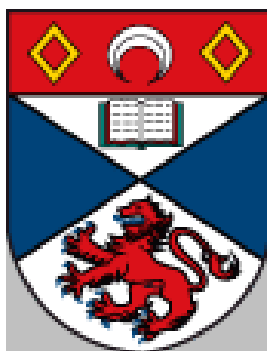


# UNIVERSITY OF ST. ANDREWS



## **A Salience Theory of Choice Errors**

**Paola Manzini and Marco Mariotti**

**No. 1003**

## **DISCUSSION PAPER SERIES**

**SCHOOL OF ECONOMICS & FINANCE**

**St. Salvator's College  
St. Andrews, Fife KY16 9AL  
Scotland**

# A Saliency Theory of Choice Errors <sup>\*</sup>

Paola Manzini

Marco Mariotti<sup>†</sup>

April 2010

## Abstract

We study a psychologically based foundation for choice errors. The decision maker applies a preference ranking after forming a ‘consideration set’ prior to choosing an alternative. Membership of the consideration set is determined both by the alternative specific *saliency* and by the *rationality* of the agent (his general propensity to consider all alternatives). The model turns out to include a logit formulation as a special case. In general, it has a rich set of implications both for exogenous parameters and for a situation in which alternatives can affect their own saliency (saliency games). Such implications are relevant to assess the link between ‘revealed’ preferences and ‘true’ preferences: for example, less rational agents may paradoxically express their preference through choice more truthfully than more rational agents.

**J.E.L. codes:** D0.

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

---

<sup>\*</sup>We thank Mauro Papi and Ivan Soraperra for helpful comments and discussions. Financial support from ESRC grant RES-000-22-3474 is gratefully acknowledged.

<sup>†</sup>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

A vast class of models of choice assume deterministic behaviour. This holds both for the classical ‘rational’ model (e.g. Samuelson [22], Richter [17]) and for more recent models of boundedly rational choice. But empirical economists have always had to confront the noisiness of the data. This raises the need to graft an appropriate *error structure* on the model, and therefore leads to the construction of a probabilistic choice model. Pioneering theoretical contributions in this area have been Luce [7] and Block and Marshak’s [2] and Marshak’s [15] Random Utility Maximization (RUM) model. In this paper we present a simple model of probabilistic choice from discrete choice sets (including RUM as a special case), with two main features:

- 1) The stochastic components of the model is given a precise interpretation.
- 2) Some parameters governing those components are endogenised via an equilibrium process.

The RUM model culminated in its most influential version, McFadden’s ([10], [11]) *conditional logit* (or *multinomial logit*) discrete choice model, in which the probability  $p(a_i, A)$  that alternative  $a_i$  is selected from a choice set  $A$  takes the form

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

where  $u(a_j)$  expresses the ‘systematic utility’ of alternative  $a_j$ .<sup>1</sup> This model is a case within a general class in which alternative  $a_i$  generates a ‘random utility stimulus’  $u(a_i) + \varepsilon_i$ , where  $\varepsilon_i$  is an error term, and is chosen over alternative  $a_j$  if  $u(a_i) - u(a_j) > \varepsilon_j - \varepsilon_i$ . In this perspective, the crucial step consists of defining an appropriate probabilistic structure on the errors. The logit model follows from the  $\varepsilon_i$  taking on i.i.d. Gumbel (or extreme value type I) distributions<sup>2</sup>. A probit model would follow instead by assuming normal distributions. In general, the basic constraint is that larger errors are made with smaller

---

<sup>1</sup>In applications, it is usually assumed further that utility is a linear function of the alternative attributes.

<sup>2</sup>Gumbel distribution function (with parameters  $\mu$  and  $\sigma$ ):  $F(x) = \exp\left(-e^{\frac{x-\mu}{\sigma}}\right)$ .

probabilities, from which it follows that better alternatives are chosen with higher probability. Closely related ideas have also found their way in modelling strategic behaviour, for the first time with McKelvey and Palfrey's ([13], [14]) notion of Quantal Response Equilibrium (QRE)<sup>3</sup>.

