A Salience Theory of Choice Errors *

Paola Manzini Marco Mariotti[†]

This version: July 2011

Abstract

A decision maker applies a preference ranking after forming a 'consideration set' of alternatives. Both alternative *salience* and agent *rationality* (his general propensity to consider all alternatives) determine the consideration set. The model includes an ordinal version of the standard logit formulation and it has a rich set of implications both when parameters are exogenous and when alternatives can affect their own salience (*salience games*). For example, an increase in rationality may lead to choosing a low quality alternative with a higher probability. In some salience games, salience reveals quality in equilibrium.

J.E.L. codes: D0.

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

*We thank Chris Flynn, Jacob Goeree, Rosa Matzkin, Mauro Papi, Ivan Soraperra, Rani Spiegler and Alex Tetenov for helpful comments and discussions, as well as the audience at the ESWC2010 in Shanghai and the Collegio Carlo Alberto in Turin. Financial support from ESRC grant RES-000-22-3474 is gratefully acknowledged.

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

1 Introduction

Decision makers make mistakes. A consumer buying a new PC is not aware of all the latest models and specifications and ends up making a selection he later regrets;[1] a doctor short of time for formulating a diagnosis overlooks the relevant disease for the given set of symptoms;[2] the most appropriate procedure escapes the attention of a pilot in the heat of an emergency. In these examples the decision-maker has no problem in *evaluating* the alternatives he considers (unlike, for example, a consumer who is uncertain about the quality of a product). Yet, for various reasons the agent fails to *consider* all relevant alternatives. This is the topic of this paper: we model an agent who with some probability makes an error - the agent fails to pick the optimal alternative from a set, because with some probability he neglects some alternatives.

The vast majority of theoretical models of economic choice assume that behaviour is deterministic. The primitive in these models is a choice function c that indicates the selection $c(A)$ the agent makes from any menu A. This holds true both for the classical 'rational' model of preference maximisation (Samuelson [41], Richter [36]) and for more recent models of boundedly rational choice. But empirical economists cannot escape confronting the noisiness of the data: individual choice responses typically exhibit variability, in both experimental and market settings. This raises the need, for practical applications, to graft an appropriate *error structure* on the model, and therefore leads to the construction of a probabilistic choice model. For this reason we develop a stochastic model of choice. The primitive is now a probability function p that indicates the probability $p(a_i, A)$ that alternative a_i is selected from a menu A. Pioneering theoretical contributions in this area have been Luce [23] and Block and Marshak's [7] and Marshak's [32] Random Utility Maximization (RUM) model. The RUM model culminated in its most influential version, McFadden's ([27], [28]) *multinomial logit* (or *conditional logit*) discrete choice model. While this specification is extremely versatile in applications, it lacks a firm psychological foundation. From this point of view, our model can be seen a simple model of probabilistic choice from discrete choice sets in which:

[1] Goeree [13] makes this argument and quantifies this phenomenon with empirical data.

[2] We discuss this medical case in more detail in section 3.

1) The error component of the model is derived from explicit psychological assumptions.

2) Some parameters governing those components are endogenised via an equilibrium process.

To make this point more precise, recall that in the multinomial logit model the probability $p(a_i, A)$ takes the form

$$p(a_i, A) = \exp\left(u\left(a_i\right)\right) / \sum_{a_j \in A} \exp\left(u\left(a_j\right)\right),$$

where $u(a_j)$ expresses the 'systematic utility' of alternative a_j.[3] This model is a case within a general class of additive RUM in which alternative a_i generates a 'random utility stimulus' $u(a_i) + \varepsilon_i$, where ε_i is an error term, and is chosen over alternative a_j if $u(a_i) - u(a_j) > \varepsilon_j - \varepsilon_i$. Different error specifications generate different models: the logit model follows from the ε_i being assumed i.i.d. Gumbel (or extreme value type I) distributed random variables.[4] A probit model would follow instead by assuming normal distributions. In general, the basic constraint is that larger errors are made with smaller probabilities, from which it follows that better alternatives are chosen with higher probability.[5]

Beside the basic constraint of error monotonicity in utility, the assumption of a specific error structure lacks a clear psychological foundation, and deviations from utility maximisation follow from the black box of exogenous random errors grafted on the utility term.[6]

[3] In applications, it is usually assumed further that utility is a linear function of the alternative attributes.

[4] Gumbel distribution function (with parameters μ and σ): $F(x) = \exp\left(-e^{\frac{x-\mu}{\sigma}}\right)$.

[5] Closely related ideas have also found their way in modelling strategic behaviour, for the first time with McKelvey and Palfrey's ([30], [31]) notion of Quantal Response Equilibrium (QRE) - see Goeree, Holt and Palfrey [14] for an overview.

[6] We recall that the Gumbel distribution function can be seen as the limit distribution function of (a suitable transformation of) the maximum value statistics for a sample of N i.i.d. random variables, as N tends to infinity. The statistics needs to be appropriately transformed when taking the limit since, obviously, letting G be the common distribution function of the random variables and G^M the distribution function of $\max(X_i)_{i=1,\ldots,N}$, we have $\lim_{N\to\infty}\left(G^M(x)\right) = \lim_{N\to\infty}\left(G(x)\right)^N = 0$ for any x unless $G(x) = 1$. Even this brief account of the error structure behind the conditional logit model makes it transparent how opaque its economic or psychological interpretation is.

In our approach, instead of imputing departures from rationality to unexplained errors in the perception of systematic utility, we formulate *directly* a model of boundedly rational choice which includes some more elementary stochastic components. In this way the error structure becomes fully transparent, being a core part of the theory of bounded rationality we are proposing. The stochastic components of the model are extremely simple and the variability around rational choice behaviour can be given a precise psychological interpretation. As we shall see, this framework, rather than contradicting the multinomial logit model, provides an (ordinal) extension of it.[7] Moreover, we view our approach as complementary to axiomatic justifications of choice[8] of the type started by Luce [23] - they formulate abstract a priori restrictions on the probabilistic choice function. The plausibility of such restrictions should be evaluated in concrete choice models (for example our model, which we think is plausible, fails Luce's axiom - see section 3.3).

Our setup hinges on the notion of a *consideration set*. As in the opening examples, an agent does not rationally evaluate *all* objectively available alternatives in A, but only a (possibly strict) subset of them, $C(A)$. This subset is the consideration set. Once a $C(A)$ has been formed, a final choice is made by maximising a preference relation over $C(A)$, which in this paper we assume to be standard (complete and transitive). This two-step conceptualisation of the act of choice is rooted and well-accepted in psychology and marketing science, but it has begun to diffuse in economics. Several recent models of boundedly rational choice adopt it in one way or the other (Manzini and Mariotti [24] and especially Masatlioglu, Nakajima and Ozbay [26] and Eliaz and Spiegler [11], [12]).

The composition of the consideration set $C(A)$ is stochastic. For each alternative the probability of being considered depends on two types of parameters (probabilities). The first parameter ρ ('*rationality*'), expresses the *general* propensity of the agent to consider all alternatives: this may depend for example on the time the agent has at his disposal to make a decision or on the effort put into it. The second type of parameter, σ_i ('*salience*'), captures all the *alternative-specific* factors that affect the probability of membership of the consideration set. We focus on salience as the main interpretation of σ_i because of

[7] Mattson and Weibull [33] provide an alternative model-based approach to probabilistic choice. Their work is discussed in in section 6..

[8] See notably the recent contribution by Gul, Natenzon and Pesendorfer [16], discussed in section 6.

its focal role, through the mechanism of *attention,* in human cognition (e.g., Anderson [3]): salience can be viewed as the success in attracting attention.[9] But σ_i can also be interpreted, in some contexts, to include other idiosyncratic factors affecting inclusion (such as psychological, political or racial biases). For example, voters may not consider certain parties as relevant for cultural reasons or out of habit.[10]

We present two arguably natural models (called AND and OR) that depend on the specific way the parameters combine to determine the probability of membership of $C(A)$. We show that for a special case (that of equal salience across alternatives) both models can be expressed in a logit format. However, we use only *ordinal* preference information. Contrast this with the logit model which uses, as explained above, inequalities of the type $u(a_i) - u(a_j) > \varepsilon_j - \varepsilon_i$. These inequalities are only invariant to common cardinal (affine) transformations of the u and the errors, and therefore contain cardinal information.

One advantage of expressing rationality in parametric form is that it makes it easy to study the impact of rationality on the probability of a given alternative being chosen. We find that some of the 'intuitive' properties of error do not hold. One might expect, for example, that as rationality increases the agent will tend to choose each nonoptimal alternative with lower and lower probability: but we show that this is not necessarily the case.[11] Both the AND and the OR models predict either monotonic or single peaked relationships between the rationality parameter ρ and the probability of choice, for all alternatives, no matter how 'bad'.[12] Unexpected effects may also occur (in the OR model)

[9] The role of salience and attention is also beginning to be incorporated in formal economic models of choice. In particular, Masatlioglu, Nakajima and Ozbay [26] offer a revealed preference method to detect attention, while Tyson [44] studies salience as a method of resolution of the 'pseudoindifferences' faced by a satisficing agent.

[10] See Wilson [47] for a consideration set approach to political competition. It is reported there that African Americans tend to ignore Republican candidates in spite of the overlap betwen their policy preferences and the stance of the Republicans, and even if they are dissatisfied with the Democratic candidate.

[11] In fact, there is also evidence that is consistent with choice probabilities falling as rationality increases. We report it in section 3.

[12] The AND model predicts, for all alternatives and all parameters, an interval (and possibly the full interval) in which the probability of choosing any given alternative increases with rationality. The OR model predicts an increasing interval for some alternatives and some parameter configurations, but forbids

in respect of the *odds* (probability ratio) of choosing a better alternative over a worse alternative, which may decrease with ρ. And in both models the choice odds fail Luce's [23] Independence of Irrelevant Alternatives test (which implies, together with other assumptions, the logit model).[13] As for salience, when it is exogenously determined, higher salience determines an increase in choice probability. This accords with evidence, as e.g. that reported in Ambler et. al [2], who find that in an incentivised experiment choice probability increased with salience.[14] However, in our model salience interacts with the other parameter in a non-obvious way.

