University of
St Andrews

School of Economics & Finance Discussion Papers

Imperfect Attention and Menu Evaluation

Paola Manzini and Marco Mariotti

School of Economics and Finance Discussion Paper No. 1319
21 Oct 2013 (revised 8 Apr 2015)
JEL Classification: D01
Keywords: attention, bounded rationality, stochastic choice

School of Economics & Finance
Discussion Paper Series
issn 0962-4031
http://ideas.repec.org/s/san/wpecon.html
info: econ@st-andrews.ac.uk

Modelling Imperfect Attention*

Paola Manzini[†] Marco Mariotti[‡]

This version: April 2015

Abstract

We propose a novel method to model an agent who is imperfectly attentive in the sense that she may consider only some of the alternatives available. Our methodology departs from the standard 'revealed preference' one: we make plausible assumptions on the values to the imperfectly attentive agent of different choice situations. We derive in this way a simple reduced-form model that is compatible with several cognitive processes underlying choice: the agent stochastically forms a consideration set by noticing each alternative with a given probability and then maximises a deterministic utility function over the consideration set.

J.E.L. codes: D0.

Keywords: attention, bounded rationality, stochastic choice

*We are grateful to David Dillenberger, Yorgos Gerasimou, Pietro Ortoleva, Kemal Ozbek and Dan Sasaki for insightful comments, to various seminar audiences (notably attendees at BRIC2014 in Barcelona), and to ESRC for financial support through grant ES/J012513/1.

[†]School of Economics and Finance, University of St. Andrews, Castlecliffe, The Scores, St. Andrews KY16 9AL, Scotland, U.K. e-mail: paola.manzini@st-andrews.ac.uk.

[‡]School of Economics and Finance, Quen Mary University of London, Mile End Road, London E1 4NS, U.K. (e-mail: m.mariotti@qmul.ac.uk).

1 Introduction

1.1 Motivation

Cognitive shortcomings and biases that cause a decision maker to ignore some of the available options of choice are varied and well documented. For example, there are hundreds of new computer models every year which the prospective purchaser does not have the time to analyse in detail. So, as carefully evidenced in Sovinsky Goeree [23], the *consideration set* of the consumer is typically a strict subset of all computer models that are available on the market. In this paper we gather all these phenomena under the umbrella term of 'imperfect attention' and tackle the question of how to model an imperfectly attentive agent.

Recent evidence on consideration set formation is suggestive and highlights some broad themes while still not agreeing on the details. The typical consumer in a supermarket faces a choice between several varieties and brands in each product category, possibly under time pressure, and does not see all of them. Reutskaya, Camerer, Nagel and Rangel [21] (henceforth RCNR) simulate the supermarket situation in a laboratory. One of their key findings is that consumers roughly optimise on the (sub)set of options they see;[1] the latter emerges out of a 'hybrid' random search process in which the stopping rule is not of the 'satisficing' type (in the sense of Simon [22]). Arguably, however, a consideration set might be arrived at in different ways in different contexts. Indeed, Caplin, Dean and Martin [4] (henceforth CDM), using a different experimental methodology, find evidence for a menu search strategy in which the stopping rule of the satisficing type: the agent stops considering options as soon as a reservation value is hit.

Here, we steer clear of seeking to specify the details of the cognitive process (e.g., a sequential search strategy or an instantaneous visual focusing mechanism) that makes alternatives in the menu noticed. Our aim is instead to build a model of an imper-

[1] RCNR find evidence supporting a form of 'soft' optimisation on the stochastic consideration set: the agent optimises with small errors which are second-order compared to those deriving from not seeing all items.

fectly attentive agent that is of a *reduced-form* type, open to multiple interpretations, and versatile for use in applications. The analytical representation we obtain can be interpreted as a general two-stage choice process (as suggested by the observations both by RCNR and CDM): namely, there is first a *consideration* phase (a search phase in RCNR and CDM) and then an *optimisation* phase on the consideration set.[2] However, the details of whatever operations are performed by the agent to form the consideration set are not modelled. Instead they are collapsed into a set of parameters, as in the direct econometric specification of Sovinsky Goeree [23] or in our own axiomatically derived model in Manzini and Mariotti [16].

A core innovation of our paper is the methodology we use to derive and characterise the model, which aims to complement the existing 'revealed preference' style of approach. In the latter, direct consistency restrictions are posed on choice from menus, or on enriched choice data as in CDM.[3] We submit that a useful alternative way to model imperfect attention is to think instead of the *values* that different choice situations have for the agent. Based on comparisons between these values, we will infer a choice model *indirectly*, through the structure of the formula that will be shown to represent menu values. Often the analyst can make grounded hypotheses about such comparative values, even when he is unable to specify exactly the cognitive process that leads the agent to choice. The crucial property we will assume is that the inattentive agent is made better off by the appearance of top alternatives. Here, 'top' means that the agent would be willing to choose them if he was aware of all alternatives. In many contexts, that good alternatives (jobs, consumer products, hospitals...) are beneficial is both a reasonable premise of the analysis and one that is empirically testable, albeit with non-choice data. For example, it seems safe to assume that a new high-quality hospital on average benefits prospective patients. This assumption can be tested using data such as clinical outcomes and self-reported satisfaction.

[2]Note that in the satisficing model optimisation on the set of seen alternatives is built in the stopping rule, so the CDM search process could also be viewed in the same light. Masatlioglu, Nakajima and Ozbay [17] follows a similar two stage procedure in a deterministic environment. We discuss related literature fully in section 5.

[3]See section 5 for remarks on the growing revealed preference literature on imperfect attention.

1.2 Preview: approach and results

We make a distinction akin to the standard one between the direct utility function u and the indirect utility function V of classical consumer theory (we use the language of utility in this preview for the sake of exposition – in the analysis proper the primitive will be a preference rather than a utility).[4] Notice that these objects could be used straightforwardly (if not very usefully) to *define* a rational consumer. By the construction of V, a rational consumer is simply one for whom $u(a) = V(x)$ for any bundle a and any budget x that contains a as one of its best bundles.

More usefully, we can define an inattentive agent along similar lines. Let us distinguish between the value $V(x)$ of choosing from a menu x (i.e., the value resulting from whatever choice process the inattentive agent uses), and the value $u(a)$ of an alternative a.[5] We assume that imperfect attention has (at least) the two following simple consequences:

(1) If the agent is imperfectly attentive, and if a is a best alternative in menu x, then $u(a) > V(x)$.

(2) If the agent is imperfectly attentive, and if a is at least as good as any alternative in menu x, then $V(x \cup \{a\}) > V(x)$.

In the context of the computer example, (1) says that if a is the best model for the inattentive consumer, then she is better off with a than if she has to discover a from the set of models available in the market. And (2) says that if a new model a becomes available in the market, and a is the best model for the consumer, then she is better off.

Property (1) should be part of any definition of imperfect attention, as it captures a core aspect that is common to many of its possible causes. In fact, it is hard to deny that whatever it is that makes attention imperfect, it has precisely the effect that an agent is worse off when choosing from a menu compared to having an optimal option in that menu. Property (2), on the other hand, specifies the form of imperfect attention we are

[4]That is,. $V(x) = \arg\max_{a \in x} u(a)$, where x is a budget and a a bundle of commodities.

[5]In fact, in the analysis we shall not postulate any numerical utility function, but just a welfare ordering of choice situations and alternatives - u and V are used here just for ease of exposition.

interested in modelling. It essentially excludes those cognitive processes for which the new top alternative is not only not paid attention to, but also serves as a potent distractor from other top alternatives (e.g. because of cognitive overload). As explained in the next section, (2) is consistent with some leading models of, and evidence on, attention formation.

(1) and (2) are the only properties of imperfect attention that we shall assume. This is not meant as an exhaustive definition of imperfect attention. The main point of the paper is to show that these two properties by themselves lead the analyst to a certain model of imperfect attention, that has some convenient simplifying features compared to other models that appear plausible theoretically. If it is deemed that other features of imperfect attention are essential beside (1) and (2), they can be captured by a further specialisation of that model.

The implications of (1) and (2) (plus risk rationality in the form of the von Neumann Morgenstern axioms) are intriguing. Choices from menus can be represented as the outcome of *deterministic* utility maximisation over a *stochastic* consideration set. Moreover, we can pin down three distinctive simplifying aspects of the stochastic process that generates the consideration set:

- *Stochastic independence in consideration*: the probability of considering any group of alternatives in a menu can be taken to be the product of the probabilities of considering each of those alternatives individually - there is no correlation in consideration.
- *Menu independence in consideration*: the probability of considering an alternative can be taken to be a property of the alternative, independent of the menu.
- *Value independence in consideration*: the probability of considering an alternative can be taken to be independent of its value.

Specifically, the value structure of menus satisfies the Expected Utility assumptions in addition to (1) and (2) if and only choices *from* menus are made as follows. Each alternative a in menu x is considered with a probability $\alpha(a)$ (the *attention parameter*) and is assigned a utility value $u(a)$. Then the agent picks, according to some given order, one of the highest utility alternatives among those which he has considered. For example, if there are three alternatives a, b and c with $u(a) > u(b) = u(c)$ the value of menu $x = \{a, b, c\}$ is

$$\begin{aligned} V(x) &= \alpha(a) u(a) \\ &\quad + (1 - \alpha(a))(1 - (1 - \alpha(b))(1 - \alpha(c))) u(b) \\ &\quad + (1 - \alpha(a))(1 - \alpha(b))(1 - \alpha(c)) u(\varnothing) \end{aligned}$$

where $u(\varnothing)$ is the utility of not consuming any of the options in x.[6] Then the implicit choice from x is determined either by considering a (with the probability given by the attention parameter) and choosing it irrespective of whatever else is considered; or missing a and picking one of b or c if considered; or missing everything in x.

