \left\{ s \in L_i^1 : p\left(s, L_i^1\right) > p\left(s, L_i^1 \cup \{x_{i-1}\}\right) \right\}$$

where observe that by lemmas 1 and 2 we have $0 < |S_i| \leq 2$.

If $S_i = \{c\}$ for some $c \in X$, then let $x_i = c$.

If $|S_i| = 2$, then let $S_i = \{c, d\}$ with $c \neq d$ and consider two cases:

(1.i) $p\left(c, L_i^1\right) > p\left(c, L_i^1 \cup \{x_j\}\right)$ (resp., $p\left(d, L_i^1\right) > p\left(d, L_i^1 \cup \{x_j\}\right)$) for all $j = 1, ...i-1$ and $p\left(d, L_i^1\right) = p\left(d, L_i^1 \cup \{x_j\}\right)$ (resp., $p\left(c, L_i^1\right) = p\left(c, L_i^1 \cup \{x_j\}\right)$) for some $j \in \{1, ..., i-2\}$ (that is, one alternative is impacted by *all* predecessors x_j in L_i^1 while the other alternative is not).

In this case let $x_i = c$ (resp., $x_i = d$).

(1.ii) $p\left(c, L_i^1\right) > p\left(c, L_i^1 \cup \{x_j\}\right)$ and $p\left(d, L_i^1\right) > p\left(d, L_i^1 \cup \{x_j\}\right)$ for all $j = 1, ...i-1$ (both alternatives are impacted by all predecessors x_j in L_i^1).

In this case, if there is $s \in S_i$ such that $p\left(s, L_i^1\right) = p\left(x_j, L_j^1\right) \neq 1$ for some $j < i$, let $x_i = s$. Otherwise, let $x_i = c$ (where note that by construction $p\left(x_j, L_j^1\right) > 0$).

Finally, set $\mu(r_1) = \alpha$ where $\alpha = \min\left\{p\left(x_j, L_j^1\right), j = 1, ..., n\right\}$ if $\min\left\{p\left(x_j, L_j^1\right), j = 1, ..., n\right\} < 1$. Otherwise, if $\min\left\{p\left(x_j, L_j^1\right), j = 1, ..., n\right\} = 1$, the two rankings are the same and the assignment problem is trivial.**.**

Proceed in an analogous way for the construction of r_2, starting from y_2, that is for all $i = 2, ...n$ define recursively

$$L_i^2 = X \setminus \bigcup_{j=1}^{i-1} y_j$$

$$T_i = \left\{ t \in L_i^2 : p\left(t, L_i^2\right) > p\left(t, L_i^2 \cup \{y_{i-1}\}\right) \right\}$$

and as before $0 < |T_i| \leq 2$.

If $|T_i| = \{e\}$ for some $e \in X$, then let $y_i = e$, while if $|T_i| = 2$, then letting $T_i = \{e, f\}$:

(2.i) $p\left(e, L_i^2\right) > p\left(e, L_i^2 \cup \{y_j\}\right)$ (resp. $p\left(e, L_i^2\right) > p\left(e, L_i^2 \cup \{y_j\}\right)$) for all $j = 1, ...i-1$, and $p\left(f, L_i^2\right) = p\left(f, L_i^2 \cup \{y_j\}\right)$ (resp. $p\left(e, L_i^2\right) = p\left(e, L_i^2 \cup \{y_{i-g}\}\right)$) for some $j \in \{1, ..., i-2\}$.

In this case let $y_i = e$ (resp., f).

(2.ii) $p\left(e, L_i^2\right) > p\left(e, L_i^2 \cup \{y_j\}\right)$ and $p\left(f, L_i^2\right) > p\left(f, L_i^2 \cup \{y_j\}\right)$ for all $j = 1, ...i-1$.

In this case let $y_i = f$ (resp., $y_i = e$) whenever $e \succ_1 f$ (resp., $f \succ_1 e$) (i.e. we require consistency with the construction of the first order). Finally let $\mu(r_2) = 1 - \mu(r_1)$.