This approach has proved to be extremely useful for experimental and empirical economics. But it seems fair to say that - beside the basic constraint of error monotonicity in utility - the error structure is not given a clear psychological foundation. Rather, it is chosen for analytical convenience from a standard set of statistical distributions, and then is added on to a deterministic model. In other words, economic (or psychological) theory stops at the level of utility maximisation. 'Bounded rationality' follows from exogeneous random errors. This makes the error structure rather difficult to assess and interpret.<sup>4</sup>

We propose here a different route: we formulate *directly* a boundedly rational model of choice which includes some stochastic components. The main advantage of doing so is that the error structure becomes fully transparent, being part of the core model itself. The stochastic components of this model are extremely simple and, especially, can be given a precise psychological interpretation.

The model focusses on the notion of a *consideration set*. The agent does not rationally evaluate all objectively available alternatives, but only a (possibly strict) subset of them, the consideration set. Once a consideration set has been formed, a choice is made by means of a preference relation, which in this paper we assume to be standard (complete and transitive). This two-step conceptualisation of the act of choice is rooted in psychology

---

<sup>3</sup>See Goeree, Holt and Palfrey [4] for an overview.

<sup>4</sup>In his Nobel lecture, McFadden [12] recounts of how he first developed the model in response to a specific practical problem, and then sought a theoretical foundation, which he found in the Gumbel error specification within RUM theory. We recall here 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  $\max(X_i)_{i=1,\dots,N}$ , we have  $\lim_{N \rightarrow \infty} (G^M(x)) = \lim_{N \rightarrow \infty} (G(x))^N = 0$  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.

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 [8], Eliaz and Spiegler [3], Masatlioglu, Nakajima and Ozbay [9])<sup>5</sup>.

In our model the formation of the consideration set is stochastic. It is in fact the only stochastic component of the model. For an alternative, the probability of membership of the consideration set depends on two types of parameters (probabilities). The first parameter  $\rho$  ('rationality'), expresses the general propensity of the agent to consider all alternatives. The second parameter  $\sigma_i$  ('salience') is alternative specific. We present two natural models (called AND and OR) that depend on the specific way the parameters combine to determine the probability of membership of the consideration set.

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, equations of the type  $u(a_i) - u(a_j) > \varepsilon_j - \varepsilon_i$ . These equations 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 makes it easy to study how the probability of a given alternative being chosen varies with rationality. 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. Both the AND and the OR models predict either monotonic or single peaked relationships between  $\rho$  and the probability of choice, for all alternatives. The AND model predicts, for all alternatives and all parameters, an interval (and possibly the full interval) in which the probability of choice increases with rationality. The OR model predicts an increasing interval for some alternatives and some parameter configurations, but forbids increasingness on the whole range except for the best alternative. Unexpected effects may also occur (in the OR model) in respect of the *odds* (probability ratio) of choosing a better alternative over a worse alternative, which may decrease with  $\rho$ . And in both

---

<sup>5</sup>These three models are examples of the modern 'sophisticated models of choice behavior which are deterministic' we alluded to in the opening.

models the odds fail Luce’s [7] Independence of Irrelevant Alternatives test (which implies, together with other assumptions, the logit model)<sup>6</sup>. Salience enters these relationships in a non-obvious way.

In the second part of the paper we endogenise salience. We consider situations in which alternatives can influence their own salience. There are many examples that fit this case. In electoral contests, politicians make statements to get noticed by the voters, not only to persuade them. Voters may not consider certain parties for cultural reasons or out of habit<sup>7</sup> (see Wilson [24] for a consideration set approach to political competition). 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 [3] work mentioned before is the first to study in detail this type of competitive situation).<sup>8</sup>

Our main result is that, when alternatives can fully control their own salience (‘absolute salience’), in equilibrium - under very general assumptions - both models have a ‘the showiest is the best’ feature: the equilibrium ordering of salience fully 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.

All these results have a bearing for the inferences we draw on true preferences using revealed preferences reconstructed from choice data. We briefly comment on such

---

<sup>6</sup>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.