The analysis is contained in section 3, which follows the introduction of the formal setup. In section 4 we also study the consequences of dropping property (2) and using only the core property (1). This results in a more general model similar to the one above, in which the attention parameters become menu-dependent. However, this menu-dependence is still constrained: adding to a menu a better alternative than the existing ones cannot increase the attention paid to them. This is again a natural feature when attention derives from some types of search process. Finally, in section 4.2 we discuss some of the restrictions that are implicit in the very structure of the value function V above, independently of any additional constraints imposed on the attention parameters. We discuss the connections with existing literature in section 5, and section 6 concludes.

[6]The device of allowing the possibility of 'not choosing' is used in several other contexts, see e.g. Gerasimou [10] and Kreps [14] for the deterministic case and Brady and Rehbeck [2] and Corbin and Marley [5] for the stochastic case, beside Manzini and Mariotti [16]. In the empirical IO literature using discrete choice models, it is also standard to introduce an outside option to allow for the possibility that the data do not contain all brands or models that have a positive market share (as in Sovinsky Goeree [23]).

2 The Model

Let X be a finite set of alternatives, denoted $a, b, c, \ldots$ and let $\mathcal{X} = 2^X$. An element $x \in \mathcal{X}$ is called a *menu*.

Let $\Delta(X \cup \mathcal{X})$ be the set of lotteries on $X \cup \mathcal{X}$. To simplify notation we identify the degenerate lotteries in $\Delta(X \cup \mathcal{X})$ with elements of $X \cup \mathcal{X}$.

A nonempty $x \in \mathcal{X}$ is interpreted as the situation in which the agent has to choose an element from x. A non-degenerate element of $\Delta(X \cup \mathcal{X})$ is interpreted as a risky situation in which the agent either will have to pick an element from some menu or will be given some alternative, with the identity of the menu or the alternative to be determined randomly.

Special attention must be paid to $\varnothing$. While this object will be part of the evaluation, it is *not* interpreted as an object of choice. In other words, the agent never faces a situation in which she has to choose between alternatives in a menu and $\varnothing$. The situation $\varnothing$ takes place when no alternative in X is available to the agent (which in our interpretation of the model happens when no alternative is considered). The agent does not choose $\varnothing$: it is rather something that happens to her; it is a situation rather than an alternative.

We impose properties on a binary relation $\succsim$ on $\Delta(X \cup \mathcal{X})$, where $g' \succsim g$ for any $g, g' \in \Delta(X \cup \mathcal{X})$, interpreted as the agent being better off in situation g' than in situation g. We shall call $\succsim$ a *preference*, though it should remain clear that $\succsim$ is not meant to be associated with a process of choice between menus by the agent; our agent chooses alternatives from menus, while $\succsim$ describes (ordinally) how well off she is from those choices.

We consider the following properties for a preference $\succsim$ (with $\succ$ and $\sim$ denoting the asymmetric and symmetric parts, respectively):

A0 - Choice Is Valuable: $x \succ \varnothing$ for all $x \in \mathcal{X}$.

A1 - Order: $\succsim$ is a weak order.

A2 - Continuity: For all $g, g', g'' \in \Delta(X \cup \mathcal{X})$ such that $g'' \succsim g \succsim g'$, there exists

$\alpha \in [0,1]$ such that $\alpha g'' + (1-\alpha) g' \sim g$.

A3 - Independence: For all $g, g', g'' \in \Delta(X \cup \mathcal{X})$ and $\alpha \in [0,1]$: $g \succsim g' \Leftrightarrow \alpha g + (1-\alpha) g'' \succsim \alpha g' + (1-\alpha) g''$.

A4 - Endowment is better than choice For all $x \in \mathcal{X}$ and $a \in X$: $a \succsim b$ for all $b \in x \Rightarrow a \succ x$.

A5 - Top options are valuable: For all $x \in \mathcal{X}$ and $a \in X$: $a \notin x$ and $a \succsim b$ for all $b \in x \Rightarrow \{a\} \cup x \succ x$.

A0 is just a definition of the range of choice situations we consider. The menus faced by the agent contain alternatives that are valuable compared to the situation in which choice is not available. They are 'opportunity sets'.

A1-A3 are a version of the vNM axioms applied to the particular domain of menus and alternatives.

Finally, A4 and A5 are the two axioms that define imperfect attention in this paper. As discussed in the Introduction A4 is a core property of imperfect attention, whatever causes it.

A5, on the other hand is compatible with many but not all consideration processes. For A5 to fail strictly, a top alternative should simultaneously not be noticed and make other good alternatives not noticed. This may involve some fairly elaborate, but possible, cognitive process related to 'frame' elements. For example, the introduction of the top alternative in a list occurs simultaneously with a demotion of other top alternatives to the lower ranks in the list. Cognitive overload, impairing the agent's general ability to pay attention to alternatives, is another possibility. Similarity effects might also be invoked: the new top alternative might be similar to some existing poor alternatives and draw attention to them. On the other hand, if the consideration set is determined by a *random search process*, it is hard to see how a new top alternative could harm. Suppose the agent searches randomly and sets a reservation value, at which he stops searching. Whatever the reservation value that is set, if the agent finds the new top alternative a before stopping, she will choose it (or an alternative of equivalent value); and if not, she will not be any worse off than if a had not been in the menu.

Thus an agent using the satisficing search process evidenced in CDM should satisfy A5. A similar reasoning applies to the 'hybrid' search model proposed in RCNR, in which the agent stops, after seeing any new item, with a probability that is increasing with the value of the best alternative seen up to that point and with the time elapsed.

Alternatively, suppose that the consideration set is determined by an *attention filter* (Masatlioglu, Nakajima and Ozbay [17]). This term refers to the property that removing an alternative that is not paid attention to does not affect the attention paid to other alternatives. If the new top alternative is noticed, it (or an equally good alternative) will be chosen,[7] so that the menu value is not decreased; and if it is not noticed, it cannot affect the agent's consideration set because of the attention filter property: so in this case, too, adding the top alternative cannot hurt.

We shall establish a link between preferences that satisfy the above properties and a numerical representation of menu values suggesting the two-stage stochastic process of choice - first consider, then choose - discussed in the introduction.

A strict total order $\hat{\succ}$ of X *refines* $\succsim$ if $a \succ b \Rightarrow a \hat{\succ} b$. In the definition below recall that we identify degenerate lotteries on an outcome with the outcome itself.

Definition 1 *An attention representation for* $\succsim$ *is a triple* $(\hat{\succ}, u, \alpha)$, *with* $\hat{\succ}$ *a strict total order of* X *that refines* $\succsim$, $u : \Delta(X \cup \mathcal{X}) \rightarrow \mathcal{R}$ *a vNM utility function representing* $\succsim$, *and* $\alpha : X \cup \mathcal{X} \rightarrow (0,1)$, *such that, for all* $x \in \mathcal{X}$:[8]

$$u(x) = \sum_{a \in x} \prod_{b \in x : b \hat{\succ} a} (1 - \alpha(b, x)) \alpha(a, x) u(a) + \prod_{a \in x} (1 - \alpha(a, x)) u(\varnothing) \qquad (1)$$

In this representation u is an *evaluation function*. The function α is an *attention function* that assigns a value to the attention received by each alternative in each menu: the interpretation is that any alternative a has a probability $1 - \alpha(a, x)$ of being missed by the agent in menu x. The strict ordering $\hat{\succ}$ is merely a tie-breaking device that resolves indifferences between the stochastic set of alternatives that are considered, with no impact on the value of the menu. Under these interpretations, the agent maximises u on the set of alternatives that are both feasible and considered.

[7]Recall that the agent we are seeking to model is a preference maximiser on the consideration set.
[8]We use the convention that the product over the empty set is equal to one.

2.1 Comments on the menu choice literature

We should clarify a potential misconception concerning the meaning of the relation $\succsim$. When comparing menus, we are *not* assuming the multistage decision structure of the menu choice literature initiated by Kreps [13]. That is, we are not considering an agent who first chooses a menu and then, after some event (e.g., the onset of temptation, or the arrival of information) picks an element from the menu she has chosen. What we do assume is that comparisons between menus (or between a menu and an alternative) are made in the abstract, as part of a definition: they are made *outside* of anyone's menu choice process, but using the agent's own preferences. Such comparisons have exactly the same status they have, for example, in a statement such as 'A misanthrope is better off living on his own than in a commune'. This statement is meaningful as there is a matter of fact about it: either a misanthrope is better off in a commune, or he is not. Moreover, this statement (partly) defines what a misanthrope is.

Similarly, as we have argued, there is a matter of fact about the assertions that for an agent there is, or there is not, a difference between having alternative a and choosing from a menu x in which a is its best alternative. Such statements are meaningful, but they are different from the preference statements of the menu choice literature.

As a consequence, the preference for an alternative a over a menu x containing a as one of its best alternatives should not be seen in this paper as a 'preference for commitment' as in the two-stage menu choice literature. It is hard to even make sense of what commitment might mean when the agent is inattentive. If the agent may neglect some alternatives in the menu, she will not feel the need to commit: she will just believe the menu to be one thing when in reality it is another thing. On the other hand, if we assumed that somehow the agent perceived all alternatives at the stage of evaluating menus, it is not clear why the later stage of a choice from menu may be vitiated by the lack of consideration of some alternatives.

There are some possible, but rather contrived, interpretations along the lines of the menu choice literature. In a first interpretation, the agent is aware of the alternatives when she chooses between menus but not when she chooses from a menu; she is aware of this structure; and she is extremely sophisticated in taking account of it when choosing between menus. A second, and better, choice-based interpretation attributes the preference to a fully informed and sympathetic third party (a parent, a doctor) who chooses menus using the agent's preferences.