2. ***Showing that the algorithm is well-defined***

We show that cases *(1.i)* and *(1.ii)* are exhaustive, which means showing that if $|S_i| = 2$ at least one alternative in S_i is impacted by *all* its predecessors in r_1. We proceed by induction on the index i. If $i = 2$ there is nothing to prove. Now consider the step $i = k+1$. If $|S_{k+1}| = 1$ again there is nothing to prove, so let $|S_{k+1}| = 2$, with $S_{k+1} = \{c, d\}$. By construction $p\left(c, L_{k+1}^1\right) > p\left(c, L_{k+1}^1 \cup \{x_k\}\right)$ and $p\left(d, L_{k+1}^1\right) > p\left(d, L_{k+1}^1 \cup \{x_k\}\right)$, so that by Lemma 1 and Lemma 2 it must be

$$p\left(c, L_{k+1}^1\right) = 1 - p\left(d, L_{k+1}^1\right) > 0 \tag{3}$$

By contradiction, suppose that there exist $u, v \in \{x_1, ...x_{k-1}\}$ for which u does not impact c in L_{k+1}^1 and v does not impact d in L_{k+1}^1, i.e.

$$p\left(c, L_{k+1}^1 \cup \{u\}\right) = p\left(c, L_{k+1}^1\right) \tag{4}$$

and

$$p\left(d, L_{k+1}^1 \cup \{v\}\right) = p\left(d, L_{k+1}^1\right) \tag{5}$$

(note that by construction we have $u, v \neq x_k$). We can rule out the case $u = v$. For suppose $u = v = x_j$ for some $j < k$. Then, recalling (3),

$$p\left(d, L_{k+1}^1 \cup \{x_j\}\right) = p\left(d, L_{k+1}^1\right) = 1 - p\left(c, L_{k+1}^1\right) = p\left(c, L_{k+1}^1 \cup \{x_j\}\right)$$

and thus $p\left(x_j, L^1_{k+1} \cup \{x_j\}\right) = 0$, which contradicts Monotonicity and $p\left(x_j, L^1_j\right) > 0$ with $L^1_{k+1} \cup \{x_j\} \subset L^1_j$.

Since by construction and Monotonicity $p\left(v, L^1_{k+1} \cup \{v\}\right) > 0$, Lemma 1, Lemma 2 and (5) imply

$$p\left(c, L^1_{k+1} \cup \{v\}\right) = 0 \tag{6}$$

Similarly, $p\left(v, L^1_{k+1} \cup \{v\}\right) > 0$ (by construction and Monotonicity), Lemma 1, Lemma 2 and (4) imply

$$p\left(d, L^1_{k+1} \cup \{u\}\right) = 0 \tag{7}$$

(i.e. u impacts d in $L^1_{k+1} \cup \{u\}$ and v impacts c in L^1_{k+1}). Now consider the menu $L^1_{k+1} \cup \{u, v\}$. By (6), (7) and Monotonicity it must be:

$$p\left(c, L^1_{k+1} \cup \{u, v\}\right) = 0 = p\left(d, L^1_{k+1} \cup \{u, v\}\right) \tag{8}$$

It cannot be that $p\left(u, L^1_{k+1} \cup \{u, v\}\right) = 1$, for otherwise Monotonicity would imply $p\left(u, L^1_{k+1} \cup \{u\}\right) = 1$, contradicting $p\left(c, L^1_{k+1} \cup \{u\}\right) = p\left(c, L^1_{k+1}\right) > 0$. Similarly, it cannot be that $p\left(v, \{u, v\} \cup L^1_{k+1}\right) = 1$. Finally, if either $p\left(v, L^1_{k+1} \cup \{u, v\}\right) = 0$ or $p\left(u, L^1_{k+1} \cup \{u, v\}\right) = 0$, then in view of (8) it would have to be $p\left(w, L^1_{k+1} \cup \{u, v\}\right) > 0$ for some $w \in L^1_{k+1}$. But this is impossible since $c \neq w \neq d$ by (8), and then by Monotonicity the contradiction $p\left(w, L^1_{k+1}\right) > 0$ would follow. Therefore it must be

$$p\left(u, L^1_{k+1} \cup \{u, v\}\right) = 1 - p\left(v, L^1_{k+1} \cup \{u, v\}\right)$$

It follows that both

$$p\left(u, L^1_{k+1} \cup \{u\}\right) = p\left(u, L^1_{k+1} \cup \{u, v\}\right) \tag{9}$$

and

$$p\left(v, L^1_{k+1} \cup \{v\}\right) = p\left(v, L^1_{k+1} \cup \{u, v\}\right) \tag{10}$$

i.e. neither does u impact v in $L^1_{k+1} \cup \{v\}$, nor does v impact u in $L^1_{k+1} \cup \{u\}$. Suppose w.l.o.g. that v is a predecessor of u in r_1, and let $u = x_j$. By the inductive hypothesis v impacts u in L^1_j, so that by Contraction Consistency it also impacts u in $L^1_{k+1} \cup \{u\} \subset L^1_j$, i.e.

$$p\left(u, L^1_{k+1} \cup \{u\}\right) > p\left(u, L^1_{k+1} \cup \{u, v\}\right)$$

a contradiction with (9). A symmetric argument applies if u is a predecessor of v using (10).