But what determines the salience of an alternative? In the second part of the paper we endogenise salience: its value now arises as the equilibrium of some 'salience game', i.e. an abstract situation in which alternatives can influence their own prominence. There are many examples that fit this description. In electoral contests, politicians make statements to get noticed by the voters, not only to persuade them. In animal mate competition the alternatives are male animals, the chooser is a female, and salience is controlled via natural selection (e.g. endowing peacocks with more or less showy tails), or by human activities (hair-styling, body-building, wealth-accumulation). In an I.O. context the alternatives are products, and salience is controlled via marketing strategies (Eliaz and Spiegler's [11] work mentioned before is the first to study in detail this type of competitive situation).[15] We look for the Nash equilibria of a game in which alternatives select simultaneously the salience-determining variables. We show that in such games equilibria in pure strategies

increasingness on the whole range except for the best alternative.

[13]In its core version. The nested logit, for example, allow for violations of IIA. A probit model also allows for such violations. See e.g. Agresti [1] for an overview of statistical methods for categorical data.

[14]In their experiment, subjects were showed footage of a supermarket they had visited before, and taken on a virtual shopping trip, along which they had to make several choices out of a sequence of sets of three alternatives, for which they would receive a purchase voucher. The experimenters also asked subjects to indicate the tallest in various sets of packages (to distinguish between the cognitive efforts required by simpler choices), and also asked subjects to rank the various items shown in terms of familiarity, using these data to construct a salience index.

[15]Examples in our discipline of factors affecting salience might be the choice of research topic, or the title of a paper (an author wants to be read, reader has only a limited attention, catchy titles confer salience - we may imagine that 'choosing' here means 'remembering': the reader only puts effort in remembering the best quality paper among those he has considered).

exist under general conditions.

Our main result is that when alternatives can fully control their own salience ('absolute salience'), in equilibrium - under very general assumptions - both AND and OR specifications exhibit a 'the showiest is the best' feature: the equilibrium ordering of salience reflects the preference ordering over alternatives. However, when salience is relative (so that alternatives can control salience only partially), there exists fully perverse equilibria in both models. In such equilibria the worst alternative has the highest probability of being chosen.

The apparently counterintuitive nature of some of our results has an obvious bearing for the inferences we draw on true preferences using 'revealed preferences' reconstructed from choice data. We briefly comment on such implications in the concluding section.

2 Salience and rationality

2.1 The Model

There is a countable (possibly finite) choice set of alternatives $A = \{a_1, ..., a_n, ...\}$. This set is fixed throghout except when otherwise stated. The agent has a strict preference ordering $\succ$ on A. We will often refer to the position of an alternative in the ranking as its *quality*, with a lower i indicating a higher quality, so that $a_i \succ a_j$ iff $i < j$.

The preference $\succ$ is applied only to a *consideration set* $C(A) \subseteq A$ of alternatives (the set of alternatives he actively considers). We allow for $C(A)$ to be empty, in which case the chooser picks a default option a^* (e.g. walking away from the shop, remaining without a partner, abstaining from voting).

The membership of $C(A)$ for the alternatives in A is probabilistic. The probability of membership combines two components (probabilities): an idiosyncratic component $\sigma_i \in (0, 1)$ which is specific to the alternative and an alternative-independent component $\rho \in (0, 1)$. We call the probability σ_i the *salience* of alternative a_i, and a list $(\sigma_1, ..., \sigma_n, ...)$ a *salience profile*. As observed before, while we use the term salience throughout for simplicity and because it tallies with leading examples, there are situations in which σ_i is not associated with *awareness* of the alternative by the agent, but rather with the *resistance* of the agent to consider the alternative for choice (e.g. for ideological reasons).

The probability ρ measures the general propensity of the agent to consider all alternatives. All else being equal, an agent with a higher ρ is more likely to apply his preference to the entire choice set A and ρ can thus be interpreted as the agent's *degree of rationality.*

We consider two elementary probability models, both of which determine $C(A)$ in two stages, as depicted in figure 1. The sets at the terminal nodes indicate the destination of the alternative.

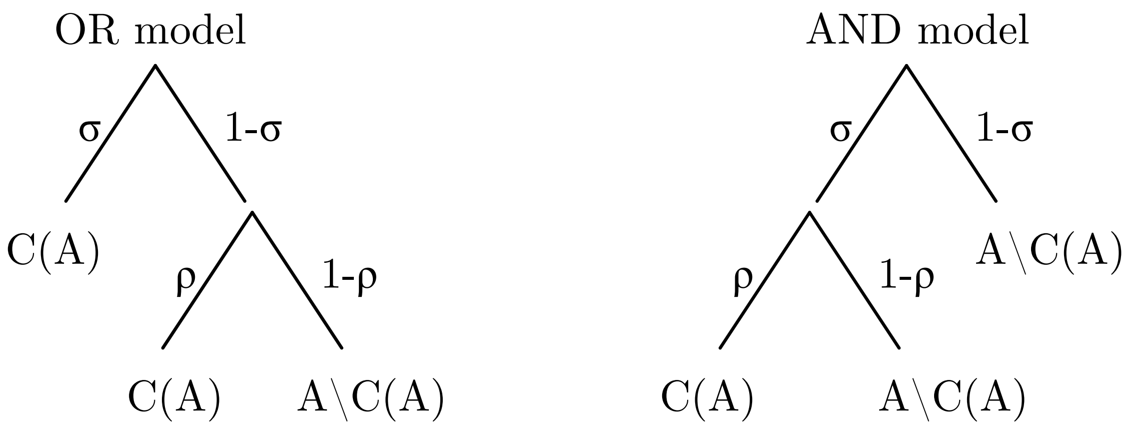

Figure 1: Structure of the two models

OR model: any alternative a_i is drawn into the consideration set with probability σ_i, and then all alternatives which haven't been drawn in this way are considered with probability ρ. So

$$prob\left(a_i \in C\left(A\right)\right) = \sigma_i + \left(1 - \sigma_i\right)\rho$$

AND model: any alternative a_i is provisionally drawn into the consideration set with probability σ_i, and it remains there with probability ρ. So

$$prob\left(a_i \in C\left(A\right)\right) = \rho\sigma_i$$

In both models, once the probabilistic consideration phase has been completed and a set $C(A)$ has been formed, the agent chooses (if $C(A)$ is nonempty) the alternative a_i with the properties that

$$a_i \in C\left(A\right) \text{ and } a_i \succ a_j \text{ for all } a_j \in C\left(A\right) \setminus \{a_i\}$$

If $C(A)$ is empty he chooses the default option a^*.[16]

Finally, observe that the models are invariant to permuting the order in which salience and rationality are applied.

2.2 Shopping as a goal or shopping with a goal? The OR model, the AND model and attention

The literature on salience and attention is vast and spans several related but different fields, from the the neurosciences to psychology and consumer research. Very many models have been proposed to capture the various aspects of the relationship between salience and attention, and we do not expect to settle the question here. We want however to briefly show how our simple AND and OR model can be grounded, at least qualitatively, in influential theories from these neighbouring fields. The reader uninterested in this justification can safely skip the section.

The 'load theory of selective attention and cognitive control' introduced by Lavie et al. [22] posits that attention is driven by the interaction of two mechanisms activated by the perceptual load imposed by the *relevant* 'stimuli': a relatively passive one that prevents irrelevant stimuli to be perceived; and an 'active' one that draws from cognitive functions to actively shut out irrelevant stimuli that are perceived. If the perceptual load is high enough, then all the irrelevant information is ignored by the decision maker; on the other hand if the load is not sufficiently high, then competing stimuli will have to be suppressed by the more active mechanism that draws from cognitive functions to achieve this objective. Put it differently, whereas attending to relevant stimuli can reduce distractor processing, the opposite happens if attention is clogged by other tasks (i.e. the agent is distracted, leaving room for salience to act).[17]

[16] The default option could be replaced by a more complex procedure to arrive at a choice, notably uniform randomisation over A.

[17] In the authors' words: "distractors can be excluded from perception when the level of perceptual load in processing task-relevant stimuli is sufficiently high to exhaust perceptual capacity, leaving none of this capacity available for distractor processing. However, in situations of low perceptual load, any spare capacity left over from the less demanding relevant processing will spill over to the processing of irrelevant distractors. Thus, in this model, early selection is predicted for situations of high perceptual

Let ρ denote an index of the agent's capability to shut out irrelevant stimuli: with high ρ the agent is able to prevent any of these from reaching him, but with a low ρ there is a lot of 'spare capacity' that allows irrelevant stimuli (i.e. salience) to play a role. The full range of cases is therefore approximated by the OR model, where

$$prob\left(a_i \in C\left(A\right)\right) = \rho + \left(1-\rho\right)\sigma_i = \sigma_i + \left(1-\sigma_i\right)\rho$$

Implicit in this setup is that rationality and salience contribute in analogous ways in directing an agent to attend to an alternative. This would apply to a situation of '*shopping as a goal*', where the decision maker is not after any particular item, and has an 'open mind' as to what he'll end up choosing.[18] On the other hand, there are other situations of '*shopping for a goal*' where the decision maker's choice is task directed, and we should expect some interaction between attention and salience: a venture capitalist must be able to recognize the profitability of an eye catching business opportunity, and the weary shopper looking for a birthday present must be able to recognise the suitability of an eye catching gift for the intended recipient. In these circumstances rationality and salience interact. There are many sophisticated ways in which this interaction could be modeled, and we do not even attempt to settle the debate raging in the neuroscience and psychological literature on how this interaction should be nuanced. We limit ourselves to arguing that the AND model provides a very simple setup to think of the interplay between salience and rationality which is qualitatively distinct from the set of situations for which the OR model is a more appropriate modeling tool.

load, whereas late selection is predicted for situations of low perceptual load". This excerpt refers to two competing theories, the so called late and early theories of attention. Put it in crude terms, the early view posits that focused attention (corresponding to rationality in our framework) will eliminate competing irrelevant stimuli, while the late view states that attention can only influence later and higher cognitive functions (rationality in our setup). See Kirschner, Ayres and Chandler [20] and references therein for a thorough review of these issues. See also Lavie [21].

[18] The 'shopping as a goal' vs. 'shopping with a goal' terminology was introduced in the highly influential Babin Darden and Griffin [4], who introduced the 'personal shopping value scale' based on 'hedonic' and 'utilitarian' values in shopping.

2.3 The Logit Retrouvé?

There is a surprising formal relation between our two models and the logit model. We begin by discussing a benchmark situation, in which both of our models collapse to a logit form. Let $p_{AND}(a_i, A)$ and $p_{OR}(a_i, A)$ denote the probabilities of choice conditional on the agent picking an element in A in the AND and in the OR model, respectively.