3 Analysis

The goal of this section is to show that, together with the Expected Utility axioms, A4 and A5 characterise a particularly simple version of the attention representation, in which the attention function α is constant across menus.

Let us say that, in an attention representation $(\hat{\succ}, u, \alpha)$, α is *menu independent* if, for all $x, y \in \mathcal{X}$ and for all $a \in x \cap y$, $\alpha(a, x) = \alpha(a, y)$. In this case we write $\alpha(a)$ instead of $\alpha(a, x)$.

Theorem 1 *A preference $\succsim$ satisfies A0-A5 if and only if it has an attention representation $(\hat{\succ}, u, \alpha)$ in which α is menu independent.*

The logic of the proof[9] is simple. The axioms enable an iterated 'peeling off' procedure that makes any menu x indifferent to a lottery over two outcomes, one of them being a sub-menu obtained by removing an alternative in x, as follows. Suppose for simplicity that preferences are strict and number the alternatives in x from best to worst as $x = \{a_1, ...a_K\}$. By A5 it must be that $x \succ x \setminus \{a_1\}$. Given that $a_1 \succ x$ by A4, we can then construct (thanks to Continuity) a lottery $\alpha a_1 + (1 - \alpha) x \setminus \{a_1\}$ for some unique $\alpha \in (0, 1)$ that is indifferent to x, so that by the vNM axioms this can be represented as $u(x) = \alpha u(a_1) + (1 - \alpha) u(x \setminus \{a_1\})$ for some vNM utility u. Then we can iterate the process applying the argument successively to $x \setminus \{a_1\}$, $x \setminus \{a_1, a_2\}$,.... Finally, we show that the resulting formula, with menu independent parameters, also works on menus that are not related by set inclusion.

Some aspects of the representation should be noted.

- **Stochastic independence and value independence in consideration**

We have highlighted the 'menu independence in consideration' feature of the representation, but no less important are the features of stochastic independence and of

[9] See Appendix A.1.

value independence: not only are the attention parameters defined independently of the menu, but also of the evaluation function u, and in each menu the probability of the consideration set is the product of the attention parameters of the alternatives in it. Here these features are implications of theoretical hypotheses, but it is interesting to note that there is at least some direct empirical support for both. In particular, Van Nierop et al. [19] find evidence of the lack of correlation in attention, and RCNR of the independence between values and attention.

- **Indifferent alternatives do not imply indifferent menus.**

Two menus x and y of the same cardinality that are composed of indifferent alternatives[10] are not necessarily indifferent to the agent. For example, suppose that two yogurt types a and b are equally good for a consumer, and that they are both strictly better than a third type c. It can still be the case that $\{a, c\} \succ \{b, c\}$. In the interpretation we are giving preferences, this is not a puzzling phenomenon: even with menu independence of the attention parameters, the discrepancy in the values of menus that contain indifferent alternatives can be explained by the different levels of attention received by the two distinct alternatives a and b in the two menus. In the example, $\alpha(a) > \alpha(b)$ would rationalise the preference - perhaps yogurt a has a more catchy label and the consumer is less likely to choose the inferior c through inattention. Similarly, even if indifferent between a taxi and a bus, you may fail to spot bus number 38 out of many buses outside a station, while any taxi would do.

- **But why couldn't it be...**

We have proposed two properties as characteristic of imperfect attention, Endowment is better than choice and Top options are valuable**.** But it could be argued that these two properties may also be features of other forms of bounded rationality. For example, consider an agent who is prone to 'implementation errors' (as e.g. in Mattsson and Weibull [18], Ahn and Sarver [1] and Koida [11]). One could say that in this case, too, it is better to have an alternative than to pick it from a menu in which it is the best, and also that adding a top alternative makes the agent better off.

We agree with argument. The interpretation of the primitives in a representation

[10] That is, there exists a bijection f from x to y with $f(a) \sim a$ for all $a \in x$.

must at least in part come from the context to which the representation is meant to be applied, or from additional information: it cannot be exhausted a priori in a handful of properties. For an analogy, do the Savage axioms characterise the standard 'rational' subjective expected utility maximiser, or do they rather characterise a 'behavioural' optimistic (or pessimistic) agent whose probabilities depend on the stakes? As shown by Dillenberger, Postlewaite and Rozen [6], both interpretations are admissible. Similarly, an agent satisfying the Weak Axiom of Revealed Preference can equally well characterise both a standard utility maximiser and a boundedly rational agent who uses a sequential heuristic to make decisions (Mandler, Manzini and Mariotti [15]).

- **A formula for the attention parameters**

In the representation of theorem 1, we have that for all $x \in \mathcal{X}$ and with $a \hat{\succ} b$ for all $b \in x \setminus \{a\}$:

$$\begin{aligned}
u(x) &= \sum_{b \in x} \prod_{c \in x: c \hat{\succ} b} (1-\alpha(c))\, \alpha(b)\, u(b) + \prod_{b \in x} (1-\alpha(b))\, u(\varnothing) \\
&= \alpha(a)\, u(a) + \\
&\quad + (1-\alpha(a)) \left(\sum_{b \in x \setminus \{a\}} \prod_{c \in x \setminus \{a\}: c \hat{\succ} b} (1-\alpha(c))\, \alpha(b)\, u(b) + \prod_{b \in x \setminus \{a\}} (1-\alpha(b))\, u(\varnothing) \right) \\
&= \alpha(a)\, u(a) + (1-\alpha(a))\, u(x \setminus \{a\})
\end{aligned}$$

and therefore

$$\alpha(a) = \frac{u(x) - u(x \setminus \{a\})}{u(a) - u(x \setminus \{a\})}.$$

This formula is interesting in two respects. First, it shows that in this case the $\alpha(a)$ are *uniquely defined*, since they are invariant to any positive affine transformation of u, and u (as a vNM utility) is unique precisely up to such transformations. Observe that for any $a \in x$ there exists an x with $a \in x$ and $a \hat{\succ} b$ for all $b \in x \setminus \{a\}$, for example $x = \{a\}$.

Secondly, the formula provides an interpretation of the attention parameters in terms of utility. The attention paid to a is measured by the ratio between the incremental utility of *having the opportunity of choosing* a (from a menu in which a is best), and the incremental utility of *having* a instead.

4 Menu dependence

4.1 The role of the 'Top options are valuable' assumption

While we have argued for A4 (Endowment is better than choice) as the core property of imperfect attention, the crucial property yielding the menu independence of the attention representation is A5 (Top options are valuable). In this section we show that even without that property, menu dependence in attention can be substantially limited.

In principle, one may imagine that adding an alternative to a menu increases the attention paid to an existing alternative. This could be the case, for example, through a similarity effect, when the new alternative is similar to the existing one. Or, a product offered by a multiproduct firm (e.g. a program of a media company) may draw attention to other products offered by the same firm.[11] But if preferences satisfy A0-A4, even when violating A5, this effect can be excluded when the new alternative is better than the existing ones, as we show in theorem 2 below.

Let us say that an attention representation $(\hat{\succ}, u, \alpha)$ satisfies *monotonicity in consideration* if

$$b \hat{\succ} a \text{ for all } a \in x \Rightarrow \alpha(a, x) \geq \alpha(a, x \cup \{b\}) \text{ for all } a \in x. \tag{2}$$

Monotonicity in consideration (and much more) is implied for example when the attention parameters have the 'Luce structure' $\alpha(a, x) = \frac{\lambda(a)}{\sum_b \lambda(b)}$, where λ is a strictly positive real valued function of the alternatives. Another situation where monotonicity in consideration is a natural property is when the attention representation is the reduced form of a satisficing search process. Then adding a top alternative a should make it less likely that the other alternatives will be seen, since the agent will stop whenever he finds a. On the other hand, monotonicity in consideration excludes the effect studied in Payro and Ülkü [20], whereby the agent may consider inferior alternatives just because they are similar to the top alternative in the menu.[12]

[11] As in Eliaz and Spiegler [8].

[12] Payro and Ülkü [20] do not explicitly mention consideration effects, but this is a possible interpretation of their choice model.

Theorem 2 *A preference $\succsim$ satisfies A0-A4 if and only if it has an attention representation $(\hat{\succ}, u, \alpha)$ that satisfies monotonicity in consideration.*

The logic of the proof[13] has the same flavour of that for theorem 1, although it is markedly less straightforward because of the need to take into account the case $x\backslash\{a_1\} \succsim x$ (which is ruled out by A5 in theorem 1).

The representation in theorem 2 allows the addition of a top alternative to reduce the attention paid to existing alternatives. Then it could happen, for example, that $a \succsim b$ and

$$\begin{aligned}\alpha(b,\{a,b\}) &< \alpha(b,\{b\})\frac{1}{(1-\alpha(a,\{a,b\}))} - \frac{\alpha(a,\{a,b\})}{(1-\alpha(a,\{a,b\}))}\frac{u(a)}{u(b)} \\ &\Leftrightarrow \alpha(b,\{b\})\,u(b) > \alpha(a,\{a,b\})\,u(a) + (1-\alpha(a,\{a,b\}))\,\alpha(b,\{a,b\})\,u(b) \\ &\Leftrightarrow \{b\} \succ \{a,b\}\end{aligned}$$

That is, adding a top alternative to a menu may decrease the value of that menu. This will be the case if adding the new top alternative reduces by a sufficient amount the attention paid to the best existing alternatives, and the attention paid to the new top alternative is not too large. This is a direct violation of A5. The perhaps surprising message of theorems 1 and 2 taken together, then, is that excluding (through A5) the negative contribution of a top alternative to the value of a menu does not merely limit the impact on the attention parameters that a top alternative may have: it forces these parameters to be completely menu-independent!