<sup>7</sup>In Wilson [24], for example, it is reported that African Americans tend to ignore Republican candidates in spite of the overlap between their policy preferences and the stance of the Republicans, and even if they are dissatisfied with the Democratic candidate.

<sup>8</sup>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).

implications in the concluding section.

## 2 Saliency and rationality

### 2.1 The Model

There is a countable (possibly finite) choice set of alternatives  $A = \{a_1, \dots, a_n, \dots\}$ . 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  $i < j$ .

While a standard rational consumer explores the entire choice set  $A$  and picks the maximal element according to  $\succ$ , here  $\succ$  is applied only to a *consideration set*  $C(A) \subseteq A$  of alternatives (the set of alternatives he actively considers). We allow for the consideration set 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).

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  $\sigma_i$  the *saliency* of alternative  $a_i$ , and a list  $(\sigma_1, \dots, \sigma_n, \dots)$  a *saliency profile*. Note that while we use the term saliency throughout for simplicity and because it tallies with leading examples, there are situations in which  $\sigma_i$  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  $\rho$  measures the general propensity of the agent to consider all alternatives. All else being equal, an agent with a higher  $\rho$  is more likely, *coeteris paribus*, to apply his preference to the entire choice set  $A$  and  $\rho$  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.

**OR model:** any alternative  $a_i$  is drawn into the consideration set with probability  $\sigma_i$ , and

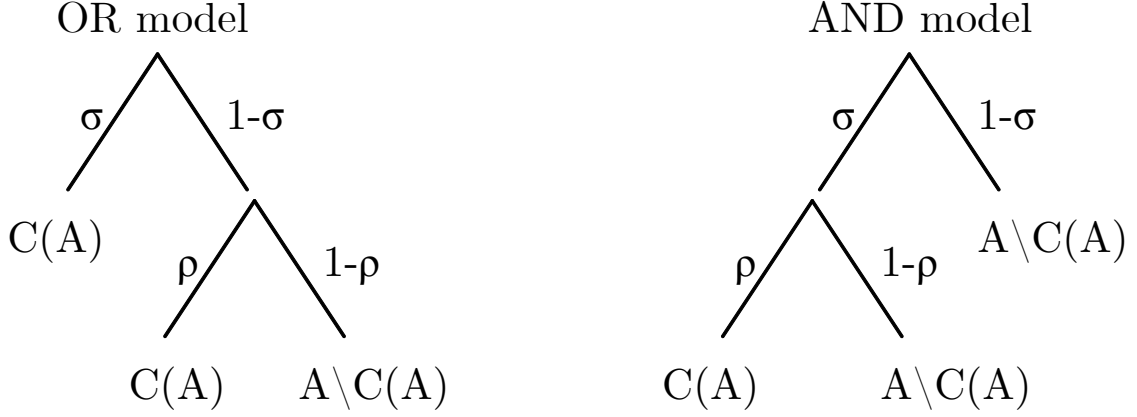


Figure 1: Structure of the two models

then all alternatives which haven't been drawn in this way are considered with probability  $\rho$ . So

$$\text{prob}(a_i \in C(A)) = \sigma_i + (1 - \sigma_i) \rho$$

**AND model:** any alternative  $a_i$  is provisionally drawn into the consideration set with probability  $\sigma_i$ , and it remains there with probability  $\rho$ . So

$$\text{prob}(a_i \in C(A)) = \rho \sigma_i$$

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

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

If  $C(A)$  is empty he chooses the default option  $a^*$ .<sup>9</sup>

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

## 2.2 The Logit Retrouvé?

There is a formal relation between these models and the RUM models. We begin by discussing a benchmark situation, in which both models collapse to a logit form. Let

---

<sup>9</sup>The default option could be replaced by a more complex procedure to arrive at a choice, notably uniform randomisation over  $A$ .



$p_{AND}(a_i, A)$  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** *There exists a utility function  $u : A \rightarrow \text{Re}$  representing  $\succ$  on  $A$ , a salience profile, 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*

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

To see this, let's write down the choice probabilities explicitly. In the OR model, the probability  $p_{OR}(a_i)$  that  $a_i \in A$  is chosen is<sup>10</sup>:

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

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

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

that is,

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

---

<sup>10</sup>Use the convention that  $\prod_{j=1}^0 (\cdot) = 1$ .