Proposition 1 *Suppose $\sigma = \sigma_i$ for all i for some $\sigma \in (0,1)$. Then there exists a utility function $u : A \to \mathbb{R}$ representing $\succ$ on A and coefficients $\alpha, \alpha', \beta, \beta'$ such that $p_{AND}(a_i, A)$ and $p_{OR}(a_i, A)$ can be written in logit form, that is*

$$\begin{aligned} p_{OR}(a_i, A) &= \exp(\alpha + \beta u(a_i)) / \sum_j \exp(\alpha + \beta u(a_j)) \\ p_{AND}(a_i, A) &= \exp(\alpha' + \beta' u(a_i)) / \sum_j \exp(\alpha' + \beta' u(a_j)), \end{aligned}$$

In the Appendix we prove a more general result, but it is useful to see why the statement holds by writing down the choice probabilities explicitly in the two models, which we will soon need anyway. In the OR model, the probability $p_{OR}(a_i)$ that $a_i \in A$ is chosen is:[19]

$$p_{OR}(a_i) = \underbrace{\sigma_i \prod_{j<i} \left((1-\sigma_j)(1-\rho)\right)}_{\substack{\text{probability that } a_i \text{ enters } C \text{ in the} \\ \text{first stage and the alternatives better} \\ \text{than } a_i \text{ do not enter } C}} + \underbrace{(1-\sigma_i)\rho \prod_{j<i} \left((1-\sigma_j)(1-\rho)\right)}_{\substack{\text{probability that } a_i \text{ enters } C \text{ in the} \\ \text{second stage and the alternatives better} \\ \text{than } a_i \text{ do not enter } C}}$$

that is

$$p_{OR}(a_i) = (\sigma_i + (1-\sigma_i)\rho)(1-\rho)^{i-1} \prod_{j<i} (1-\sigma_j)$$

In the AND model, the probability $p_{AND}(a_i)$ that $a_i \in A$ is chosen is

$$p_{AND}(a_i) = \underbrace{\sigma_i \prod_{j<i} (1-\rho\sigma_j)}_{\substack{\text{probability that } a_i \text{ enters } C(A) \text{ in the} \\ \text{first stage and the alternatives better} \\ \text{than } a_i \text{ do not enter } C(A)}} \times \underbrace{\rho \prod_{j<i} (1-\rho\sigma_j)}_{\substack{\text{probability that } a_i \text{ remains in } C(A) \text{ in} \\ \text{the second stage and the alternatives} \\ \text{better than } a_i \text{ do not enter } C(A)}}$$

[19] Use the convention that $\prod_{j<i}(.) = 1$ for $i = 1$.

that is,

$$p_{AND}(a_i) = \rho\sigma_i \prod_{j<i} (1 - \rho\sigma_j)$$

Observe that for interior values of the parameters ρ and σ_i, $i = 1, 2...$, the probabilities $p_{OR}(a_i)$ and $p_{AND}(a_i)$ are strictly positive for any i. Consider now the case of a common level of salience $\sigma \in (0, 1)$, with $\sigma = \sigma_i$ for all i. Then the probability distributions are log-linear in quality

$$\begin{aligned} \log p_{OR}(a_i) &= \alpha - \beta (i - 1) \\ \log p_{AND}(a_i) &= \alpha' - \beta' (i - 1), \end{aligned}$$

with $\alpha = \log(\sigma + \rho - \rho\sigma)$, $\beta = -\log(1 - \rho)(1 - \sigma)$, $\alpha' = \log \rho\sigma$, and $\beta' = -\log(1 - \rho\sigma)$. It follows that, defining an (ordinal) utility function u representing $\succ$ on A by $u(a_i) = 1 - i$, we can write the probabilities of choice conditional on the agent picking an element in A as in the statement of Proposition 1.

Proposition 1 clarifies the sense in which the model offers a psychological foundation for the multinomial logit specification. The logit structure simply emerges from the elementary random mechanism postulated for membership of the consideration set $C(A)$ - and from the assumption of maximisation over $C(A)$ - and not from a particular specifcation of additive utility errors. The OR and AND models can be seen in this perspective as a class of 'distortions' of the multinomial logit model. The distortions arise from differences in salience between the alternatives. For example in the OR model

$$\log p_{OR}(a_i) = \log(\sigma_i + (1 - \sigma_i)\rho) + (i - 1)\log(1 - \rho) + \sum_{j<i} \log(1 - \sigma_j)$$

and i may enter non-linearly in the expression through the term $\sum_{j<i} \log(1 - \sigma_j)$.

It is important, however, to bear in mind that the logit structure only holds for (affine transformations of) the particular utility specification assumed. As our models use *ordinal* preference information as primitive, the probability of choice can only be invariant with respect to that type of information. Some allowed utility transformations will destroy the loglinear relationships. In other words, at a fundamental level the log-linearity holds with respect to the quality ranking index. This special case of our model thus is an ordinal version of the multinomial logit.

2.4 Basic Comparative Statics Properties

Some comparative statics properties are immediate and, in both the OR and AND model, as expected:

- ***(salience responsiveness)*** the probability of an alternative being chosen increases in the alternative's own salience and decreases in the salience of the other alternatives;

- ***(quality responsiveness)*** an increase in own quality[20] increases the probability of the alternative being chosen;

- ***(monotonicity)*** if the salience ranking is (weakly) the same as the inverse quality ranking (i.e. $i < j \Rightarrow \sigma_i \geq \sigma_j$), the probability that a better alternative is chosen is higher than the probability that a worse alternative is chosen.[21] However, the distribution of salience may scramble this association between quality and probability of being chosen.

On the other hand, because $p_{OR}(a_i)$ and $p_{AND}(a_i)$ are i^{th} degree polynomials in ρ, the effect of an increase in rationality is more subtle, as we shall study in detail below. In this respect the status of the best alternative a_1 is different from that of all the other alternatives. For a_1 an increase in rationality is always clear-cut good news in both models, with

$$\begin{aligned}
\frac{\partial p_{AND}(a_1)}{\partial \rho} &= (1 - \sigma_1) > 0 \\
\frac{\partial p_{OR}(a_1)}{\partial \rho} &= \sigma_1 > 0
\end{aligned}$$

(observe however that the two models have opposite implications concerning the effect of salience on the impact of rationality).

A further observation stems from looking at the *distribution functions*, over quality levels, indicating the probability of choosing an alternative of at least a given level i of

[20] More precisely, a permutation of the objects in the preference order which improves the ranking of the object.

[21] So in particular this holds for the case of equal salience $\sigma_i = \sigma$ for some common $\sigma \in [0, 1]$.

quality:

$$\begin{aligned} F_{OR}(i) &= 1-(1-\rho)^i \prod_{j\leq i}(1-\sigma_j) \\ F_{AND}(i) &= 1-\prod_{j\leq i}(1-\rho\sigma_j) \end{aligned}$$

from which it is evident that, in both models:

- ***(cumulative rationality responsiveness)*** For any quality level i, the probability of choosing an alternative of quality i or better is increasing in the degree of rationality.

But for individual alternatives different from the best, the probability of being chosen as a function of rationality depends in a non-obvious way on the parameters of the model.

3 Which alternatives gain[22] from rationality?

Can paying more attention ever lead to worse decisions? In the abstract, one might conjecture that the probability of a lower quality alternative being chosen - i.e. the probability that an error is committed - decreases as the agent becomes progressively more rational: for example, spending more cognitive resources for the analysis of a problem, thus increasing ρ, should always lead to an evaluation that cannot be worse than if less effort had been expended in assessing it. This is in fact *not* the case in several disparate and important domains. For instance, in the medical sciences it has been found that the number of correct diagnosis *decrease* as more time is granted to considering the case. In de Vries et al. [46] 80 clinical psychology students were given medical cases describing patients affected by various psychiatric disorders that they had to diagnose. After reading the medical notes, half of the participants were given 4 more minutes to study the cases, while the other half were given another unrelated task (word finding puzzle) that prevented them from reasoning about the cases. The participants in this second group outperformed

[22] We use the convention that being chosen is good news for an alternative, like for a mate being selected or a product being chosen by the consumer, but unlike an employee being singled out for downsizing a department.

those in the other group in terms of providing a correct diagnosis. Observe that the diagnose case is particularly tailored to our setting in a number of dimensions: there is a vast number of choice alternatives some of which may be overlooked, and the fit of diagnoses to symptoms is a matter of objective scientific description. Similar effects are found in justice decisions,[23] consumer evaluations[24] and even sports judgements.[25] That is, the probability of choosing the best alternative decreases as rationality increases. As we explain in detail below, our AND and OR models suggest one possible mechanism to account for this feature.

3.1 OR model

The probability of choice $p_{OR}(a_i)$ is a high degree polynomial in the degree of rationality ρ, and thus the behaviour of $p_{OR}(a_i)$ as a function of ρ is potentially very complicated. In this section we demonstrate how this relationship can indeed exhibit non-monotonicities, but that the number of possible patterns is nevertheless limited: there is at most one peak in the probability of choice for the OR model as rationality increases.

To see this, compute

$$\begin{aligned}\frac{\partial p_{OR}(a_i)}{\partial \rho} &= \left[(1-\sigma_i)(1-\rho)^{i-1}\prod_{j<i}(1-\sigma_j)\right] + \\ &\quad \left[(\sigma_i + (1-\sigma_i)\rho)(i-1)(1-\rho)^{i-2}\prod_{j<i}(1-\sigma_j)\right] \\ &= (1-\rho)^{i-2}(1-i\sigma_i - i(1-\sigma_i)\rho)\prod_{j<i}(1-\sigma_j)\end{aligned}$$

[23]Ham van den Bos and van Doorn [17] find that justice decisions concerning job applicants are more accurate if less conscious thought is devoted to the evaluation.

[24]See Dijksterhuis et al. [9] and Dijksterhuis and Nordgren [10], who sparked a huge literature on this subject with their Theory of Unconscious Thought. However other authors, most notably Payne et al. [35], have argued that these effects may be dependent on the specific domain/type of decison problem considered.

[25]This applies to better predictions of the outcome of soccer games made by non experts who were asked to provide a prediction immediately as compared to having more time to mull things over. See Dijksterhuis et al. [8].

which is ambiguous in sign. The decomposition in square brackets highlights the source of ambiguity. On the one hand, an increase in ρ increases the probability that a_i will be considered by the decision maker in the event, with probability $(1-\sigma_i)$, that it has not entered the consideration set because of its salience; but, on the other hand, it also increases the probability that better alternatives are considered.