A second observation concerns the uniqueness of the representation. Obviously the same attention parameters work with any positive affine transformation of u. In general, however, while thanks to the vNM axioms the evaluation function u is cardinally unique (given the preference), it may only be possible to restrict, but not to pin down uniquely, the attention functions α that are compatible with a given preference. Uniqueness is not guaranteed given the nonlinear way the attention parameters enter the representation. However, the monotonicity condition in the representation helps to put some bounds on the attention parameters, as we illustrate in example A.3 in the Appendix.

[13]See Appendix A.2.

There is of course a second source of non-uniqueness in the way that $\hat{\succ}$ breaks indifferences in $\succsim$, but this is less important since the *evaluation* of a menu is not impacted by the exact choice of $\hat{\succ}$, the overall attention enjoyed by alternatives in a menu that belong to the same indifference class being independent of the choice of $\hat{\succ}$. For menu evaluation purposes, regardless of how the alternatives within the same indifference class are ranked by $\hat{\succ}$ what matters is the probability that *some* alternative in a's indifference class is noted by the decision maker. There is no bonus for noticing more than one alternative in any indifference class, given that only the single alternative that is ultimately chosen determines value. So at least in this respect the lack of a complete identification of the attention parameters does not matter. Note however that together with Monotonicity in consideration the exact specification of $\hat{\succ}$ may change the admissible range of the attention parameters $\alpha(a, x)$ (this is illustrated in Appendix A.4).

4.2 The attention representation with no restrictions

Finally, we discuss what conditions on preferences are implicit in an attention representation *per se*, regardless of any restriction imposed on the attention parameters. Even so, not all $\succsim$ have an attention representation. It is instructive to go through a couple of examples.

Suppose first that $x = \{a, b\}$ and

$$x \succ a \succsim b \succ \varnothing$$

Then, if there were an attention representation, we would have

$$\begin{aligned} u(x) &= \alpha(a, x)\, u(a) + (1 - \alpha(a, x))\, \alpha(b, x)\, u(b) \\ &+ (1 - \alpha(a, x))(1 - \alpha(b, x))\, u(\varnothing) > u(a) \geq u(b) > u(\varnothing) \end{aligned}$$

a contradiction.[14] These preferences might result when intrinsic value is attached to to the act of choice: having the choice between Pravda and Wall Street Journal might

[14]Even if we weakened the previous relations to $x \succsim a \succsim b \succ \varnothing$, insofar as these preferences deviate from those in the main text, the agent would clearly have to have perfect attention at least for one alternative, a case also excluded from the representation.

be more valuable to the agent than being given Pravda, even if ultimately the agent would choose Pravda anyway. A4, which aims to isolate the attention motive, directly rules out such preferences.

A more subtle example of preference that cannot be captured by an attention representation is as follows. Let $x = \{a, b\}$ and

$$\begin{aligned} a &\succ b \succ \{b\} \\ a &\succ x \succ \{b\} \\ x &\sim ka + (1-k)\, b \text{ for some } k \in (0,1) \end{aligned} \tag{3}$$

The preference in the second line is not essential and serves only to simplify the proof of the Observation below. The key preference is in the third line. The indifference between a menu and a mixture of the alternatives it contains could be naturally explained by *correlation in consideration*. For example, suppose that when faced with the menu x the agent considers only a with probability k and considers only b with probability $(1-k)$, leading to the indifference in 3. However, this consideration pattern obviously cannot be generated independently by means of parameters $\alpha(a, x)$ and $\alpha(b, x)$. Another explanation, that does not rely on correlation, is that with probability k the agent considers both a and b (then choosing a) and with probability $(1-k)$ he considers only b. This consideration pattern can be generated independently with attention parameters $\alpha(a, x) = k$ and $\alpha(b, x) = 1$, but this type of *full attention* for an alternative is not admissible in our framework. We show that, short of assuming correlation, full attention for b is implied by the above preferences, leading to a contradiction:

Observation: The preferences in (3) have no attention representation.

Proof. Suppose that an attention representation exists, and let u be such a representation normalised by $u(\varnothing) = 0$. Then:

$$ku(a) + (1-k)\, u(b) = u(x) = \alpha(a, x)\, u(a) + (1 - \alpha(a, x))\, \alpha(b, x)\, u(b)$$

implying $\alpha(b, x) = 1$. Since $a \succ x \succ \{b\}$, we have $u(a) > u(x) > u(\{b\})$, so that

$$u(x) = k' u(a) + \left(1 - k'\right) u(\{b\})$$

for some $k' \in (0,1)$. Then, since $u(\{b\}) = \alpha(b,\{b\})\,u(b) + 0$ by the normalised representation, we have

$$u(x) = \alpha(a,x)\,u(a) + (1-\alpha(a,x))\,u(b) = k'u(a) + \left(1-k'\right)\alpha(b,\{b\})\,u(b)$$

implying $\alpha(b,\{b\}) = 1$. It follows that $u(\{b\}) = u(b)$, so that u cannot represent $b \succ \{b\}$, a contradiction. ■

This example highlights what kind of pattern is excluded by the lack of correlation that is implicit in an representation. While no single axiom in our characterisations rules out the above preference configuration directly, our axiom system as a whole does so.

Because we have assumed vNM rationality, one can see our main representation as capturing deviations from rational behaviour that can be imputed *exclusively* to imperfect attention and not to deviations from the Expected Utility (EU) hypothesis. Since we have shown above that vNM rationality implies a lack of correlation in consideration, we can interpret any evidence of correlated consideration as a separate departure from full rationality, distinct form imperfect attention *per se*.[15] While disentangling the EU hypothesis from inattention is theoretically clean, and also a convenient simplification in applications, future research could fruitfully address the relation between inattention and violations of EU.

5 Related literature

Imperfect attention in this paper is close to the cognitive limitations in choice studied by, beside the already mentioned CDM and RCNR, Brady and Rehbeck [2] (BR), Caplin and Dean [3] (CD); Caplin, Dean and Martin [4] (CDM); Echenique, Saito and Tserenjigmid [7] (EST), Eliaz and Spiegler ([8], [9]) (ES); Manzini and Mariotti [16] (MM); Masatlioglu, Nakajima and Ozbay [17] (MNO); Sovinsky Goeree [23] (SG). In all these papers the agent fails to consider some alternatives in a menu. While ES explore in detail the consequences of imperfect attention in a strategic setting, MNO and MM

[15] See Brady and Rehbeck [2] for intuitively plausible examples of correlations in consideration.

give an abstract characterisation of consideration sets models based on a standard revealed preference method. They consider the agent's choices from a set of menus and state conditions under which the agent's choices could be interpreted as deriving from a certain type of imperfect attention. MNO focus on deterministic choices and look at the 'attention filter' restriction of the dependence of attention on the menu. MM study stochastic choices and characterise a simplified version of the choice procedure implied by theorem 2. MM main characterising axiom is a form of stochastic menu independence of choice, which says that the impact that an alternative a has on the choice probability of another alternative b is independent of which other alternatives are present in the menu beside a and b. The MM model is a particular case of the implicit choice rule suggested in this paper, because it posits a strict preference ordering. BR more substantially generalise this model by allowing menu effects an correlation in consideration. EST propose a different model, in which alternatives are perceived by the agent in a fixed order, and are chosen with Luce probabilities only if no alternative that ranks higher in the perception order has been chosen.

CD on the other hand innovate the revealed preference method by positing an enriched set of non-standard data, and assuming that the analyst can observe provisional choices and contemplation times. CDM put in practice this methodology in an experimental setting, validating the CD search-satisficing model of choice. While the innovative techniques used by CD and CDM allow in a sense the consideration to be observed directly by the analyst, SG uses careful econometric techniques of a more standard kind to infer the existence of non-trivial consideration sets in the purchase of personal computers. Yet differently, RCNR use eye-tracking techniques to identify the visual fixation on objects on the part of the agent (see also Krajbich and Rangel [12] and the literature therein for further results in this research program)

6 Concluding remarks

There are many reasons why attention may be imperfect and there are many modelling strategies for each of these reasons. The most ambitious one is to try and open the 'black box' of cognition, getting closer to observing the consideration set of the agent directly. This is the strategy followed, for example, by RCNR, in which they provide evidence for a 'hybrid' search process (tested against a satisficing search process and full rationality) using an eye-tracking methodology. CDM offer another example of this strategy using a different experimental approach, and reach a different conclusion by supporting a version of Simon's satisficing model.

On the other hand, especially in applications, until the 'open the black box' strategy is perfected and enough evidence is accumulated for a variety of choice contexts, we argue that it is useful to have reduced-form models that do not commit the researcher to a particular cognitive black box, and that are broadly consistent with various search processes, or even with different cognitive processes for the construction of a consideration set altogether. An example of this approach is the work by Masatlioglu, Nakajima and Ozbay [17], where the class of 'attention filter' based choices is axiomatised through classical revealed preference arguments. A similar example, but in a stochastic choice environment, is our own work in [16].

What the present paper adds to this second strand of literature is a different basis for such models. We think that the most appealing aspect of the methodology we have proposed is that it yields a fairly specific, simple and tractable representation of imperfect attention using - beside the working assumption of Expected Utility - only very broad properties of its *welfare consequences*. It is in this sense that our implied choice model can be seen as a reduced form of several more detailed stories that may lie behind the agent's failure to consider all alternatives.