A straightforward adaptation of the argument above shows that cases *(2.i)* and *(2.ii)* are exhaustive.

3. Showing that the algorithm retrieves the observed choice.

Let p_α be the dRUM generated by (r_1, r_2, α) as constructed above. For any alternative x denote L_x^i its (weak) lower contour set in ranking r_i, that is

$$L_x^i = \{x\} \cup \{s \in X : x \succ_i s\}$$

and note that by construction $p\left(x, L_x^i\right) > 0$. We examine the possible cases of failures of the algorithm in succession.

$\boxed{\textit{(3.i)}: p\,(a, A) = 0 \text{ and } p_\alpha\,(a, A) > 0.}$

Then $a \succ_i a'$ for some i, for all $a' \in A \setminus \{a\}$, hence $A \subseteq L_a^i$, and thus by Monotonicity and $p\left(a, L_a^i\right) > 0$ we have $p\,(a, A) > 0$, a contradiction.

$\boxed{\textit{(3.ii)}: p\,(a, A) = 1 \text{ and } p_\alpha\,(a, A) < 1.}$

Then there exists $b \in A$ such that $b \succ_i a'$ for some i, for all $a' \in A \setminus \{b\}$, hence $A \subseteq L_b^i$, and thus by Monotonicity and $p\left(b, L_a^i\right) > 0$ we have $p\,(b, A) > 0$, a contradiction.

$\boxed{\textit{(3.iii)}: p\,(a, A) = \alpha \text{ and } p_\alpha\,(a, A) = 1.}$

Let $\mathcal{I}\,(p_\alpha)$ be the set of all pairs (a, A) satisfying the conditions of this case. Fix an $(a, A) \in \mathcal{I}\,(p_\alpha)$ that is maximal in $\mathcal{I}\,(p_\alpha)$ in the sense that $(b, B) \in \mathcal{I}\,(p_\alpha) \Rightarrow a \succ_i b$ for some i.

Because $p_\alpha\,(a, A) = 1$ we have $A \subseteq L_a^1 \cap L_a^2$. If $p\left(a, L_a^1 \cap L_a^2\right) = 1$ we have an immediate contradiction, since by Monotonicity $p\,(a, A) = 1$. So it must be $p\left(a, L_a^1 \cap L_a^2\right) \in \{\alpha, \min\{\alpha, 1-\alpha\}\}$. We show that this also leads to a contradiction.

Suppose first that $L_a^1 \setminus L_a^2 \neq \varnothing$, and take $z \in L_a^1 \setminus L_a^2$ such that $a \succ_1 z$ and $z \succ_2 a$. By construction this implies $p\left(a, L_a^2\right) > p\left(a, L_a^2 \cup \{z\}\right)$, so that by Contraction Consistency $p\left(a, L_a^1 \cap L_a^2\right) > p\left(a, \left(L_a^1 \cap L_a^2\right) \cup \{z\}\right)$.

By Monotonicity $p\left(a, \left(L_a^1 \cap L_a^2\right) \cup \{z\}\right) \geq p\left(a, L_a^2\right) > 0$ (where recall $z \in L_a^1$, so that $L_a^1 \cup \{z\} = L_a^1$). Since $p\left(a, L_a^1 \cap L_a^2\right) = 1$ would lead to a contradiction, the only possibility is that $p\left(a, L_a^1 \cap L_a^2\right) = 1 - \alpha$ and $p\left(a, \left(L_a^1 \cap L_a^2\right) \cup \{z\}\right) = \alpha$ with $1 - \alpha > \alpha$ (if $1 - \alpha =$

$\alpha = \frac{1}{2}$, then it would have to be $p\left(a, L_a^1 \cap L_a^2\right) = 1$, a contradiction). Since however $p(a, A) = \alpha$, coupled with $A \subseteq L_a^1 \cap L_a^2$ this would contradict Monotonicity. A symmetric argument applies if $L_a^2 \setminus L_a^1 \neq \varnothing$.

Finally, consider the case $L_a^1 = L_a^2 = L_a$ for some $L_a \subset X$. Then $X \setminus L_a^1 = X \setminus L_a^2$. Define $U_a = X \setminus L_a$ (where observe that $u \in U_a \Rightarrow u \succ_i a$, $i = 1, 2$). Observe that $U_a \neq \varnothing$, for otherwise given the construction it could not be $p(a, A) < 1$ and $p_\alpha(a, A) = 1$. Since Monotonicity and $A \subseteq L_a^1 \cap L_a^2$ require $p\left(a, L_a^1 \cap L_a^2\right) < 1$ to avoid a contradiction with $p(a, A) = \alpha$, there exists a $w \in \left(L_a^1 \cap L_a^2\right)$ for which $p\left(w, L_a^1 \cap L_a^2\right) > 0$.