Defining the threshold value

$$\frac{1-i\sigma_i}{(1-\sigma_i)\,i} \equiv \overline{\rho}$$

we have

$$\frac{\partial p_{OR}(a_i)}{\partial \rho} > 0 \Leftrightarrow \rho < \overline{\rho}$$

The threshold $\overline{\rho}$ ranges in $(-\infty, 1]$ and attains its maximum setting $i = 1$. Therefore $\frac{\partial p_{OR}(a_i)}{\partial \rho}$ is either non-monotonic but single-peaked or monotonic on $(0,1)$, and $p_{OR}(a_i)$ attains a maximum, as a function of $\rho \in (0,1)$, at $\overline{\rho}$ whenever $\overline{\rho} \in (0,1)$. For $p_{OR}(a_i)$ to peak at positive levels of ρ, it must be that $i < \frac{1}{\sigma_i}$ for otherwise $\overline{\rho} \leq 0$. Quality and salience are substitutes to maintain a given $\overline{\rho}$.

The set of alternatives can thus be partitioned into in three types (according to when an increase in rationality is good news for the alternative), which we record as:

Proposition 2 *$p_{OR}(a_i)$ has at most one peak as a function of the degree of rationality $\rho \in (0,1)$. There are three cases:*

(1) For the top quality alternative, $(i = 1)$ the probability of choice $p_{OR}(a_1)$ is always increasing in ρ;

(2) For alternatives a_i with $\frac{1}{\sigma_i} > i > 1$, the probability of choice $p_{OR}(a_i)$ is strictly increasing in ρ on an initial range;

(3) For alternatives a_i with $\sigma_i i \geq 1$, the probability of choice $p_{OR}(a_i)$ is strictly decreasing in ρ.

To summarise in words, for certain parameter values, the degree of rationality which maximises the probability that a given alternative (different from the top one but of sufficient good quality) is chosen, is an *intermediate* one. An increase in rationality is good news only for:

- the top alternative, always;

- alternatives displaying a combination of good quality and low salience, at sufficiently low rationality levels.

Note that only own salience, and not the salience of the other alternatives, affects the value of $\overline{\rho}$ and thus the sign of the derivative. The effect of own salience on $\overline{\rho}$ (namely $i-i^2$) is negative for $i > 1$ and zero for $i = 1$. A lower quality (increase in i) reduces the threshold $\overline{\rho}$. The quality effect and the salience effect, as well as examples of choice probabilities peaking at intermediate degrees of rationality of second (or worse) rate alternatives, are visualised in figure 2. Here, with respect to a baseline case (black line, $i = 2$, $\sigma_i = 0.1$) quality is decreased (to $i = 6$) in the $p_{OR}(a_i)$ represented by the gray solid line while salience is increased (to $\sigma_i = 0.3$) in the $p_{OR}(a_i)$ represented by the grey dashed line.

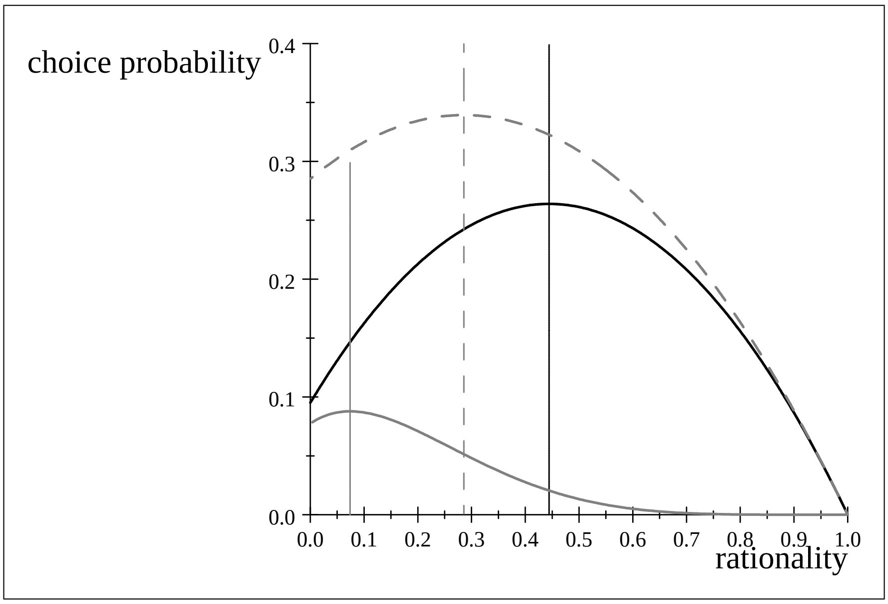

Figure 2: Comparative statics in the OR model: increasing σ_i shifts $p_{OR}(a_i)$ upwards and to the left; increasing i shifts $p_{OR}(a_i)$ downwards and to the left.

3.2 AND model

The response of the probability of choice in the AND model as rationality increases is qualitatively different from that of the OR model, as we now demonstrate, though in this case too the choice probability can fall as rationality increases over certain ranges.

In order to highlight the role of ρ it is convenient to rewrite the model with the following notation. Let $S(m,k)$ denote the ordered set of combinations of k elements from the set $\{\sigma_1,...,\sigma_m\}$, where $|S(m,k)| = \binom{m}{k}$, with all of the elements in $S(m,k)$ listed in ascending order lexicographically. Finally, let $s_{m,k} = \left\{1,2,...,\binom{m}{k}\right\}$ denote the corresponding index set and let $S(m,k)(i)$ denote the $i-th$ element of $S(m,k)$.

To see why this notation is useful, let for example $i=4$ and compute the probability $p_{AND}(a_4)$ that alternative a_2 is selected. This is given by

$$\begin{gathered} p_{AND}(a_4) = \sigma_4\rho\left((1-\sigma_1\rho)(1-\sigma_2\rho)(1-\sigma_3\rho)\right) \\ = \sigma_4\rho\left(1-\sigma_1\rho-\sigma_2\rho-\sigma_3\rho+\sigma_2\sigma_3\rho^2+\sigma_1\sigma_2\rho^2+\sigma_1\sigma_3\rho^2-\sigma_1\sigma_2\sigma_3\rho^3\right) \\ = \sigma_4\rho\left(1-\rho(\sigma_1+\sigma_2+\sigma_3)+\rho^2(\sigma_1\sigma_2+\sigma_1\sigma_3+\sigma_2\sigma_3)-\rho^3\sigma_1\sigma_2\sigma_3\right) \end{gathered}$$

The relevant index sets are $s_{3,1} = \{1,2,3\}$, $s_{3,2} = \{1,2,3\}$ and $s_{3,3} = \{1\}$, so that e.g. $S(3,2)(1) = \sigma_1\sigma_2$, $S(3,2)(2) = \sigma_1\sigma_3$, $S(3,2)(3) = \sigma_1\sigma_3$, and so on. Then we can rewrite $p_{AND}(a_4)$ as

$$p_{AND}(a_4) = \sigma_4\rho\left(1+\sum_{j=1}^{3}\left((-\rho)^j\sum_{k\in s_{3,j}}S(3,j)(k)\right)\right)$$

In general, defining

$$A(i,j) = \sum_{k\in s_{i-1,j}} S(i-1,j)(k),$$

the probability that a_i is chosen can be rewritten as:

$$p_{AND}(a_i) = \sigma_i\rho\left(1+\sum_{j=1}^{i-1}\left((-\rho)^j A(i,j)\right)\right)$$

We can now check how this varies with rationality:

$$\frac{\partial p_{AND}(a_i)}{\partial\rho} = \sigma_i\left(1+\sum_{j=1}^{i-1}\left((-\rho)^j A(i,j)\right)\right)+\sigma_i\rho\left(\sum_{j=1}^{i-1} j(-1)^j\left(\rho^{j-1}A(i,j)\right)\right)$$

which yields

$$\frac{\partial p_{AND}(a_i)}{\partial\rho} = \sigma_i\left(1+\sum_{j=1}^{i-1}\left((j+1)(-\rho)^j A(i,j)\right)\right)$$

Like in the OR model, the effect of a change in ρ on the probability of choice is ambiguous, but here there clearly exists $\widehat{\rho}\in(0,1)$ such that $\frac{\partial p_{AND}(a_i)}{\partial\rho} > 0$ if $\rho < \widehat{\rho}$. Unlike in the OR model, *there cannot be any sure-fire loser from an increase in rationality*: every alternative gains from increases in rationality, whatever the salience profile and the quality of the alternative, at sufficiently low levels of rationality (this happens by taking away choice probability from the default alternative a^*).

Nevertheless, as in the OR model, the pattern of behaviour cannot be too complicated: the threshold $\widehat{\rho}$, when it exists in $(0,1)$, is unique.

Proposition 3 *For all i, the probability of choice $p_{AND}(a_i)$ has at most one peak as a function of the degree of rationality $\rho \in (0,1)$, and it is strictly increasing on an initial range. Moreover $p_{AND}(a_i)$ can be strictly increasing in ρ on the entire interval $(0,1)$ even for a lower quality alternative a_i with $i > 1$.*

All (easy but mostly tedious) missing proofs for this and subsequent results are relegated to an Appendix.

The latter part of the statement highlights the major difference from the OR model that increases in rationality can be good news for inferior alternatives at *all* levels of rationality.