The main practical contribution of this paper has been to show that, if there are reasons to believe that the decision maker's cognition, while imperfect, does not lead to harm when new top alternatives are introduced (axiom A5), then the researcher is legitimised to use our reduced-form representation of imperfect attention. This way of justifying this model frees the researcher from the need to use revealed preference conditions such as those in Manzini and Mariotti [16], a task that may be difficult to perform when there is little variation in the menus from which choice is observed. A particularly stark case in question, which we are currently investigating and where consideration effects are very likely to play a role, is hospital choice: given the slow change in the set of available hospitals, typically only choices from a single menu will be observed, making revealed preference tests difficult to implement. On the other hand, the assumption that a single additional top hospital would make the prospective patients better off (our A5) seems much easier to justify a priori than the menu-independence type of properties typically used in revealed preference analysis.[16]Furthermore, as noted in the Introduction, indicators both of clinical outcomes and of patient satisfaction are data that are relatively easy to collect and that would permit the empirical testing of the welfare hypothesis in A5.

The model we have proposed has some practical advantages that should come handy in applications: values can be treated independently of 'salience'; menu-effects and correlation effects in attention can be ignored; and (in the case of Theorem 1) the attention parameters are uniquely identified.

References

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

[2] Brady, R. and J. Rehbeck (2014) "Menu-Dependent Stochastic Consideration", mimeo, University of California San Diego.

[3] Caplin, A. and M. Dean (2011) "Search, Choice and Revealed Preference", *Theoretical Economics* 6: 19-48.

[4] Caplin, A., M. Dean and D. Martin (2011) "Search and Satisficing", *American Economic Review* 101(7): 2899-2922.

[5] Corbin, R. and A.A.J Marley (1974) "Random Utility Models with Equality: An Apparent, but Not Actual, Generalization of Random Utility Models", *Journal of Mathematical Psychology*, 11: 274-293

[16] See the previous section.

[6] Dillenberger, D., K. Rozen and A. Postlewaite (2012) “Optimism and Pessimism with Expected Utility”, working paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2195566.

[7] Echenique, F., K. Saito and Tserenjigmid (2013) “The Perception-Adjusted Luce Model”, mimeo, California Institute of Technology.

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

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

[10] Gerasimou, G. (2014) “Asymmetric Dominance, Deferral and Status Quo Bias in a Behavioral Model of Choice”, mimeo, University of St Andrews.

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

[12] Krajbich, I. and A. Rangel (2011) “Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions”, *Proceedings of the National Academy of Sciences*, 108, 13852-13857.

[13] Kreps, D. M. (1979) "A Representation Theorem for ‘Preference for Flexibility’", *Econometrica*, 47: 565-577.

[14] Kreps, D.M. (2012) *Microeconomic Foundations I: Choice and Competitive Markets*, Princeton University Press, Princeton.

[15] Mandler, M., P. Manzini and M. Mariotti (2012) “One Million Answers to Twenty Questions: Choosing by Checklist”, *Journal of Economic Theory*, 147: 71-92.

[16] Manzini, P. and M. Mariotti (2014) “Stochastic Choice and Consideration Sets", *Econometrica* 82(3): 1153-76.

[17] Masatlioglu, Y., D. Nakajima and E. Ozbay (2012): “Revealed Attention”, *American Economic Review* 102(5): 2183–2205.

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

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

[20] Payro, F. and L. Ülkü (2014) “Similarity-Based Mistakes in Choice”, mimeo, ITAM.

[21] Reutskaya, E., R. Nagel, C. F. Camerer, and A. Rangel (2011) “Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study”, *The American Economic Review*, 101:900-926.

[22] Simon, H. (1955) “A Behavioral Model of Rational Choice”, *Quarterly Journal of Economics*, 69(1): 99-118.

[23] Sovinsky Goeree, M. (2008) “Limited Information and Advertising in the US Personal Computer Industry”, *Econometrica* 76: 1017-1074.

A Appendix

A.1 Proof of theorem 1

Proof. Necessity. Suppose the representation holds. Let $a \notin x$ and $a \succsim b$ for all $b \in x$. A4 holds obviously. To check A5, observe that $x \cup \{a\} = \{b \in x : b \hat{\succ} a\} \cup \{a\} \cup \{b \in x : a \hat{\succ} b\}$. Then we have

$$\begin{aligned}
u(x \cup \{a\}) - u(x) &= \sum_{b \in x : b \hat{\succ} a} \prod_{c \in x : c \hat{\succ} b} (1 - \alpha(c)) \alpha(b) u(b) \\
&\quad + \alpha(a) u(a) \prod_{b \in x : b \hat{\succ} a} (1 - \alpha(b)) \\
&\quad + (1 - \alpha(a)) \sum_{b \in x : a \hat{\succ} b} \prod_{c \in x : c \hat{\succ} b} (1 - \alpha(c)) \alpha(b) u(b) - \\
&\quad + (1 - \alpha(a)) \prod_{b \in x} (1 - \alpha(b)) u(\varnothing) \\
&\quad - \sum_{b \in x} \prod_{c \in x : c \hat{\succ} b} (1 - \alpha(c)) \alpha(b) u(b) - \prod_{b \in x} (1 - \alpha(b)) u(\varnothing) \\
&= \alpha(a) \prod_{b \in x : b \hat{\succ} a} (1 - \alpha(b)) \Bigg(u(a) - \sum_{b \in x : a \hat{\succ} b} \prod_{c \in x : a \hat{\succ} c \hat{\succ} b} (1 - \alpha(c)) \alpha(b) u(b) + \\
&\quad - \prod_{b \in x : a \hat{\succ} b} (1 - \alpha(b)) u(\varnothing) \Bigg) \\
&> 0
\end{aligned}$$

where the last inequality follows from the fact that $u(a) \geq u(b)$ for all $b \in x$ such that $a \hat{\succ} b$ and that the sum of the coefficients in the last two terms add up to less than unity.[17]

For sufficiency, let $\succ'$ be an arbitrary linear order on X and define $\hat{\succ}$ lexicographically as follows: $a \hat{\succ} b$ iff $a \succ b$ or $a \sim b$ and $a \succ' b$. Until further notice it is convenient to denote menus by numbering the alternatives in them according to $\hat{\succ}$, as $x = \{a_1, ..., a_K\}$ with $a_i \hat{\succ} a_{i+1}$ for all $i = 1, ..., K-1$. We will first show that for all $x \in \mathcal{X}$

[17] This holds since

$$\begin{aligned}
\sum_{b \in x : a \hat{\succ} b} \prod_{c \in x : a \hat{\succ} c \hat{\succ} b} (1 - \alpha(c)) \alpha(b) + \prod_{b \in x : a \hat{\succ} b} (1 - \alpha(b)) &< \\
\sum_{b \in x} \prod_{c \in x : c \hat{\succ} b} (1 - \alpha(c)) \alpha(b) + \prod_{b \in x} (1 - \alpha(b)) &= 1
\end{aligned}$$

there exist numbers $\alpha(a_1, x), ..., \alpha(a_K, x) \in (0,1)$ such that

$$\begin{aligned} x &\sim \alpha(a_1,x)\,a_1 + (1-\alpha(a_1,x))\,\alpha(a_2,x)\,a_2 + ... + \prod_{i=1}^{K-1}(1-\alpha(a_i,x))\,\alpha(a_K,x)\,a_K \\ &\quad + \prod_{i=1}^{K}(1-\alpha(a_i,x))\,\varnothing \end{aligned} \tag{4}$$

If x consists of only one element, then by A0 and A4 $a_1 \succ \{a_1\} \succ \varnothing$. By the vNM axioms and textbook arguments,

$$\{a_1\} \sim \alpha(a_1, \{a_1\})\,a_1 + (1-\alpha(a_1,\{a_1\}))\,\varnothing$$

for some unique $\alpha(a_1, \{a_1\}) \in (0,1)$ and so the result holds. Suppose then that x consists of two or more elements. We argue by induction on the cardinality of the menu, supposing that the assertion is true for all menus with fewer than K elements and letting $x = \{a_1, ..., a_K\}$ (where recall that $a_i \hat{\succsim} a_{i+1}$ for all $i = 1, ..., K-1$).

We have $a_1 \succ x \succ x \backslash \{a_1\}$ by A4 and A5. Therefore by the vNM axioms there exists a unique $\alpha(a_1, x) \in (0,1)$ such that

$$x \sim \alpha(a_1,x)\,a_1 + (1-\alpha(a_1,x))\,x\backslash\{a_1\}.$$

By the inductive hypothesis, there exist $\alpha(a_2, x\backslash\{a_1\}), ..., \alpha(a_K, x\backslash\{a_1\}) \in (0,1)$ such that

$$\begin{aligned} x\backslash\{a_1\} &\sim \alpha(a_2, x\backslash\{a_1\})\,a_2 + ... + \prod_{i=2}^{K-1}(1-\alpha(a_i, x\backslash\{a_1\}))\,\alpha(a_K, x\backslash\{a_1\})\,a_K \\ &\quad + \prod_{i=2}^{K}(1-\alpha(a_i, x\backslash\{a_1\}))\,\varnothing \end{aligned}$$

and so by Independence the desired conclusion follows by setting $\alpha(a_i, x) = \alpha(a_i, x\backslash\{a_1\})$ for all $i = 2, ..., K$. Then by the vNM theorem and A1-A3 there exists a vNM utility u on $\Delta(X \cup \mathcal{X})$ representing $\succsim$ such that

$$\begin{aligned} u(x) &= u\left(\alpha_1(a_1,x)\,a_1 + ... + \prod_{i=1}^{K-1}(1-\alpha(a_i,x))\,\alpha(a_K,x)\,a_K + \prod_{i=1}^{K}(1-\alpha(a_i,x))\,\varnothing\right) \\ &= \sum_{i=1}^{K}\prod_{j=1}^{i-1}\left(1-\alpha\left(a_j,x\right)\right)\alpha(a_i,x)\,u(a_i) + \prod_{i=1}^{K}(1-\alpha(a_i,x))\,u(\varnothing) \end{aligned}$$

(we use the convention that $\prod_{i=m}^{n} f(i) = 1$ and $\sum_{i=m}^{n} f(i) = 0$ for all functions $f : \mathbb{N} \to (0,1)$ whenever $m > n$).