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.

This observation provides a simple new psychological foundation for an often used error specification. The OR and AND models can be seen in this perspective a class of ‘distortions’ of the conditional logit model, where 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 to that type of information. Some allowed utility transformations will destroy the loglinear relationships. In other words, at a fundamental level the log-linearity only holds with respect to the quality ranking index.

### 2.3 Limiting behavior

The limiting behavior of the two models underscores notable differences. In the limit as  $\rho$  tends to 1, for any salience profile, the OR model clearly converges to the standard model of preference maximisation: the best alternative is considered with probability one, and therefore it is chosen with probability one. On the contrary, in the AND model there may still be a positive probability (depending on the salience profile) that the better alternatives are not considered, and therefore are not chosen.

Curiously, the limiting version of the AND model for high rationality coincides with the limiting version of the OR model for low rationality,

$$\lim_{\rho \rightarrow 1} p_{AND}(a) = \lim_{\rho \rightarrow 0} p_{OR}(a) = \sigma_i \prod_{j < i} (1 - \sigma_j)$$

A moment's reflection explains this with the fact that taking these opposit limits is the way to make consideration behavior be determined solely by salience in each of two models. Observe that the quality difference effect persists in the limit.

The limiting *conditional* probabilities of choice as  $\rho$  vanishes are of interest:

$$\begin{aligned} \lim_{\rho \rightarrow 0} p_{OR}(a_i, A) &= \lim_{\rho \rightarrow 0} \frac{(\sigma_i + (1 - \sigma_i)\rho)(1 - \rho)^{i-1} \prod_{j < i} (1 - \sigma_j)}{\sum_k (\sigma_k + (1 - \sigma_k)\rho)(1 - \rho)^{k-1} \prod_{j < k} (1 - \sigma_j)} \\ &= \frac{\sigma_i}{1 + \sigma_k \sum_{k < i} \sigma_k \prod_{j < k} \frac{1}{(1 - \sigma_j)} + \sum_{k > i} \sigma_k \prod_{j < k} (1 - \sigma_j)} \end{aligned}$$

and

$$\begin{aligned} \lim_{\rho \rightarrow 0} p_{AND}(a_i, A) &= \lim_{\rho \rightarrow 0} \frac{\rho \sigma_i \prod_{j < i} (1 - \rho \sigma_j)}{\sum_k \rho \sigma_k \prod_{j < k} (1 - \rho \sigma_j)} \\ &= \frac{\sigma_i}{\sum_k \sigma_k} \end{aligned}$$

These calculations highlight a further difference from the logit (or quantal) models, which collapse to random choice as when the ‘rationality’ parameter  $\beta$  of the logit tends to 0. In general the limiting behaviour for  $\rho$  that tends to 0 is not purely random choice in our models. In both the OR and the AND model the salience difference between alternatives is preserved in the limit. In addition the quality difference between alternatives is preserved in the limit in the OR model (though not in the AND model).

The AND model obviously does collapse to random choice in the limit of low rationality in the common salience case (assume here the number of alternatives to be a finite number

$n$ ).<sup>11</sup> The OR model collapses instead to random choice in the limit of low rationality provided that common salience also vanishes:

$$\begin{aligned}\sigma_i &= \sigma \in (0, 1) \text{ for all } i \Rightarrow \\ \lim_{\rho \rightarrow 0} \lim_{\sigma \rightarrow 0} p_{OR}(a_i, A) &= \frac{1}{n}\end{aligned}$$

Note that the order matters in the above repeated limit.

## 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<sup>12</sup> 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.<sup>13</sup> However, the distribution of salience may scramble this association between quality and probability of being chosen.

---

<sup>11</sup>Though not the OR model: in the common salience case we have

$$\begin{aligned}& \lim_{\rho \rightarrow 0} p_{OR}(a_i, A) \\ &= \frac{\sigma}{2 - \sigma \left( 1 - \sum_{k < i} \frac{1}{(1-\sigma)^{i-k}} \right) - (1-\sigma)^{n-i}}\end{aligned}$$

<sup>12</sup>More precisely, a permutation of the objects in the preference order which improves the ranking of the object.