Note finally that, unlike in the OR model, the *entire* salience profile is relevant to determine the impact of rationality. We display the salience and quality effect in the graph below, using the same values as for the OR model:

3.3 Choice Odds and Menu Effects

We have noted that the effect of an increase in rationality on the probability of choice of any alternative which is not the best is ambiguous. But what about the *odds* of choosing a better quality alternative over a lower quality alternative? Even if the probability of choosing an inferior alternative increases with rationality, one may conjecture that it does so at a lower speed than superior alternatives, so that the odds of making a better choice increase. This conjecture is clearly true in the AND model. Defining, for $i < j$,

$$odds_{AND}(i,j) \equiv \frac{p_{AND}(a_i)}{p_{AND}(a_j)} = \frac{\sigma_i}{\sigma_j \prod_{k=i}^{j-1}(1-\rho\sigma_k)}$$

we have immediately

$$\frac{\partial odds_{AND}(i,j)}{\partial \rho} > 0$$

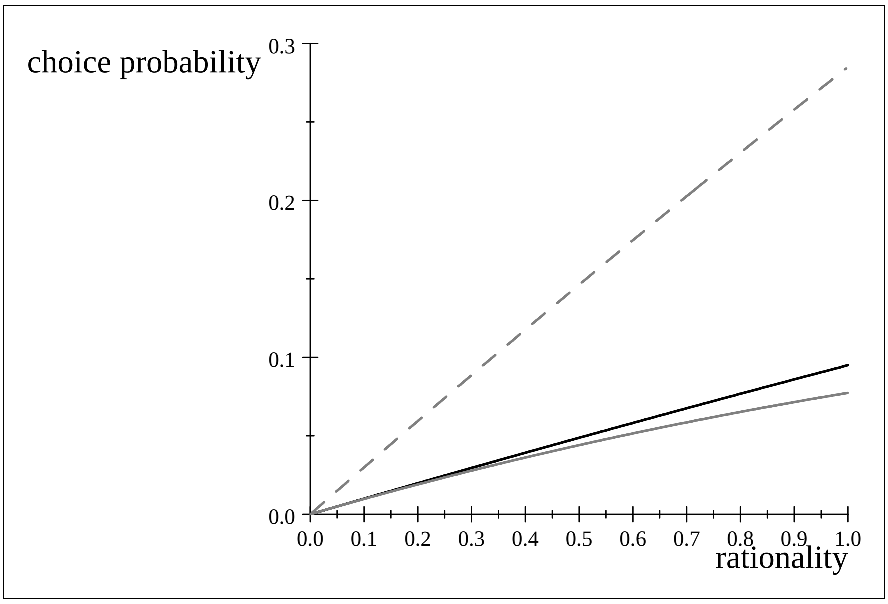

Figure 3: Comparative statics in the AND model: increasing σ_i shifts $p_{OR}(a_i)$ upwards; increasing i shifts $p_{OR}(a_i)$ downwards. With these parameter values ρ is always increasing over the $(0, 1)$ interval.

But the conjecture turns out to be false in the OR model, at low levels of rationality and low levels of salience of the inferior alternative, irrespective of the quality difference between the two alternatives.

Define, with $i < j$,

$$odds_{OR}(i,j) \equiv \frac{p_{OR}(a_i)}{p_{OR}(a_j)} = \frac{(\sigma_i + (1-\sigma_i)\rho)}{(\sigma_j + (1-\sigma_j)\rho)(1-\rho)^{j-i}\prod_{k=i}^{j-1}(1-\sigma_k)}$$

Then we have

Proposition 4 *The odds of picking a worse alternative may increase with the degree of rationality: for all i, j with $i < j$, there exists $\bar{\rho} \in (0,1)$ and $\bar{\sigma}_j \in (0,1)$ such that, for $\rho < \bar{\rho}$ and $\sigma_j < \bar{\sigma}_j$, $\frac{\partial odds_{OR}(i,j)}{\partial \rho} < 0$.*

Observe that both expressions for the odds violate Luce's [23] classical IIA axiom, which states that the choice probability ratio for two alternatives a_i and a_j is independent of the other alternatives in the choice set A. In our models,[26] this holds true only for changes in A which remove or delete alternatives each of which is either better or worse

[26]Imagining now that they apply to a collection of subsets of a universal set of objects X.

than *both* a_i and a_j. Inserting, for example, an alternative a_l with $a_i \succ a_l \succ a_j$ in the choice set would change the terms $(1-\rho)^{j-i}\prod_{k=i}^{j-1}(1-\sigma_k)$ and $\prod_{k=i}^{j-1}(1-\rho\sigma_k)$ which appear in $odds_{OR}(i,j)$ and $odds_{AND}(i,j)$, respectively. The insertion of such an intermediate alternative would make no difference regarding the probability of choice of the better alternative a_i, but would create a new event of probability $(\sigma_l+(1-\sigma_l)\rho)(\sigma_j+(1-\sigma_j)\rho)$ in the OR model and $(\sigma_l\rho)(\sigma_j\rho)$ in the AND model (namely the probabilities that a_n and a_j are both considered), in which the lower quality alternative is not chosen. As a particular implication of these observations, this means that $odds_{OR}(i,j)$ and $odds_{AND}(i,j)$ are weakly increasing with the size of the choice set.

The dependence of the choice odds on the other available alternatives is often a realistic feature, which applied economist have sought to incorporate, for example, in the multinomial logit model.[27] The blue bus/red bus problem is the standard example. Suppose the agent chooses with probabilities one third each the train (t), a red bus (r) or a blue bus (b) as a means of transport, so that the choice odds for any two alternatives are 1. Nevertheless, if r is removed from the choice set, it is natural to expect that the odds of choosing b over t become 2, rather than staying at 1 as required by IIA. In our model (once adapted to include ranking ties), the natural explanation for why the odds should change (that the agent ranks a blue bus and a red bus in the same way) immediately yields the odds change.[28]

4 Stochastic Intransitivity

When choice is random there are many ways to define analogues of transitive behaviour in deterministic models. A popular such analogue is *stochastic transitivity.* In this section we assume that the AND and the OR model are defined for a domain of binary sets $\{a_i, a_j\}$. A random choice rule is a function p that maps each pair $\{a_i, a_j\}$ in the domain

[27]By adding a nested structure to choice process (nested logit) or by allowing heteroscedasticity of the choice errors. See e.g. Greene [15].

[28]Gul, Natenzon and Pesendorfer [16] relax the the Luce model so as to accommodate the blue bus/red bus example (see section 6).

to a probability distribution over $\{a_i, a_j, a^*\}$. The rule p satisfies stochastic transitivity if

$$p(a_i, \{a_i, a_j\}) \geq \frac{1}{2}, p(a_j, \{a_j, a_k\}) \geq \frac{1}{2} \Rightarrow p(a_i, \{a_i, a_k\}) \geq \frac{1}{2}$$

Applying our models p_{OR} and p_{AND} to this framework, it is easy to see that they can account for violations of stochastic transitivity. Interestingly, in the AND model this can happen even in the limit for the agent tending to full rationality (i.e. $\rho = 1$).

To see this, for the AND model we consider the following example. Set $\sigma_1 = 0.4$, $\sigma_2 = 0.51$ and $\sigma_3 = 0.9$. We have:

$$\begin{aligned}
\lim_{\rho \to 1} p_{AND}(a_2, \{a_2, a_3\}) &= \lim_{\rho \to 1} \rho\sigma_2 = 0.51 > \frac{1}{2} \\
\lim_{\rho \to 1} p_{AND}(a_3 | \{a_1, a_3\}) &= \lim_{\rho \to 1} \rho\sigma_3 (1 - \rho\sigma_1) \\
&= 0.9(1 - 0.4) = 0.54 > \frac{1}{2}
\end{aligned}$$

but also

$$\begin{aligned}
p_{AND}(a_2, \{a_1, a_2\}) &= \lim_{\rho \to 1} \rho\sigma_2 (1 - \rho\sigma_1) \\
&= 0.51(1 - 0.4) = 0.306 < \frac{1}{2}
\end{aligned}$$

thereby violating stocahstic transitivity.

The assumption of full rationality (i.e. $\rho \to 1$) is of course not crucial in this example, since for instance we could replicate a similar intransitivity with a low value of rationality, say $\rho = 0.6$, by rescaling upwards the values for σ_i accordingly. It is interesting to note that in the AND model, stochastic transitivity will always hold provided ρ is sufficiently low, since e.g. whenever $\rho < 0.5$ no alternative can be chosen with probability higher than $\frac{1}{2}$ in binary sets , so that stochastic transitivity holds vacuously.

For the OR model, set for instance $\rho = 0.4$, $\sigma_1 = 0.1$, $\sigma_2 = 0.51$ and $\sigma_3 = 0.9$ to obtain

$$\begin{aligned}
p_{OR}(a_2, \{a_2, a_3\}) &= (\sigma_2 + (1 - \sigma_2)\rho) \\
&= (0.51 + (1 - 0.51)\,0.4) = 0.706 > \frac{1}{2} \\
p_{OR}(a_3, \{a_1, a_3\}) &= (\sigma_3 + (1 - \sigma_3)\rho)(1 - \rho)(1 - \sigma_1) \\
&= (0.9 + (1 - 0.9)\,0.4)(1 - 0.4)(1 - 0.1) = 0.5076 > \frac{1}{2}
\end{aligned}$$

but

$$
\begin{aligned}
p_{OR}(a_2, \{a_1, a_2\}) &= (\sigma_2 + (1 - \sigma_2)\rho)(1 - \rho)(1 - \sigma_1) \\
&= (0.51 + (1 - 0.51)\, 0.4)(1 - 0.4)(1 - 0.1) = 0.381\,24 < \frac{1}{2}
\end{aligned}
$$

The key to obtaining stochastic cycles in both sets of examples is that the salience ordering is exactly opposite to the quality odering of the alternatives.

5 Salience games: does salience reveal quality?

5.1 Absolute salience: The showiest is the best

We now consider the situation where alternatives can choose, possibly at a cost, the salience they possess. This is natural in several contexts. For example, a minor politician can make an outrageous statement to get noticed by the media and enter the voters' consideration set, but he will likely incur a cost in terms of credibility. One can increase expenditure on hairdressing to get noticed by potential partners. And firms, of course, have huge advertising budgets.

We are mainly interested in the question of how the equilibrium salience order 'reveals' the quality order (and in how this is reflected in choice probabilities). The answer is not obvious a priori as incentives seem to run both ways. On the one hand the best alternative has a strong incentive to get noticed: it fears no competition. On the other hand, the only weapon the inferior alternatives have to have a chance to be chosen is to increase their probability of entering the consideration set.

In this section we assume that there is a finite number of alternatives, and that the strategy set for alternative a_i is a *finite* subset S of the unit interval (below we illustrate how other domains could be considered). The payoff to each alternative is the expected probability of being chosen minus a (possibly negative) cost associated with the chosen salience level. One interpretation of this function is that alternatives either vie for one single chooser who chooses one alternative, or care about 'market share' with a continuum of identical choosers each of whom chooses one alternative. Formally, let e be a function

$e : S \to \mathrm{Re}$. The payoff to alternative i for a pure strategy profile $\sigma \in S^n$ is

$$z_i(\sigma_i, \sigma_{-i}) = (\sigma_i + (1-\sigma_i)\rho)(1-\rho)^{i-1} \prod_{j<i} (1-\sigma_j) - e(\sigma_i)$$

for the OR model and

$$z_i(\sigma_i, \sigma_{-i}) = \sigma_i \rho \prod_{j<i} (1-\rho\sigma_j) - e(\sigma_i)$$

for the AND model. We make no assumption on the function e. In particular, e may not be monotonic increasing. So e could be interpreted as painful effort, when increasing salience is costly, or as elation, when increasing salience is pleasurable at least on some range.