Finally, we need to prove that α is menu independent in this representation. Returning to the generic (unnumbered) notation for menus, we have seen already that, using the construction above:

Claim 1: For all $x \in \mathcal{X}$: $a \notin x$ and $a \succsim b$ for all $b \in x \Rightarrow \alpha(b, x) = \alpha(b, x \cup \{a\})$ for all $b \in x$.

Next, we show:

Claim 2: For all $x, y \in \mathcal{X}$: $a \notin x \cup y$ and $a \succsim b$ for all $b \in x \cup y \Rightarrow \alpha(a, x) = \alpha(a, y)$.

Proof. Since by A0 and A4 $a \succ \{a\} \succ \varnothing$, by the vNM axioms there exists a unique $\alpha_{a,\{a\}} \in (0,1)$ such that

$$\{a\} \sim \alpha_{a,\{a\}} a + \left(1 - \alpha_{a,\{a\}}\right) \varnothing$$

Similarly, considering any $x \in \mathcal{X}$ such that $a \notin x$ and $a \succsim b$ for all $b \in x$, it follows by A4 and A5 that $a \succ \{a\} \cup x \succ x$. By A2 then there exist a unique $\alpha_{a,x} \in (0,1)$ such that

$$\{a\} \cup x \sim \alpha_{a,x} a + \left(1 - \alpha_{a,\{a\}}\right) x$$

By A3 it must be

$$\begin{aligned}
& k\{a\} + (1-k)[\{a\} \cup x] \\
& \sim k\left[\alpha_{a,\{a\}} a + \left(1 - \alpha_{a,\{a\}}\right) \varnothing\right] + (1-k)\left[\alpha_{a,x} a + \left(1 - \alpha_{a,\{a\}}\right) x\right] \\
& = \left(k\alpha_{a,\{a\}} + (1-k)\alpha_{a,x}\right) a + k\left(1 - \alpha_{a,\{a\}}\right) \varnothing + (1-k)\left(1 - \alpha_{a,\{a\}}\right) x \qquad (5)
\end{aligned}$$

for any $k \in (0,1)$. Fix one such k. Since it is also the case (by A4) that $a \succ \{a\}$ and $a \succ \{a\} \cup x$, then $a \sim ka + (1-k)a \succ k\{a\} + (1-k)(\{a\} \cup x)$ by A3. In addition, also by A3, $k\{a\} + (1-k)(\{a\} \cup x) \succ k\varnothing + (1-k)x$. Therefore

$$a \succ k\{a\} + (1-k)(\{a\} \cup x) \succ k\varnothing + (1-k)x$$

and, by A2 there exists a unique $\gamma \in (0,1)$ such that

$$k\{a\} + (1-k)(\{a\} \cup x) \sim \gamma a + (1-\gamma)[k\varnothing + (1-k)x]$$

But this is simply expression (5), so that it must be

$$\left.\begin{array}{r}\left(k\alpha_{a,\{a\}}+(1-k)\,\alpha_{a,x}\right)=\gamma\\ k\left(1-\alpha_{a,\{a\}}\right)=k\,(1-\gamma)\\ (1-k)\left(1-\alpha_{a,\{a\}}\right)=(1-\gamma)\,(1-k)\end{array}\right\}\Leftrightarrow\alpha_{a,\{a\}}=\alpha_{a,x}$$

Applying the same argument to a $y\in\mathcal{X}$ such that $a\notin y$ and $a\succsim b$ for all $b\in y$ yields $\alpha_{a,y}=\alpha_{a,\{a\}}=\alpha_{a,x}$, proving the claim. □

To conclude the proof of sufficiency, take $x,y\in\mathcal{X}$ and $a\in x\cap y$ (if $a\notin x\cap y$ for all a then there is nothing to prove). Let $x_L=\{b\in x:a\succ b\}$, enumerate arbitrarily the elements other than a in $x\backslash x_L$, that is let $x\backslash x_L=\{a,b_1,...b_n\}$, and let $x_i=x_L\cup\{b_1,...b_i\}$ for all $i=1,...n$ where $n=|x\backslash x_L|-1$. Similarly, let $y_L=\{c\in y:a\succ c\}$, $y\backslash y_L=\{a,c_1,...c_m\}$ and let $y_i=y_L\cup\left\{c_1,...c_j\right\}$ for all $j=1,...m$ where $m=|y\backslash y_L|-1$. Claim 2 implies that $\alpha\,(a,\{a\}\cup x_L)=\alpha\,(a,\{a\}\cup y_L)$. By Claim 1 we have that $\alpha\,(a,\{a\}\cup x_L)=\alpha\,(a,\{a\}\cup x_1)$, and by induction $\alpha\,(a,\{a\}\cup x_i)=\alpha\,(a,\{a\}\cup x_{i+1})$ for all $i\leq n-1$, where of course $x=\{a\}\cup x_n$, so that

$$\alpha\,(a,x)=\alpha\,(a,\{a\}\cup x_L)=\alpha\,(a,\{a\}\cup y_L)\tag{6}$$

A similar reasoning applied to $\alpha\,(a,\{a\}\cup y_L)$ yields

$$\alpha\,(a,y)=\alpha\,(a,\{a\}\cup y_L)=\alpha\,(a,\{a\}\cup x_L)\tag{7}$$

and then by (6) and (7) we conclude $\alpha\,(a,x)=\alpha\,(a,y)$. ■

A.2 Proof of theorem 2

Proof. Necessity. Suppose $\succsim$ on $\Delta\,(X\cup\mathcal{X})$ has an attention representation $(\hat{\succ},u,\alpha)$ in which α satisfies the monotonicity condition. Since $\alpha\,(a,x)\in(0,1)$ for all $x\in\mathcal{X}$ and $a\in x$, it follows

$$u\,(x)=\sum_{a\in x}\prod_{b\in x:b\hat{\succ}a}(1-\alpha\,(b,x))\,\alpha\,(a,x)\,u\,(a)+\prod_{a\in x}(1-\alpha\,(a,x))\,u\,(\varnothing)>u\,(\varnothing)$$

so that A0 holds. The necessity of A1-A3 is standard and thus omitted. Finally, let $a \succsim b$ for all $b \in x$. Then

$$u(x) = \sum_{c\in x} \prod_{c\in x: c\hat{\succ} b} (1-\alpha(c,x))\,\alpha(b,x)\,u(b) < u(a)$$

since the left hand side on the inequality is a convex combination of values which do not exceed $u(a)$ and the sum of the weights on $\max_{b\in x} u(b) \leq u(a)$ is strictly less than unity (given that $\prod_{c\in x}(1-\alpha(c,x))$, the weight on $u(\varnothing)$, is strictly positive), so that A4 holds.

For sufficiency, we can proceed as in the proof of theorem 1 if $x \succ x\setminus\{a_1\}$. Otherwise, we need to consider the two alternative cases.

Case 1: $x\setminus\{a_1\} \succ x$. Together with A0 this implies $x\setminus\{a_1\} \succ x \succ \varnothing$, and by the vNM axioms there exists a unique $\beta \in (0,1)$ with

$$x \sim \beta x\setminus\{a_1\} + (1-\beta)\varnothing. \tag{8}$$

Moreover by A4 $a_1 \succ x\setminus\{a_1\} \succ x$, so that there exists a unique $\alpha \in (0,1)$ such that

$$x\setminus\{a_1\} \sim \alpha a_1 + (1-\alpha)x.$$

Having defined α and β in this way, we claim that equation (4) (with the stated properties on the coefficients) holds by setting the coefficients recursively as follows:

$$\begin{aligned}
\alpha(a_1,x) &= \alpha\beta \\
\alpha(a_k,x) &= \gamma\alpha(a_k, x\setminus\{a_1\}) \text{ with} \\
\gamma &= \frac{(1-\alpha)\beta^2 \prod_{i=2}^{k-1}(1-\alpha(a_i,x\setminus\{a_1\}))}{1-\alpha\beta-(1-\alpha)\beta^2\left(\sum_{i=2}^{k-1}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)}
\end{aligned} \tag{9}$$

for all $k = 2, ...K$.