Let $x_1, ..., x_m$ be the predecessors of a in the ranking r_1 and let $y_1, ..., y_m$ be the predecessors of a in the ranking r_2 (where obviously $\{x_1, ..., x_m\} = U_a = \{y_1, ..., y_m\}$). So $a = x_{m+1} = y_{m+1}$ (note that since $L_a^1 = L_a^2$, a must have the same position in both rankings).

By construction x_m and y_m impact a in L_a, that is $p(a, L_a) > p(a, L_a \cup \{x_m\})$ and $p(a, L_a) > p(a, L_a \cup \{y_m\})$.

We claim that there exists a $z \in U_a$ that does not impact w in L_a, that is $p(w, L_a) = p(w, L_a \cup \{z\})$. To see this, if $p(w, L_a) = p(w, L_a \cup \{x_m\})$ (resp. $p(w, L_a) = p(w, L_a \cup \{y_m\})$) simply set $z = x_m$ (resp. $z = y_m$). If instead $p(w, L_a) > p(w, L_a \cup \{x_m\})$ and $p(w, L_a) > p(w, L_a \cup \{y_m\})$ then, since $a = x_{m+1} = y_{m+1}$, it must be $p(w, L_a) = p(w, L_a \cup \{z\})$ for some $z \in U_a$ by the construction of the algorithm (otherwise, given that x_m and y_m impact w in L_a, it should be $w = x_{m+1}$ or $w = y_{m+1}$, contradicting $a = x_{m+1} = y_{m+1}$). Fix such a z.

Monotonicity and the construction imply that $p(z, L_a \cup \{z\}) = 1 - p(w, L_a \cup \{z\}) = 1 - p(w, L_a)$. But since $z \in U_a$ we have $z \succ_i a'$, $i = 1, 2$, for all $a' \in L_a$, so that $p_\alpha(z, L_a \cup \{z\}) = 1$ and therefore $z \in \mathcal{I}(p_\alpha)$. And since in particular $z \succ_i a$, $i = 1, 2$, the initial hypothesis that (a, A) is maximal in $\mathcal{I}(p_\alpha)$ is contradicted. Since we have made no assumptions on α, this same reasoning applies to the case $p(a, A) = 1 - \alpha$ and $p_\alpha(a, A) = 1$.

$\boxed{\textit{(3.iv)}\text{: } p(a, A) = \alpha \text{ and } p_\alpha(a, A) = 0.}$

By construction there exist alternatives b and c (where possibly $b = c$) such that $b \succ_1 a'$ for all $a' \in A \setminus \{b\}$ and $c \succ_2 a'$ for all $a' \in A \setminus \{c\}$, so that $A \subseteq L_b^1$ and $A \subseteq L_c^2$. If

$b \neq c$, then by Monotonicity $p(b, A) > 0$ and $p(c, A) > 0$, contradicting Lemma 1. Thus let $b = c$, so that $b \succ_i a'$ for all $a' \in A \setminus \{b\}$ for $i = 1, 2$, and by construction $p\left(b, L_b^1 \cap L_b^2\right) > 0$. If $p\left(b, L_b^1 \cap L_b^2\right) = 1$ then by Monotonicity we have a contradiction, since $A \subseteq L_b^1 \cap L_b^2$, so that $p\left(b, L_b^1 \cap L_b^2\right) = p(b, A) = 1$. So the only possibility is that $p\left(b, L_b^1 \cap L_b^2\right) = 1 - \alpha$, which is not possible as it would require $p(b, A) = 1 - \alpha$ and $p_\alpha(b, A) = 1$, a configuration ruled out by case 3.3. Since we have made no assumptions on α, this same reasoning applies to the case $p(a, A) = 1 - \alpha$ and $p_\alpha(a, A) = 0$.

$\boxed{\textit{(3.v)}: p(a, A) = \alpha \text{ and } p_\alpha(a, A) = 1 - \alpha.}$

Observe that this can only be the case when $\alpha \neq \frac{1}{2}$. By construction $A \subseteq L_a^2$. By Negative Expansion applied to L_a^2 and $A \subseteq L_a^2$ it must be $p\left(a, L_a^2\right) = \alpha$, so that by construction it cannot be $p_\alpha(a, A) = 1 - \alpha$. ■