Proposition 5 *In both the AND and the OR model there exists an equilibrium in pure strategies.*

The proof makes it clear that this pure strategy existence result is robust to generalisations. It continues to hold when S is a compact subset of $[0,1]$ and $e(.)$ is continuous, or possibly discontinuous but increasing. The next characterisation result holds even more generally, for any structure of $S \subseteq [0,1]$ and any $e(.)$.

When salience can be chosen endogenously, in equilibrium at any level of rationality the salience order coincides (weakly) with the preference order.

Proposition 6 (The showiest is the best) *Suppose $a_i \succ a_j$ and let $(\sigma_1, ..., \sigma_n)$ be a pure strategy equilibrium. Then, for all ρ, $\sigma_i \geq \sigma_j$ both in the AND and in the OR model.*

Note that lower quality alternatives do not have any intrinsic disadvantage, in terms of salience enhancing technology, with respect to higher quality alternatives, so our result is not supported by any kind of signalling argument. The reason why lower quality alternatives produce less salience in equilibrium does not derive from lower levels of resources or lower unit costs of salience production: every alternative can choose from exactly the same set at exactly the same cost or benefit. The result is purely a function of the cognitive process postulated for the agent.

5.2 Relative salience: the ugly duckling can get picked most often

In the previous section each alternative could select its own salience independently of the salience of the other alternative. In this sense salience was *absolute*. This is appropriate in some contexts, e.g. if repeated ads in favour of an alternative merely have the function of making the agent aware of the alternative ('did you know that people who read book A also read book B?'; 'have you considered using a scooter to go to work'?), with σ_i representing either the probability that the agent is aware or the proportion of aware agents within a population. In other contexts, however, alternatives can only control variables that affect salience in a *relative* way. If everybody else dresses in green you will be salient by dressing in yellow, and viceversa. If all other candidates converge on a given political message, you will be salient by deviating from that message. We call this the case of relative salience.

We show that in this case the neat equilibrium ordering obtained in proposition 6 breaks down. As a consequence, it is even possible that, in equilibrium, the worst alternative is selected with the highest probability.

Suppose now that each alternative a_i selects a 'position' $v_i \in [0, 1]$, and that own salience is determined by the entire profile of the v_i's. In particular, we assume that an alternative's salience is conferred by its difference, in terms of position, from the 'average alternative' (excluding itself)

$$\sigma_i = \left(v_i - \frac{\sum_{j \neq i} v_j}{(n-1)} \right)^2 \in [0, 1]$$

The alternatives aim as usual at maximising the probability of being chosen, where the probability is computed according to either the AND or the OR model.

Proposition 7 *In the AND model there exist (for some n) pure strategy Nash equilibria in which, for any ρ, the worst alternative has the highest probability of being chosen.*

Proposition 8 *In the OR model there exist (for some n) pure strategy Nash equilibria in which, for ρ sufficiently small, the worst alternative has the highest probability of being chosen.*

The difference in the statements of these two results stems from a difference in the limiting behaviours of the AND the OR model: at degrees of rationality near one, the OR model - but not the AND model - approximates well the standard utility maximisation model.[29]

5.3 What about prices?

In our salience games we have completely eschewed the price setting issue: after all, it may be reasonable to expect different producers of a product of a given quality to use both prices and salience/advertising as a strategic variable. Ignoring prices however is not a major drawback of our analysis given our aims, as we are considering a situation where the ranking of the alternative is given - so we could see the ranking of alternatives as already embedding a quality/price combination. If our agents could consider all of the available alternatives, they would have no difficulty in picking the best one according to their preferences, which would incorporate any trade-offs between intrinsic product quality and price. We need this as a premise since the question we have addressed is: if the decision maker is not perfectly aware of the relevant choice set because of perceptual or other limitations, how do rationality and salience interact in determining choice probabilities?

Our results show that 'choice errors' can lead decision makers to demand products, which would otherwise not be demanded, even if these products are low ranked because of e.g. high price or low quality or a combination of the two. This insight is in line with field evidence (see e.g. Goeree [13]).[30]

6 Related literature

As we explained in the introduction the logit model can be seen as a very specific version of an additive RUM model (models in which there is a probability distribution over the

[29] This statement follows from a straighforward calculation of the limits.

[30] In the market for personal computers, Goree [13] finds that estimated markups over costs are higher than what a model in which agents consider all of the avaiable alternatives would predict, with top firms advertising more than average and earning more than average markups. She also posits the lack of full information on what the full range of available alternatives is as an explanation for her results.

utility functions that the agent maximises). The Luce [23] model, according to which

$$p(a_i, A) = \frac{u(a_i)}{\sum_{a_i \in A} u(a_i)}$$

for some systematic utility u, is also a RUM model (Block and Marschak [7]). Recently, Gul, Natenzon and Pesendorfer [16] have significantly advanced our understanding of the Luce model. First, in a domain which is 'rich' in a certain sense, they show that the Luce model is equivalent to the following Independence property: $p(a_i, A \cap C) \geq p(a_j, B \cap C)$ implies $p(a_i, A \cap D) \geq p(a_j, B \cap D)$ for all sets A, B, C and D such that $(A \cup B) \cap (C \cup D) = \varnothing$. Second, they generalise the Luce model in such a way as to accommodate red bus/blue bus type of violations of Luce's Independence (see section 3.3), as well as some well-known choice anomalies. We view such type of axiomatic analyses as complementary to the procedural approach adopted in this paper. Observe also how that type of analysis imposes restrictions across choice sets. Here, we have focussed entirely on a given choice set. Stochastic choice is also the focus of the anticipated choice model characterised in Koida [19]. In this model, though, the emphasis is on how a decision maker's (probabilistic) mental states drive the choice of an alternative from each menu, in turn determining in non-obvious ways the agent's preferences over menus. In our setup we instead concentrate on mistakes before choice is made.

Turning to the procedural approach, Mattsson and Weibull [33] obtain an elegant foundation for (and generalisation of) the logit. In their model the agent (optimally) pays a cost to get close to implementing any desired outcome (see also Voorneveld [45]). More precisely, the agent has to exert more effort the more distant the desidered probability distribution from a given default distribution. When the agent makes an optimal trade-off between the expected payoff and the cost of decision control, the resuting choice probabilities are a 'distortion' of the logit model, in which the degree of distortion is governed by the default distribution. In one way our paper shares with this work the broad methodology mentioned above, to focus on a detailed model to explain choice errors. However, it is also very different in that the latter assumes a (sophisticated form of) rational behaviour on the part of the agent. One may then wonder whether 'utility-maximisation errors' might not occur at the stage of making optimal tradeoffs between utility and control costs, raising the need to model those errors. A second major difference stems from the fact that, as noted, our model uses purely *ordinal* preference information.

Recently, Rubinstein and Salant [40] have proposed a general framework to describe an agent who expresses different preferences under different *frames* of choice. The link with this paper is that the set of such preferences is interpreted as a set of deviations from a true (welfare relevant) preference. However, their analysis takes a very different direction from ours in that it eschews any stochastic element. The probability model is, on the contrary, at the core of our theory.

There are also different plausible ways to model consideration sets and the competition for them. The already mentioned work by Eliaz and Spiegler [11] studies in great detail the competition between two firms, who choose marketing strategies to make their products enter the consideration sets of a continuum of identical consumers. The choice model at the heart of this work is analogous to the more abstract one of Masatlioglu and Nakajima [25], which is deterministic. Eliaz and Spiegler [11] also perform comparative statics exercises that relate to changes in rationality. One the main findings is that in the equilibria corresponding to some parameter configurations firms do not increase their profits compared to a situation in which consumers are fully rational (informed). More comparisons are made either by introducing in the population of boundedly rational consumers some rational consumers, or by changing the 'consideration function' (the function that determines the consideration set of consumers). One implication is that industry profits are a non-monotonic function of changes in rationality thus defined.[31]

7 Concluding remarks

The broad aim of this paper was to open the black box of 'bounded rationality' and statistical distributions in explaining choice errors that make an agent deviate from 'true' utility maximisation. The methodology has been that of studying a concrete model of bounded rationality, focussing on a specific decision *procedure*, instead of restricting a priori the error structure, either through axioms or through the selection of a particular probability distribution. This has led to some unexpected properties of choice probabilities. We think

[31] See also Eliaz and Spiegler [12] for a further analysis of related topics.

the general approach we have taken is validated by its ability to provide a plausible explanation for otherwise puzzling choice observations in a number of varied fields from clinical medicine to law, consumer behaviour and sports. A second novelty of the model, which we believe to be of wider interest, is to suggest an equilibrium mechanism to endogenise its parameters.

A general message from our paper is that 'revealed preferences' do not necessarily become a better guide to discovering true preferences as the rationality of the agent increases (namely when the agent has a higher probability of being better informed about the available alternatives). Suppose you can observe or infer the degrees of rationality, ρ and ρ', under which two sets of choices were made. Suppose that alternative x is chosen more frequently over alternative y in condition ρ than in condition ρ': it does not necessarily follow from the fact that $\rho > \rho'$ that it is more likely that the chooser deems x better than y. Less rational agents (or agents choosing under worse informational conditions) may express their preference through choice more truthfully than more rational agents.[32] Our model thus underscores the potentialsubtlety of the link between preference and choice: this is a topic to be addressed in further research.

References

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

[2] Ambler, T., S. Braeutigam, J. Stins, S. Rose and S. Swithenby (2004) "Salience and Choice: Neural Correlates of Shopping Decisions", *Psychology and Marketing*, 21(4): 247–261.

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

[4] Babin, B. J., W. R. Darden and M. Griffin (1994) "Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value", *The Journal of Consumer Research*, 20(4): 644-656.

[32] Note that, because the frequencies of choice are affected by salience, they cannot in general be taken as a guide to the quality ranking, except in special cases such as equal salience across alternatives or endogenous absolute salience.

[5] Bernheim, B. D., 2009, "Behavioral Welfare Economics", *Journal of the European Economic Association*, 7: 267–319.