<sup>13</sup>So in particular this holds for the case of equal salience  $\sigma_i = \sigma$  for some common  $\sigma \in [0, 1]$ .

Because  $p_{OR}(a_i)$  and  $p_{AND}(a_i)$  are  $i^{th}$  degree polynomials in  $\rho$ , the effect of an increase in rationality is more subtle. 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 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 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 Who gains from rationality?

#### 3.1 OR model

In this section we discuss how the probability of choice for an alternative can vary non-monotonically (but with at most one peak) in the OR model as rationality increases.

Compute

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

which is ambiguous in sign. The decomposition highlights the source of ambiguity. On the one hand, an increase in  $\rho$  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; on the other hand, it also increases the probability that better alternatives are considered.

Defining

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

we have

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

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

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  $\rho \in (0, 1)$ . It is always increasing for  $i = 1$ . For any  $i > 1$ ,  $p_{OR}(a_i)$  is strictly increasing on an initial range iff  $i\sigma_i < 1$ , and for  $\rho$  sufficiently high,  $p_{OR}(a_i)$  is strictly decreasing.

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.

An agent with higher rationality may be *less* attracted by a good alternative (but not the best) if it has high salience.

Note that only own salience, and not the salience of the other alternatives, affects the value of  $\bar{\rho}$  and thus the sign of the derivative. The effect of own salience on  $\bar{\rho}$  (namely  $i - i^2$ ) is negative for  $i > 1$  and zero for  $i = 1$ . A lower quality (increase in  $i$ ) reduces the threshold  $\bar{\rho}$ . The quality effect and the salience effect, as well as examples of choice probabilities peaking at intermediate degrees of rationality for 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.

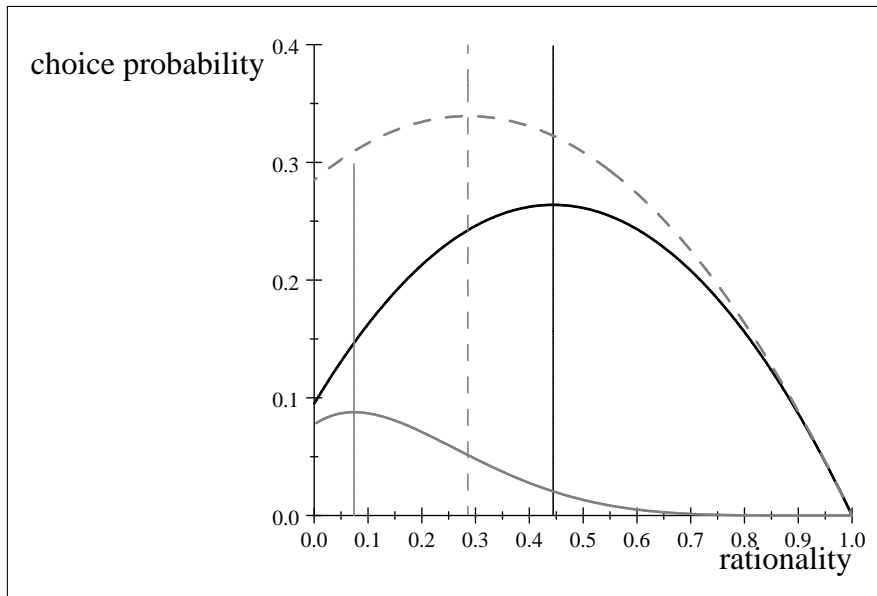


Figure 2: Comparative statics in the OR model: increasing  $\sigma_i$  shifts  $p_{OR}(a_i)$  upwards and to the left; increasing  $i$  shifts  $p_{OR}(a_i)$  downwards and to the left.