Step 1: the $\alpha(a_k,x)$ defined in equation (9) satisfy expression (4). By Independence applied to formula (8), given the definition of α, it must be:

$$x \sim \beta\alpha a_1 + \beta(1-\alpha)x + (1-\beta)\varnothing. \tag{10}$$

In turn, using the expression for x from condition (8) and Independence in expression (10) we have:

$$x \sim \alpha\beta a_1 + (1-\alpha)\beta^2 x\setminus\{a_1\} + (1-\beta)(1+\beta(1-\alpha))\varnothing$$

so that by the inductive hypothesis and Independence:

$$\begin{aligned} x \sim\ & \alpha\beta a_1 + (1-\alpha)\beta^2\left(\alpha(a_2, x\setminus\{a_1\})a_2 + ... + \prod_{i=2}^{K-1}(1-\alpha(a_i, x\setminus\{a_1\}))\alpha(a_K, x\setminus\{a_1\})a_K\right) \\ & + (1-\beta)(1+\beta(1-\alpha))\varnothing + (1-\alpha)\beta^2\prod_{i=2}^{K}(1-\alpha(a_i, x\setminus\{a_1\}))\varnothing \end{aligned} \tag{11}$$

Using (9) for any $k \geq 2$, we have

$$1-\alpha(a_k, x) = 1 - \frac{(1-\alpha)\beta^2\prod_{i=2}^{k-1}(1-\alpha(a_i,x\setminus\{a_1\}))}{1-\alpha\beta-(1-\alpha)\beta^2\left(\sum_{i=2}^{k-1}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)}\alpha(a_k, x\setminus\{a_1\}) =$$

$$= \frac{1-\alpha\beta-(1-\alpha)\beta^2\left(\sum_{i=2}^{k-1}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)-(1-\alpha)\beta^2\left(\prod_{i=2}^{k-1}(1-\alpha(a_i,x\setminus\{a_1\}))\right)\alpha(a_k,x\setminus\{a_1\})}{1-\alpha\beta-(1-\alpha)\beta^2\left(\sum_{i=2}^{k-1}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)} =$$

$$= \frac{1-\alpha\beta-(1-\alpha)\beta^2\left(\sum_{i=2}^{k-1}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)+\alpha(a_k,x\setminus\{a_1\})\prod_{i=2}^{k-1}(1-\alpha(a_i,x\setminus\{a_1\}))\right)}{1-\alpha\beta-(1-\alpha)\beta^2\left(\sum_{i=2}^{k-1}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)} =$$

$$= \frac{1-\alpha\beta-(1-\alpha)\beta^2\sum_{i=2}^{k}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)}{1-\alpha\beta-(1-\alpha)\beta^2\left(\sum_{i=2}^{k-1}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)}$$

so that the numerator of $1-\alpha(a_k, x)$ is equal to the denominators of $1-\alpha(a_{k+1}, x)$ and $\alpha(a_{k+1}, x)$. Consequently, the product $\alpha(a_k, x)\prod_{i=1}^{k-1}(1-\alpha(a_i, x))$ is a telescoping product, yielding

$$\alpha(a_k, x)\prod_{i=1}^{k-1}(1-\alpha(a_i, x)) = (1-\alpha)\beta^2\prod_{i=2}^{k-1}(1-\alpha(a_i, x\setminus\{a_1\}))$$

which is precisely the coefficient of a_k in the lottery on the right hand side of (11). Note (see Step 3 below) that in this case monotonicity in consideration is satisfied with inequality.

Step 2: $\alpha(a_k, x) > 0$ for all $k = 2, ...K$. It is obvious that $\alpha(a_1, x) > 0$ given the admissible values of α and β. For $k = 2, ...K$, note that the numerator is positive, and that (given the admissible values of α and β) we have $0 < (1-\alpha)\beta^2 < (1-\alpha\beta)$. So the denominator is positive given that

$$\sum_{i=2}^{k-1} \alpha\left(a_i, x\backslash\{a_1\}\right) \prod_{j=2}^{i-1}\left(1-\alpha\left(a_j, x\backslash\{a_1\}\right)\right) < 1$$

To prove this last inequality, observe that (keeping an eye on the summation and product indexes):

$$\begin{aligned}
&\sum_{i=2}^{k-1} \alpha\left(a_i, x\backslash\{a_1\}\right) \prod_{j=2}^{i-1}\left(1-\alpha\left(a_j, x\backslash\{a_1\}\right)\right) < 1 \\
\Leftrightarrow\ & \alpha\left(a_2, x\backslash\{a_1\}\right) + \sum_{i=3}^{k-1} \alpha\left(a_i, x\backslash\{a_1\}\right) \prod_{j=2}^{i-1}\left(1-\alpha\left(a_j, x\backslash\{a_1\}\right)\right) < 1 \\
\Leftrightarrow\ & \sum_{i=3}^{k-1} \alpha\left(a_i, x\backslash\{a_1\}\right) \prod_{j=2}^{i-1}\left(1-\alpha\left(a_j, x\backslash\{a_1\}\right)\right) < 1-\alpha\left(a_2, x\backslash\{a_1\}\right) \\
\Leftrightarrow\ & \sum_{i=3}^{k-1} \alpha\left(a_i, x\backslash\{a_1\}\right) \prod_{j=3}^{i-1}\left(1-\alpha\left(a_j, x\backslash\{a_1\}\right)\right) < 1 \\
\Leftrightarrow\ & \alpha\left(a_3, x\backslash\{a_1\}\right) + \sum_{i=4}^{k-1} \alpha\left(a_i, x\backslash\{a_1\}\right) \prod_{j=3}^{i-1}\left(1-\alpha\left(a_j, x\backslash\{a_1\}\right)\right) < 1 \\
\Leftrightarrow\ & \sum_{i=4}^{k-1} \alpha\left(a_i, x\backslash\{a_1\}\right) \prod_{j=3}^{i-1}\left(1-\alpha\left(a_j, x\backslash\{a_1\}\right)\right) < 1-\alpha\left(a_3, x\backslash\{a_1\}\right) \\
\Leftrightarrow\ & \sum_{i=4}^{k-1} \alpha\left(a_i, x\backslash\{a_1\}\right) \prod_{j=4}^{i-1}\left(1-\alpha\left(a_j, x\backslash\{a_1\}\right)\right) < 1 \\
\Leftrightarrow\ & \ldots \\
\Leftrightarrow\ & \alpha\left(a_{k-1}, x\backslash\{a_1\}\right)\left(1-\alpha\left(a_{k-1}, x\backslash\{a_1\}\right)\right) < 1
\end{aligned}$$

where the last inequality holds true by the inductive hypothesis, since $|x\backslash\{a_1\}| = K-1$).

Step 3: $\alpha(a_k, x) < 1$. It is obvious that $\alpha(a_1, x) < 1$ given the admissible values of α and β. For the other coefficients we show that

$$\frac{\alpha(a_k, x)}{\alpha\left(a_k, x\backslash\{a_1\}\right)} < 1 \text{ for all } k \leq K, \tag{12}$$

which implies the result (since $\alpha\left(a_k, x\backslash\{a_1\}\right) < 1$ by the inductive hypothesis on the cardinality of x). We proceed by induction on k (given K). If $k=2$, then from the second line in (9) we have

$$\frac{\alpha(a_2, x)}{\alpha\left(a_2, x\backslash\{a_1\}\right)} = \frac{(1-\alpha)\beta^2}{1-\alpha\beta} < 1.$$

Now suppose that $\frac{\alpha(a_k,x)}{\alpha(a_k,x\setminus\{a_1\})} < 1$ for all k for which $2 \leq k \leq k'-1 < K$, and consider $k = k'$. Then

$$\frac{\alpha\left(a_{k'},x\right)}{\alpha\left(a_{k'},x\setminus\{a_1\}\right)} = \frac{(1-\alpha)\beta^2\prod_{i=2}^{k'-1}(1-\alpha(a_i,x\setminus\{a_1\}))}{1-\alpha\beta-(1-\alpha)\beta^2\left(\sum_{i=2}^{k'-1}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)} < 1$$

$$\Leftrightarrow (1-\alpha)\,\beta^2\prod_{i=2}^{k'-1}\left(1-\alpha\left(a_i,x\setminus\{a_1\}\right)\right)$$

$$< 1-\alpha\beta-(1-\alpha)\,\beta^2\left(\sum_{i=2}^{k'-1}\alpha\left(a_i,x\setminus\{a_1\}\right)\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)$$

$$\Leftrightarrow (1-\alpha)\,\beta^2\left(1-\alpha\left(a_{k'-1},x\setminus\{a_1\}\right)\right)\prod_{i=2}^{k'-2}\left(1-\alpha\left(a_i,x\setminus\{a_1\}\right)\right)$$

$$< 1-\alpha\beta-(1-\alpha)\,\beta^2\left(\begin{array}{c}\sum_{i=2}^{k'-2}\alpha\left(a_i,x\setminus\{a_1\}\right)\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\\ +\alpha\left(a_{k'-1},x\setminus\{a_1\}\right)\prod_{j=2}^{k'-2}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\end{array}\right)$$

$$\Leftrightarrow (1-\alpha)\,\beta^2\prod_{i=2}^{k'-2}\left(1-\alpha\left(a_i,x\setminus\{a_1\}\right)\right)$$

$$< 1-\alpha\beta-(1-\alpha)\,\beta^2\left(\sum_{i=2}^{k'-2}\alpha\left(a_i,x\setminus\{a_1\}\right)\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)$$

$$\Leftrightarrow \frac{(1-\alpha)\beta^2\prod_{i=2}^{k'-2}(1-\alpha(a_i,x\setminus\{a_1\}))}{1-\alpha\beta-(1-\alpha)\beta^2\left(\sum_{i=2}^{k'-2}\alpha(a_i,x\setminus\{a_1\})\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j,x\setminus\{a_1\}\right)\right)\right)} = \frac{\alpha\left(a_{k'-1},x\right)}{\alpha\left(a_{k'-1},x\setminus\{a_1\}\right)} < 1$$

where $\frac{\alpha\left(a_{k'-1},x\right)}{\alpha\left(a_{k'-1},x\setminus\{a_1\}\right)} < 1$ holds by the inductive hypothesis on k. Thus, condition (12) holds.