[6] Bernheim, B. D. and A. Rangel, 2009, "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics", *Quarterly Journal of Economics*, 124: 51-104.

[7] Block, H.D. and J. Marschak (1960) "Random Orderings and Stochastic Theories of Responses", in Olkin, Ingram, Sudish G. Gurye, Wassily Hoeffding, Willima G. Madow, and Henry B. Mann (eds.) Contributions to Probability and Statistics, Stanford University Press, Stanford, CA.

[8] Dijksterhuis, A., M. W. Bos, A. van der Leij and R. B. van Baaren (2009) "Predicting Soccer Matches After Unconscious and Conscious Thought as a Function of Expertise", *Psychological Science*, 20: 1381-1387.

[9] Dijksterhuis, A., M. W. Bos, L. F. Nordgren and R. B. van Baaren (2006) "On Making the Right Choice: The Deliberation-Without-Attention Effect", *Science*, 311(5763): 1005-1007.

[10] Dijksterhuis, A. and L. F. Nordgren (2006) "A Theory of Unconscious Thought", *Perspectives on Psychological Science*, 1(2): 95-109.

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

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

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

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

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

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

[17] Ham, J.,K. van den Bos and E. A. van Doorn (2009) "Lady Justice Thinks Unconsciously: Unconscious Thought Can Lead to More Accurate Justice Judgments", *Social Cognition*, 27(4): 509-521.

[18] Hauser, J., and Wernerfelt, B. (1990) "An evaluation cost model of consideration sets" *Journal of Consumer Research*, 16(4), 393–408.

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

[20] Kirschnera, P.A., P. Ayres and P. Chandler (2011) "Contemporary cognitive load theory research: The good, the bad and the ugly", *Computers in Human Behavior*, 27(1): Pages 99-105.

[21] Lavie, N. (2005) "Distracted and confused?: Selective attention under load", *Trends in Cognitive Sciences*, 9(2): 75-82.

[22] Lavie, N., A. Hirst, J.W. de Fockert and E. Viding (2004) "Load theory of selective attention and cognitive control", *Journal of Experimental Psychology* 133 (3): 339–54.

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

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

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

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

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

[28] McFadden, D.L. (1974) "The Measurement of Urban Travel demand", *Journal of Public Economics* 3: 303-28.

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

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

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

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

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

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

[35] Payne J. W., A. Samper, J. R. Bettman, M. F. Luce (2008) "Boundary Conditions on Unconscious Thought in Complex Decision Making", *Psychological Science*, 19(11): 1118-23.

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

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

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

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

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

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

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

[43] Sugden, R. (2004) "The Opportunity Criterion: Consumer Sovereignity Without the Assumption of Coherent Preferences", *American Economic Review* 94: 1014-1033.

[44] Tyson, C. J. (2010) "Salience and satisficing", mimeo, Queen Mary U. of London.

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

[46] Marieke de Vries, M., C. L. M. Witteman, R. W. Holland and A. Dijksterhuis (2010) "The Unconscious Thought Effect in Clinical Decision Making: An Example in Diagnosis", *Medical Decision Making*, 30: 578–581.

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

8 Appendix: Proofs

Proof of Proposition1. We note a more general result which holds for a class of models including OR and AND as special cases. Let the probability that an alternative belongs to $C(A)$ be given by a function $f:(0,1)\times(0,1)\to(0,1)$, with $f(\sigma_i,\rho)=prob(a_i\in C(A))$. Suppose $\sigma=\sigma_i$ for all i for some $\sigma\in(0,1)$. Then there exists a utility function $u:A\to\mathbb{R}$ representing $\succ$ on A and coefficients γ,δ such that

$$p(a_i,A)=\exp(\gamma+\delta u(a_i))/\sum_j\exp(\gamma+\delta u(a_j))$$

This is true since:

$$\begin{aligned}p(a_i) &= prob(a_i\in C(A))\prod_{j<i}(1-prob(a_j\in C(A)))\\ &= f(\sigma_i,\rho)\prod_{j<i}(1-f(\sigma_j,\rho))\end{aligned}$$

Then the argument in the text continues to hold. Setting $\sigma_i=\sigma$ for all i, for some $\sigma\in(0,1)$, we have $p(a_i)=f(\sigma,\rho)(1-f(\sigma,\rho))^{i-1}$. Since $f(\sigma_i,\rho)\in(0,1)$ we can write $\log p(a_i)=\gamma-\delta(i-1)$, where $\gamma=\log f(\sigma_i,\rho)$ and $\delta=-\log(1-f(\sigma_j,\rho))$. Defining $u(a_i)=1-i$, we have $p(a_i)=\exp^{\gamma+\delta u(a_i)}$. ■

Proof of Proposition2: In the text.

Proof of Proposition 3: Define as in the text

$$A(i,j)=\sum_{k\in s_{i-1,j}}S(i-1,j)(k)$$

We have already observed that $p_{AND}(a_i)$ is strictly increasing on the initial range of definition. We study the sign of $\frac{\partial p_{AND}(a_i)}{\partial\rho}$, which depends on the sign of the expression

$$\begin{aligned}&1+\sum_{j=1}^{i-1}\left((j+1)(-\rho)^j A(i,j)\right) &&(*)\\ =\ &\rho^2\left(\frac{1}{\rho^2}+\sum_{j=1}^{i-1}\left((j+1)(-\rho)^{j-2}A(i,j)\right)\right)\end{aligned}$$

We show that there exists a single value $\widehat{\rho} \in (0,1)$ at which $\frac{\partial p_{AND}(a_4)}{\partial \rho}$ vanishes. Suppose to the contrary that there were two such values $\widehat{\rho} \in (0,1)$ and $\widehat{\widehat{\rho}} \in (0,1)$, say with $\widehat{\rho} < \widehat{\widehat{\rho}}$. Then, using the LHS side expression in equation * and the definition of $\widehat{\rho}$ and $\widehat{\widehat{\rho}}$,

$$\sum_{j=1}^{i-1} \left((j+1)(-\widehat{\rho})^j A(i,j) \right) = -1 = \sum_{j=1}^{i-1} \left((j+1)\left(-\widehat{\widehat{\rho}}\right)^j A(i,j) \right)$$

On the other hand, using the RHS in equation * and the definition of $\widehat{\rho}$ and $\widehat{\widehat{\rho}}$,

$$\sum_{j=1}^{i-1} \left((j+1)(-\widehat{\rho})^{j-2} A(i,j) \right) = -\frac{1}{(\widehat{\rho})^2}$$

and

$$\sum_{j=1}^{i-1} \left((j+1)\left(-\widehat{\widehat{\rho}}\right)^{j-2} A(i,j) \right) = -\frac{1}{\left(\widehat{\widehat{\rho}}\right)^2}$$

Therefore

$$\sum_{j=1}^{i-1} \left((j+1)(-\widehat{\rho})^{j-2} A(i,j) \right) < \sum_{j=1}^{i-1} \left((j+1)\left(-\widehat{\widehat{\rho}}\right)^{j-2} A(i,j) \right)$$

so that

$$(\widehat{\rho})^2 \sum_{j=1}^{i-1} \left((j+1)(-\widehat{\rho})^{j-2} A(i,j) \right) < \left(\widehat{\widehat{\rho}}\right)^2 \sum_{j=1}^{i-1} \left((j+1)\left(-\widehat{\widehat{\rho}}\right)^{j-2} A(i,j) \right)$$

$$\Leftrightarrow \sum_{j=1}^{i-1} \left((j+1)(-\widehat{\rho})^{j} A(i,j) \right) < \sum_{j=1}^{i-1} \left((j+1)\left(-\widehat{\widehat{\rho}}\right)^{j} A(i,j) \right)$$

a contradiction. Therefore there is at most one value of $\rho \in (0,1)$ at which $\frac{\partial p_{AND}(a_i)}{\partial \rho}$ vanishes, from which (since $\frac{\partial p_{AND}(a_i)}{\partial \rho}$ is a polynomial and is strictly increasing on the initial range of definition) the first part of the statement follows. The plot in the text shows examples for which $p_{AND}(a_i)$ is strictly increasing on the whole interval $(0,1)$: as is evident from the formula for $p_{AND}(a_i)$, this can be obtained by setting values of $\sigma_1, ..., \sigma_{i-1}$ sufficiently low (note that $A(i,j)$ approximates $\sigma_i \rho$ when $\sigma_1, ..., \sigma_{i-1}$ are close to zero). ■

Proof of Proposition 4: Differentiate logarithmically and rearrange to obtain

$$\frac{\partial \log odds_{OR}(i,j)}{\partial \rho} = \frac{(1-\sigma_i)}{(\sigma_i + (1-\sigma_i)\rho)} + \frac{(j-i)}{(1-\rho)} - \frac{(1-\sigma_j)}{(\sigma_j + (1-\sigma_j)\rho)}$$

which is negative for ρ and σ_j small enough. Conclude by noting that $odds_{OR}(i,j)$ is a positive function so that $sign \frac{\partial \log odds_{OR}(i,j)}{\partial \rho} = sign \frac{\partial odds_{OR}(i,j)}{\partial \rho}$. ■

Proof of Proposition 5: Consider the OR model. At a pure strategy equilibrium, alternative a_1 simply solves the one-person problem

$$\max_{\sigma_1 \in S} (\sigma_i + (1-\sigma_i)\rho) - e(\sigma_i)$$

Let σ_1^* be a solution to this problem. Now suppose inductively that for each $i < j$ the game restricted to alternatives $a_1, ...a_{j-1}$ has a pure strategy equilibrium $\left(\sigma_1^*, ..., \sigma_{j-1}^*\right)$. Then the game between alternatives $a_1, ...a_j$ (of which $1, ..., j-1$ are indifferent to the choice of alternative j) has the pure strategy equilibrium $\left(\sigma_1^*, ..., \sigma_j^*\right)$, where σ_j^* is a solution to the problem

$$\max_{\sigma_j \in S} (\sigma_j + (1-\sigma_j)\rho)(1-\rho)^{j-1} \prod_{i<j} (1-\sigma_i^*) - e(\sigma_j)$$

So for any n the game has a pure strategy equilibrium.