### 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.

In order to highlight the role of  $\rho$  it is instructive to rewrite the model with the following notation. Let  $S(m, k)$  denote the ordered set of combinations of  $k$  elements from the set  $\{\sigma_1, \dots, \sigma_m\}$ , where  $|S(m, k)| = \binom{m}{k}$ , with all of the elements in  $S(m, k)$  listed in ascending order lexicographically. Finally, let  $s_{m,k} = \{1, 2, \dots, \binom{m}{k}\}$  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

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

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.  $S(3, 2)(1) = \sigma_1 \sigma_2$ ,  $S(3, 2)(2) = \sigma_1 \sigma_3$ ,  $S(3, 2)(3) = \sigma_2 \sigma_3$ , 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:

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

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  $\rho$  on the probability of choice is ambiguous, but here there clearly exists  $\hat{\rho} \in (0, 1)$  such that  $\frac{\partial p_{AND}(a_i)}{\partial \rho} > 0$  if  $\rho < \hat{\rho}$ . But 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 (by taking away choice probability from the default alternative  $a^*$ ).

We now show that the threshold  $\hat{\rho}$ , when it exists in  $(0, 1)$ , is unique.

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

All (easy but mostly tedious) missing proofs are relegated to a separate section.

The latter part of the statement highlights a major difference from the OR model: in the AND model increases in rationality can be good news for inferior alternatives at *all* levels of rationality, something which cannot happen in the OR model.

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

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

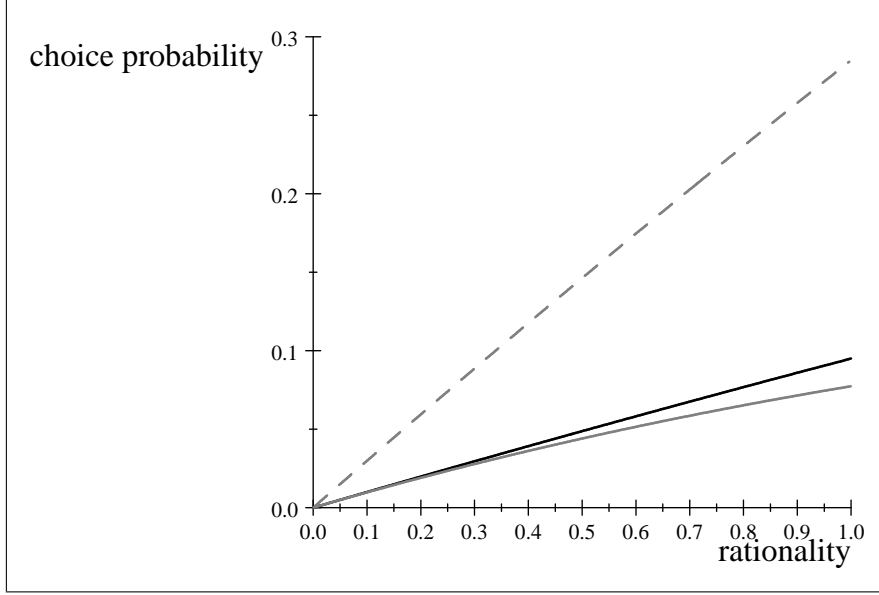


Figure 3: Comparative statics in the AND model: increasing  $\sigma_i$  shifts  $p_{OR}(a_i)$  upwards; increasing  $i$  shifts  $p_{OR}(a_i)$  downwards. With these parameter values  $\rho$  is always increasing over the  $(0, 1)$  interval.

But the conjecture is 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** *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 [7] 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<sup>14</sup>, this holds true only for changes in  $A$  which remove or delete alternatives each of which is either better or worse 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

<sup>14</sup>Imagining now that they apply to a collection of subsets of a universal set of objects  $X$ .

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  $(\sigma_i + (1 - \sigma_i)\rho)(\sigma_j + (1 - \sigma_j)\rho)$  in the OR model and  $(\sigma_i\rho)(\sigma_j\rho)$  in the AND model (namely the probabilities that  $a_i$  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 odds on the other available alternatives is often a realistic feature, which applied economists have sought to incorporate, for example, in the conditional logit model.<sup>15</sup> 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.

## 4 Saliency games

### 4.1 Absolute saliency: The showiest is the best

We now imagine that alternatives can choose, possibly at a cost, the saliency 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.