Case 2: $x\setminus\{a_1\} \sim x$. Then $a_1 \succ x \succ \varnothing$ and A2 imply that there exists a unique $\alpha \in (0,1)$ with

$$x \sim \alpha a_1 + (1-\alpha)\,\varnothing.$$

Applying Independence repeatedly, the above and $x\setminus\{a_1\} \sim x$ imply that, for all $\beta \in [0,1]$,

$$x \sim \beta\left(\alpha a_1 + (1-\alpha)\,\varnothing\right) + (1-\beta)\,x\setminus\{a_1\}$$

so that by the inductive hypothesis

$$
\begin{aligned}
x &\sim \alpha\beta a_1 + (1-\beta)\left(\alpha\left(a_2, x\backslash\{a_1\}\right)a_2 + ... + \prod_{i=2}^{K-1}\left(1-\alpha\left(a_i, x\backslash\{a_1\}\right)\right)\alpha\left(a_K, x\backslash\{a_1\}\right)a_K\right) \\
&\quad +\beta(1-\alpha)\varnothing + (1-\beta)\prod_{i=2}^{K}\left(1-\alpha\left(a_i, x\backslash\{a_1\}\right)\right)\varnothing \qquad (13)
\end{aligned}
$$

Fix β so that $\beta \in (0,1)$. Then, similarly to case 1, with α and β so defined condition (4) (with the stated properties on the coefficients) holds by setting recursively

$$
\begin{aligned}
\alpha(a_1, x) &= \alpha\beta \qquad (14) \\
\alpha(a_k, x) &= \frac{(1-\beta)\prod_{i=2}^{k-1}\left(1-\alpha\left(a_i, x\backslash\{a_1\}\right)\right)}{1-\alpha\beta-(1-\beta)\left(\sum_{i=2}^{k-1}\alpha\left(a_i, x\backslash\{a_1\}\right)\prod_{j=2}^{i-1}\left(1-\alpha\left(a_j, x\backslash\{a_1\}\right)\right)\right)}\alpha\left(a_k, x\backslash\{a_1\}\right)
\end{aligned}
$$

A straightforward adaptation of Step 2 and Step 3 in the proof of case 1 shows that $\alpha(a_k, x) \in (0,1)$ for all $k = 1, ...K$. To see that (14) retrieves the coefficients in (4) correctly, again a straightforward adaptation of the proof of Step 1 in case 1 shows that the product $\alpha(a_k, x)\prod_{i=1}^{k-1}(1-\alpha(a_i, x))$ is a telescoping product, yielding

$$
\alpha(a_k, x)\prod_{i=1}^{k-1}(1-\alpha(a_i, x)) = (1-\beta)\prod_{i=2}^{k-1}\left(1-\alpha\left(a_i, x\backslash\{a_1\}\right)\right)
$$

namely the coefficient of a_k in the lottery on the right hand side of (4). ■

A.3 An example of the non-uniqueness of the attention parameters with menu-dependence

Let $X = \{a, b\}$, suppose that preferences satisfy A0-A4 and are such that $a \succ b \succ \{a, b\} \succ \{a\} \succ \{b\}$, and suppose that u represents preferences with

$X \cup \mathcal{X}$	a	b	$\{a,b\}$	$\{a\}$	$\{b\}$	$\varnothing$
$u(\cdot)$	U	pU	qpU	rU	sU	0

where $U > 0$, $p, q, r, s \in (0,1)$ and $s < r < qp$. Since $\alpha(a, \{a\})U = u(\{a\}) = rU$ and

$\alpha(b,\{b\})\, pU = u(\{b\}) = sU$ we determine the parameters

$$\begin{aligned} \alpha(a,\{a\}) &= r \\ \alpha(b,\{b\}) &= \frac{s}{p} \end{aligned}$$

The other constraint is

$$\begin{aligned} & \alpha(a,X)\,U + (1-\alpha(a,X))\,\alpha(b,X)\,pU = u(X) = qpU \\ \Leftrightarrow \quad & \alpha(b,X) = \frac{qp-\alpha(a,X)}{(1-\alpha(a,X))\,p} \end{aligned}$$

Since $\alpha(b,X) \in (0,1)$, it must be that

$$\frac{qp-\alpha(a,X)}{(1-\alpha(a,X))\,p} \in (0,1) \Leftrightarrow \alpha(a,X) < qp$$

(observing that the numerator is less than the denominator if and only if $(q-1)\,p < (1-p)\,\alpha(a,X)$, which holds true always). Moreover, since the monotonicity condition on α imposes that $\alpha(b,X) \leq \alpha(b,\{b\})$, it must also be

$$\frac{qp-\alpha(a,X)}{(1-\alpha(a,X))\,p} \leq \frac{s}{p} \Leftrightarrow \alpha(a,X) \geq \frac{qp-s}{1-s}$$

In short, then, we have the restriction

$$\alpha(a,X) \in \left[\frac{qp-s}{1-s}, qp\right)$$

As we have used all the conditions, we cannot pin down a unique value for $\alpha(a,X)$.

A.4 An example of the effect of $\hat{\succ}$ on the attention function

Let $X = \{a,b\}$, $a \sim b \succ \{a,b\} \succ \{a\} \sim \{b\}$, with u representing these preferences and defined as

$X \cup \mathcal{X}$	a	b	$\{a,b\}$	$\{a\}$	$\{b\}$	$\varnothing$
$u(\cdot)$	U	U	pU	pqU	pqU	0

with $U > 0$, $p, q \in (0,1)$.

Since $\alpha(a,\{a\})U = u(\{a\}) = u(\{b\}) = \alpha(b,\{b\})U = pqU$ we determine the parameters $\alpha(a,\{a\}) = pq = \alpha(b,\{b\})$. The other constraint is

$$\begin{aligned} & \alpha(a,X)U + (1-\alpha(a,X))\alpha(b,X)U = u(X) = pU \\ \Leftrightarrow \quad & \alpha(b,X) = \frac{p-\alpha(a,X)}{1-\alpha(a,X)} \end{aligned} \tag{15}$$

with $\alpha(a,X) < p$ to ensure $\alpha(b,X) > 0$.

Suppose first that $a\hat{\succ}b$. As in example A.3, since the monotonicity condition on α requires that $\alpha(b,X) \leq \alpha(b,\{b\})$, it must also be

$$\frac{p-\alpha(a,X)}{1-\alpha(a,X)} \leq pq \Leftrightarrow \alpha(a,\{a,b\}) \geq \frac{p(1-q)}{1-pq}$$

In short, then, if $a\hat{\succ}b$ we have $\alpha(a,\{a,b\}) \in \left[\frac{p(1-q)}{1-pq}, p\right) \neq \varnothing$ and $\alpha(b,X) = \frac{p-\alpha(a,X)}{1-\alpha(a,X)}$.

Now consider the alternative case $b\hat{\succ}a$. As the two attention parameters $\alpha(a,X)$ and $\alpha(b,X)$ are completely symmetric, we obtain $\alpha(b,X) \in \left[\frac{p(1-q)}{1-pq}, p\right)$ and $\alpha(a,X) = \frac{p-\alpha(b,X)}{1-\alpha(b,X)} \Leftrightarrow \alpha(b,X) = \frac{p-\alpha(a,X)}{1-\alpha(a,X)}$. That is, while equation (15) establishes the same condition regardless of whether $a\hat{\succ}b$ or $b\hat{\succ}a$, the monotonicity condition imposes a different range of values for the attention parameters. For instance, setting $p = 0.6$ and $q = 0.8$, $\alpha(a,X) = 0.2$ and $\alpha(b,X) = 0.5$, the requirements for the case $a\hat{\succ}b$ fail (since for that case $\alpha(a,X) \in [0.23, 0.6)$) while those for $b\hat{\succ}a$ hold (since $0.5 = \alpha(b,X) \in [0.23, 0.6)$ and $\alpha(a,X) = \frac{0.6-0.5}{1-0.5} = 0.2$).

More in general, for any $\varepsilon \in (0, \min\{1-p, p(1-q)\})$, the case $a\hat{\succ}b$ allows for $\alpha(a,X) = p-\varepsilon$ and $\alpha(b,X) = \frac{p-(p-\varepsilon)}{1-(p-\varepsilon)} = \frac{\varepsilon}{1-p+\varepsilon} > 0$; since however $\frac{\varepsilon}{1-p+\varepsilon} < \frac{p(1-q)}{1-pq}$ given our condition on ε, this value falls outside the range for $\alpha(b,X)$ when $b\hat{\succ}a$.

Finally, we note that taking e.g. the case $a\hat{\succ}b$, if we also required $\alpha(a,X) \leq \alpha(a,\{a\}) = pq$, we would have

$$\begin{aligned} \frac{p(1-q)}{1-pq} \quad < \quad & pq \Leftrightarrow \frac{1-q}{1-pq} < q \Leftrightarrow 1-2q+pq^2 < 0 \\ \Leftrightarrow \quad & p < \frac{2q-1}{q^2} \end{aligned}$$

If q is sufficiently small, the rhs in the last inequality is negative, so that the condition cannot hold. For instance with $U = 12$, $p = \frac{1}{4}$ and $q = \frac{1}{3}$ utilities are $u(\varnothing) = 0$,

$u(a) = u(b) = 12 > 0$, $u(\{a,b\}) = 3$, $u(\{a\}) = 1 = u(\{b\})$. Then $\alpha(a,\{a\}) = \alpha(b,\{b\}) = \frac{1}{12}$, while from

$$u(X) = 3 = \alpha(a,X)\,12 + (1-\alpha(a,X))\,\alpha(b,X)\,12$$

we obtain

$$\alpha(b,X) = \frac{1-4\alpha(a,X)}{4(1-\alpha(a,X))}$$

which is positive since by monotonicity it must be $\alpha(a,X) \leq \alpha(a,\{a\}) = \frac{1}{12} < \frac{1}{4}$. On the other hand, since by monotonicity we must also have $\alpha(b,X) \leq \frac{1}{12}$, it follows that

$$\frac{1-4\alpha(a,X)}{4(1-\alpha(a,X))} \leq \frac{1}{12} \Leftrightarrow \alpha(a,X) \geq \frac{2}{11} > \frac{1}{12}$$