A similar logic applies to the AND model. ■

Proof of Proposition 6: Consider the OR model. By contradiction, suppose that $a_i \succ a_j$ but $\sigma_i < \sigma_j$. We use a revealed preference argument. Because σ_i is optimal for alternative a_i, it must provide a weakly higher expected payoff than σ_j, that is:

$$\begin{aligned} & (\sigma_i + (1-\sigma_i)\rho)(1-\rho)^{i-1} \prod_{k<i} (1-\sigma_k) - e(\sigma_i) \\ \geq \ & (\sigma_j + (1-\sigma_j)\rho)(1-\rho)^{i-1} \prod_{k<i} (1-\sigma_k) - e(\sigma_j) \end{aligned}$$

or

$$\begin{aligned} & ((\sigma_i + (1-\sigma_i)\rho) - (\sigma_j + (1-\sigma_j)\rho))(1-\rho)^{i-1} \prod_{k<i} (1-\sigma_k) \\ \geq \ & e(\sigma_i) - e(\sigma_j) \end{aligned}$$

Since $\sigma_i < \sigma_j$ and $\rho < 1$, we have

$$(\sigma_i + (1-\sigma_i)\rho) - (\sigma_j + (1-\sigma_j)\rho) = (1-\rho)(\sigma_i - \sigma_j) < 0.$$

If $e(\sigma_i) - e(\sigma_j) \geq 0$ we have an immediate contradiction, so let $e(\sigma_i) - e(\sigma_j) < 0$. Since $a_i \succ a_j$ and thus $i < j$, we have that $(1-\rho)^{i-1} \prod_{k<i} (1-\sigma_k) > (1-\rho)^{j-1} \prod_{k<j} (1-\sigma_k)$. Therefore the previous displayed equation implies

$$((\sigma_i + (1-\sigma_i)\rho) - (\sigma_j + (1-\sigma_j)\rho))(1-\rho)^{j-1} \prod_{k<j} (1-\sigma_k) > e(\sigma_i) - e(\sigma_j)$$

But then

$$\begin{aligned} & (\sigma_i + (1-\sigma_i)\rho)(1-\rho)^{j-1} \prod_{k<j} (1-\sigma_k) - e(\sigma_i) \\ > \; & (\sigma_j + (1-\sigma_j)\rho)(1-\rho)^{j-1} \prod_{k<j} (1-\sigma_k) - e(\sigma_j) \end{aligned}$$

which means that alternative j would gain by deviating from σ_j to σ_i, a contradiction.

The same argument works for the AND model. If $a_i \succ a_j$ then it must be

$$(\sigma_i \rho) \prod_{k<i} (1-\rho\sigma_k) - e(\sigma_i) \geq (\sigma_j \rho) \prod_{k<i} (1-\rho\sigma_k) - e(\sigma_j)$$

or

$$(\sigma_i - \sigma_j)\rho \prod_{k<i} (1-\rho\sigma_k) \geq e(\sigma_i) - e(\sigma_j)$$

Therefore if it were $\sigma_i < \sigma_j$ we would have either an immediate contradiction (if $e(\sigma_i) - e(\sigma_j) \geq 0$); or (if $e(\sigma_i) - e(\sigma_j) < 0$):

$$(\sigma_i - \sigma_j)\rho \prod_{k<j} (1-\rho\sigma_k) > e(\sigma_i) - e(\sigma_j)$$

which contradicts the optimality of σ_j for a_j. ■

Proof of Proposition 7: We consider the case of three alternatives and show that the position profile $v^* = (0,0,1)$ is a Nash Equilibrium. For a generic profile v, the choice probabilities are given by

$$\begin{gathered} p_{AND}(a_1, v) = \left(v_1 - \tfrac{v_2+v_3}{2}\right)^2 \rho \\ p_{AND}(a_2, v) = \left(v_2 - \tfrac{v_1+v_3}{2}\right)^2 \rho \left(1 - \left(v_1 - \tfrac{v_2+v_3}{2}\right)^2 \rho\right) \\ p_{AND}(a_3, v) = \left(v_3 - \tfrac{v_1+v_2}{2}\right)^2 \rho \left(1 - \left(v_1 - \tfrac{v_2+v_3}{2}\right)^2 \rho\right) \left(1 - \left(v_2 - \tfrac{v_1+v_3}{2}\right)^2 \rho\right) \end{gathered}$$

It is seen immediately that alternative 1's best replies to $v_2 = 0$ and $v_3 = 1$ are either corner position, i.e. $v_1 = 1$ and $v_1 = 0$, so that it cannot profitably deviate from v^*. Turning now to alternative 2, check

$$\left.\frac{\partial p_{AND}(a_2, v)}{\partial v_2}\right|_{\substack{v_1=0 \\ v_3=1}} = \frac{1}{8}\rho(2v_2 - 1)\left(8 - \rho - 5\rho v_2 - 4\rho v_2^2\right)$$

Studying the sign, it is straightforward to verify that the possible maxima are at $v_2 = 0$ and, depending on the size of ρ, either $v_2 = \frac{1}{2}$ or $v_2 = 1$. The corresponding choice probabilities are:

$$\begin{aligned} p_{AND}(a_2, v^*) &= \frac{1}{4}\rho\left(1 - \frac{1}{4}\rho\right) \\ p_{AND}\left(a_2, \left(0, \frac{1}{2}, 1\right)\right) &= 0 \\ p_{AND}(a_2, (0, 1, 1)) &= \frac{1}{4}\rho(1 - \rho) \end{aligned}$$

so that, regardless of the size of ρ, alternative 2 cannot profitably deviate from v^*.

Finally consider alternative 3:

$$\left.\frac{\partial p_{AND}(a_3, v)}{\partial v_3}\right|_{\substack{v_1=0 \\ v_2=0}} = \frac{1}{8}\rho v_3\left(\rho v_3^2 - 4\right)\left(3\rho v_3^2 - 4\right)$$

The roots of the polynomial are $v_3 = 0$, $v_3 = \pm\frac{2}{\sqrt{\rho}}$ and $v_3 = \pm\frac{2}{\sqrt{3\rho}}$, so that for $v_3 \in [0, 1]$ we have that $\left.\frac{\partial p_{AND}(a_3,v)}{\partial v_3}\right|_{\substack{v_1=0 \\ v_2=0}} > 0$ for $v_3 \in \left(0, \frac{2}{\sqrt{3\rho}}\right)$ and $v_3 > \frac{2}{\sqrt{\rho}}$, while $\left.\frac{\partial p_{AND}(a_3,v)}{\partial v_3}\right|_{\substack{v_1=0 \\ v_2=0}} < 0$ for $v_3 \in \left(\frac{2}{\sqrt{3\rho}}, \frac{2}{\sqrt{\rho}}\right)$. It follows that $p_{AND}(a_3, (0, 0, v_3))$ is maximised for $v_3 = \min\left\{1, \frac{2}{\sqrt{3\rho}}\right\} = 1$. The corresponding choice probability is

$$p_{AND}(a_3, v^*) = \frac{1}{16}\rho(\rho - 4)^2$$

It is now straightforward to verify that

$$p_{AND}(a_3, v^*) > p_{AND}(a_1, v^*) > p_{AND}(a_2, v^*)$$

■

Proof of Proposition 8: Consider again the case with three alternatives. The choice probabilities are now:

$$\begin{aligned}
p_{OR}(a_1, v) &= \left(v_1 - \frac{v_2+v_3}{2}\right)^2 + \left(1 - \left(v_1 - \frac{v_2+v_3}{2}\right)^2\right)\rho \\
p_{OR}(a_2, v) &= \\
&= \left(\left(v_2 - \frac{v_1+v_3}{2}\right)^2 + \left(1 - \left(v_2 - \frac{v_1+v_3}{2}\right)^2\right)\rho\right)\left(1 - \left(v_1 - \frac{v_2+v_3}{2}\right)^2\right)(1-\rho) \\
p_{OR}(a_3, v) &= \\
&= \left(\left(v_3 - \frac{v_1+v_2}{2}\right)^2 + \left(1 - \left(v_3 - \frac{v_1+v_2}{2}\right)^2\right)\right) \times \\
&\quad \times \left(1 - \left(v_1 - \frac{v_2+v_3}{2}\right)^2\right)\left(1 - \left(v_2 - \frac{v_1+v_3}{2}\right)^2\right)(1-\rho)^2
\end{aligned}$$

Evaluating at $v^* = (0,0,1)$ yields:

$$\begin{aligned}
p_{OR}(a_1, v^*) &= \frac{1}{4}(1+3\rho) \\
p_{OR}(a_2, v^*) &= \frac{3}{4}(1-\rho)\frac{1}{4}(1+3\rho) \\
p_{OR}(a_3, v^*) &= \frac{9}{16}(1-\rho)^2
\end{aligned}$$

It is immediately apparent that $p_{OR}(a_1, v^*) > p_{OR}(a_2, v^*)$. Moreover, for $\rho \in (0,1)$, $p_{OR}(a_3, v^*) > p_{OR}(a_1, v^*)$ if and only if $\rho < \frac{5-2\sqrt{5}}{3} < \frac{1}{3}$; and $p_{OR}(a_3, v^*) > p_{OR}(a_2, v^*)$ if and only if $\rho < \frac{1}{3}$.

To verify that v^* is an equilibrium, it is immediately checked that alternative 1 cannot profitably deviate from v^*. Turning now to alternative 2, compute:

$$\left.\frac{\partial p_{OR}(a_2, v)}{\partial v_2}\right|_{\substack{v_1=0 \\ v_3=1}} = \frac{1}{8}(\rho - 1)\left(-10(1-\rho)v_2^3 - 3(1-\rho)v_2^2 + 2(7-11\rho)v_2 - (3\rho+5)\right)$$

Assume now that $\rho < \frac{1}{3}$. This implies that $p_{OR}(a_2, (0, v_2, 1))$ can only be maximised at $v_2 = 0$ or $v_2 = 1$. The corresponding choice probabilities are

$$\begin{aligned}
p_{AND}(a_2, (0,0,1)) &= \frac{3}{16}(1+3\rho)(1-\rho) \\
p_{AND}(a_2, (0,1,1)) &= 0
\end{aligned}$$

so that $v_2 = 0$ is the best reply.

Turning finally to alternative 3:

$$\left.\frac{\partial p_{OR}(a_3, v)}{\partial v_3}\right|_{\substack{v_1=0\\v_2=0}} = \frac{1}{4} v_3 (v_3 - 2)(v_3 + 2)(1 - \rho)^2$$

and it is easy to check that $p_{OR}(a_2, (0, 0, v_3))$ is maximised in $v_3 = 0$.

The second part of the statement follows trivially from inspection of the payoff functions. ∎