---

<sup>15</sup>By adding a nested structure to choice process (nested logit) or by allowing heteroscedasticity of the choice errors. See e.g. Greene [5].

We are mainly interested in the question of how the equilibrium salience order correlates with 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  $e : S \rightarrow \text{Re}$ . 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,  $e$  can be increasing or decreasing. So  $e$  could be interpreted as effort, when increasing salience is costly, or as elation, when increasing salience is pleasurable.

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

The proof makes clear that this pure strategy existence result continues to hold when  $S$  is a compact subset of  $[0, 1]$  and  $e(\cdot)$  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  $e(\cdot)$ .

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

**Proposition 6** (The showiest is the best) *Suppose  $a_i \succ a_j$  and let  $(\sigma_1, \dots, \sigma_n)$  be a pure strategy equilibrium. Then, for all  $\rho$ ,  $\sigma_i > \sigma_j$  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. The reason why they produce less salience in equilibrium does not derive from lower levels of resources or lower unit costs of salience production (as might be the case in a signalling story): every alternative can choose from exactly the same set at exactly the same cost or benefit.

## 4.2 Relative salience: the ugly duckling can get picked most often

So far we have assumed that each alternative can 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  $\sigma_i$  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** *The AND model admits (for some  $n$ ) a pure strategy Nash equilibrium in which, for any  $\rho$ , the worst alternative has the highest probability of being chosen.*

**Proposition 8** *The OR model admits (for some  $n$ ) a pure strategy Nash equilibrium in which, for  $\rho$  sufficiently small, the worst alternative has the highest probability of being chosen. For  $\rho$  sufficiently high (for any  $n$ ) the best alternative is chosen with the highest probability.*

The difference between these two results stems from the difference between the AND and the OR model we highlighted before: at degrees of rationality near one, the OR model - but not the AND model - approximates well the standard utility maximisation model.

## 5 Proofs

**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

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

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

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

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

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

Therefore

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

so that

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

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  $\sigma_1, \dots, \sigma_{i-1}$  sufficiently low (note that  $A(i, j)$  approximates  $\sigma_i \rho$  when  $\sigma_1, \dots, \sigma_{i-1}$  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  $\rho$  and  $\sigma_j$  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

$$\max_{\sigma_1 \in S_1} (\sigma_i + (1 - \sigma_i) \rho) - e(\sigma_i)$$

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

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

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  $\sigma_i$  is optimal for alternative  $a_i$ , it must provide a weakly higher expected payoff than  $\sigma_j$ , 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  $(\sigma_i + (1 - \sigma_i) \rho) - (\sigma_j + (1 - \sigma_j) \rho) < 0$ . Furthermore, 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  $\sigma_j$  to  $\sigma_i$ , a contradiction.

The same argument works for the AND model. If  $a_i \succ a_j$  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

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

which contradicts the optimality of  $\sigma_j$  for  $a_j$ . ■

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

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

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

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

Studying the sign, it is straightforward to verify that the possible maxima are at  $v_2 = 0$  and, depending on the size of  $\rho$ , 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  $\rho$ , alternative 2 cannot profitably deviate from  $v^*$ .

Finally consider alternative 3:

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

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:

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

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  $p_{OR}(a_3, (0, 0, v_3))$  is maximised in  $v_3 = 0$ .

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

## 6 Concluding remarks and related literature

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. Admittedly, we have opened just one box. Other explanations, beside consideration sets, may be relevant. Recently Rubinstein and Salant [21] have studied 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 Spiegel [3] 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 an application of Masatlioglu, Nakajima and Ozbay [9], which is deterministic. Eliaz and Spiegel [3] also perform comparative statics exercises that relate to changes in rationality. One the main findings is that in some equilibria 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.

A first general message from our paper is that ‘revealed preferences’ are not necessarily a better guide to discovering true preferences when the rationality of the agent is higher (namely when the agent has a higher probability of being better informed about the available alternatives). For example, suppose you can observe or infer the degrees of rationality,  $\rho$  and  $\rho'$ , under which two sets choices were made. Suppose that alternative  $x$  is chosen more frequently over alternative  $y$  in condition  $\rho$  than in condition  $\rho'$ : it does not necessarily follow from the fact that  $\rho > \rho'$  that  $x$  is more likely to be 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.

A second message is that the revealed preference ranking coincide (probabilistically)

with the true preference relation in only two cases: if (1) salience is exogenous, or if (2) salience is endogenous but can be fully set by the alternatives (a yellow dress is salient). But the revealed preference ranking can even reverse the true ranking when salience is endogenous and is relative (a yellow dress is salient when all other dresses are green).

## References

- [1] Agresti, A. (2002) *Categorical Data Analysis*, John Wiley and Sons, Hoboken, NJ.
- [2] 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.
- [3] Eliaz, K. and R. Spiegler (2009) “Consideration Sets and Competitive Marketing”, mimeo, New York University and University College London.
- [4] Goeree, Jacob K., Charles A. Holt and Thomas R. Palfrey (2008) “Quantal response equilibria”, in Steven N. Durlauf and Lawrence E. Blume (eds.), *The New Palgrave Dictionary of Economics*, Second Edition.
- [5] Greene, William H. (2003) “Econometric Analysis”, Prentice Hall, Upper Saddle River, NJ.
- [6] Hauser, J., & Wernerfelt, B. (1990) “An evaluation cost model of consideration sets” *Journal of Consumer Research*, 16(4), 393–408.
- [7] Luce, R. Duncan (1959) *Individual Choice Behavior*, Wiley: New York.
- [8] Manzini, Paola and Marco Mariotti (2007) “Sequentially Rationalizable Choice”, *American Economic Review*, 97 (5): 1824-1839.
- [9] Masatlioglu, Y., D. Nakajima and E. Ozbay (2009): “Revealed Attention”, mimeo.

- [10] McFadden, D.L. (1974) “Conditional Logit Analysis of Qualitative Choice Behavior”, in P. Zarembka (ed.), *Frontiers in Econometrics*, p. 105-142, Academic Press: New York.
- [11] McFadden, D.L. (1974) “The Measurement of Urban Travel demand”, *Journal of Public Economics* 3: 303-28.
- [12] McFadden, D.L. (2000) “Economic choices”, *American Economic Review*, 91 (3): 351-378.
- [13] McKelvey, R. D. and T. R. Palfrey R. (1995) “Quantal Response Equilibrium for Normal Form Games,” *Games and Economic Behavior*, 10, 6-38.
- [14] McKelvey, R. D. and T. R. Palfrey R. (1998) “Quantal Response Equilibria for Extensive Form Games,” *Experimental Economics* 1: 9-41.
- [15] Marschak, J. (1960) “Binary choice constraints and random utility indicators”, Cowles Foundation Paper 155.
- [16] Nedungadi, P. (1990). Recall and consumer consideration sets: Influencing choice without altering brand evaluations. *Journal of Consumer Research*, 17(3), 263–276.
- [17] Richter, M. K., 1966, “Revealed Preference Theory”, *Econometrica*, 34: 635-645.
- [18] Roberts, J., & Lattin, J. M. (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, 28(4), 429–440.
- [19] Roberts, J. H., & Lattin, J. M. (1997). Consideration: Review of research and prospects for future insights. *Journal of Marketing Research*, 34(3), 406–410.
- [20] Roberts, J., & Nedungadi, P. (1995). Studying consideration in the consumer decision process. *International Journal of Research in Marketing*, 12, 3–7.
- [21] Rubinstein, Ariel and Yuval Salant (2009) “Eliciting Welfare Preferences from Behavioral Datasets”, mimeo, University of Tel Aviv Cafes, Tel Aviv university, New York University and Northwestern University.

- [22] Samuelson, Paul A. (1938) “A Note on the Pure Theory of Consumer’s Behaviour”, *Economica*, 5 (1): 61-71.
- [23] Shocker, A. D., Ben-Akiva, M., Boccara, B., & Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Letters*, 2(3), 181–197.
- [24] Wilson, Carole J. (2008) “Consideration Sets and Political Choices: A Heterogeneous Model of Vote Choice and Sub-national Party Strength”, *Political Behavior* 30: 161–183.