Logit Function in Stochastic Categorizations

Kai Wang*

Abstract

Binary categorizations refer to the behaviour for which a decision maker puts the alternatives from a menu into binary categories (authorization category and rejection category). For example, a judge approves/denies parole for inmates, a doctor administers/withholds treatment to patients, a teacher passes/fails essays. Categorizations are observed to exhibit randomness in various contexts, but the stochasticity in categorizations has received little attention in economic studies. We consider stochastic categorizations using a logit categorization function, which expresses the probability of authorizing an alternative from a menu in a 'logit-like' form. We characterize the model from a simple condition on authorization frequencies. Furthermore, we derive an empirical version of our model, which not only provides insight into applications of the model, but also gives an intuitive interpretation of the stochasticity in categorizations. We use simple examples to show that our model can bring new insights and research possibilities in economic studies.

Keywords: logit model; stochastic categorizations.

JEL codes: C10, D00.

*Department of Economics, Business School, University of Sussex, Falmer, Brighton, BN1 9SL, UK, email: Kai.Wang@sussex.ac.uk.

1 Introduction

Consider the following cases:

(i) A judge approves or denies parole for inmates;

(ii) A doctor administers or withholds treatment to patients;

(iii) A teacher passes or fails essays.

These are all binary categorization problems in which a decision maker (DM) decides which alternatives belong to a certain category. The DM assigns each individual alternative from the menu into binary categories: 'authorization' category and 'rejection' category (e.g., approval & denial, treatment & withholding, pass & fail, etc.).

Analogous to the observed stochastic choice behaviour in experimental and market settings,[1] it is common to observe categorizations that exhibit variability. In Danziger, Levav, and Avnaim-Pesso (2011), it is reported that a judge was between two and six times as likely to approve parole if the inmate was one of the first three of six considered after a food break. Elstein (1988) finds that a doctor was less likely to administer a particular treatment protocol if able to recall a case of failed treatment. Along similar lines, Gibbons and Marshall (2010) show that even with a detailed marking rubric created by the teachers themselves, the same essay was marked differently. While there is extensively rich literature on stochastic choice, the stochasticity in binary categorizations has received little attention insofar.

In this paper, we assume authorization responses to be given by a function p that indicates the probability $p(a, B)$ that alternative a is 'authorized' from menu B. This authorization probability itself is new: unlike choice probability, it does not impose the adding-up constraint $\sum_{b\in B} p(a, B) = 1$. The reason is simple and clear: a DM is not required to authorize one and only one alternative from a menu as in choice making. In this paper, we study the stochasticity in categorizations by expressing the authorization probability

[1] See the discussion in Mcfadden (2000).

$p(a, B)$ in a 'logit like' form which is similar to the choice probability in Luce (1959). We name our model the *logit categorization function.* We find that a simple independence of irrelevant alternatives (IIA) axiom on the ratio of authorization frequency to rejection frequency can fully characterize the logit categorization function. This IIA axiom on the authorization to rejection ratio is intuitive and gives insightful interpretation of the behavioural patterns in stochastic categorizations in our model.

Mcfadden (1974) develops the multinomial logit model, which is shown to be equivalent to the Luce choice model in Luce (1959), but is much more popular in econometric studies as it can incorporate the characteristics of the alternatives. Inspired by the derivation of the multinomial logit model, we further derive a full-fledged version of the logit categorization function and investigate the empirical implications of the full-fledged model. In the full-fledged model, we assume that a DM relies on 'thresholds' to solve categorization problems, such that he or she 'authorizes' an alternative from a menu if the utility of the alternative passes the threshold of the menu, and 'rejects' it otherwise. Following this set-up, the probability of authorizing alternative a from menu B, is the probability that the utility of alternative a exceeds the threshold of menu B. The full-fledged model provides intuitive explanation of the source of stochasticity in categorizations, i.e., the randomness in the utilities of alternatives and the randomness in the thresholds of menus. As with the classical Luce-Multinomial Logit Choice model, we prove that the full-fledged model is equivalent to the logit categorization function. This full-fledged model is important, as it has great potential for use in applied work by making use of the observed characteristics of the available alternatives and menus. Our framework is not only applicable to real world dataset, but also adds value to the current literature. To demonstrate this, we discuss how our model can be employed to study choice overload and availability bias through simple examples.

The logit categorization function is a novel model which allows the thresholds to be not only stochastic, but also menu dependent. Analogous to the findings in classical choice models,[2] non-stochastic utilities and non-stochastic thresholds would naturally generate a fully rational deterministic categorization correspondence. However, a deterministic categorization correspondence cannot explain the variability in categorizations. Then, the binary logit model, which implicitly assumes the threshold to be menu independent, can potentially solve the stochastic categorization problem.[3] Nevertheless, it is plausible to assume that in some cases, the thresholds are dependent on the menus. For example, a consumer may be less likely to consider an alternative if there are more options, a marker may be stricter if the average qualities of the essays are better, a college admission team can have different thresholds in different admission rounds. The logit categorization function can accommodate the categorizations with menu-dependent thresholds. The standard logit model can be applied to study categorizations if it considers the object of choice to be a set of alternatives. For example, let $B = \{a, b\}$, a standard logit model can redefine the menu as $\mathcal{B} = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$, the power set of B. Then, for instance, authorizing a and b from menu B is equivalent to choosing $\{a, b\}$ from $\mathcal{B}$. [4] However, despite its appeal, this approach is computationally demanding, as the newly defined menu sizes grow exponentially.[5] This burden on the computation makes the standard logit model unapplicable if the sizes of the menus are large. In contrast, our model is compatible with large menus as the menu sizes stay at the same level.

This paper is highly inspired by the pioneering work on sequential approval in Manzini, Mariotti, and Ulku (2021) (henceforth, MMU). MMU (2021) study

[2]See, e.g., Samuelson (1938) and Richter (1966).

[3]In fact, the binary logit model has been used in prior research. For example, the consideration set model in Goeree (2008) essentially uses binary logit model to denote the probability that an alternative is put into the consideration set (authorization).

[4]Manski and Sherman (1980) use this method to distinguish between households with one car or two cars.

[5]Specifically, menu B with J number of elements is redefined as the power set $\mathcal{B}$ with 2^J number of elements.

sequential approval behaviour (e.g., 'liking' a post on social media, 'starring' a website page on an internet browser, or 'swiping right' for a potential partner on a dating app), which essentially can be taken as categorizations. We decide to use the terminology 'authorization' instead of 'approval' in reference to the work of MMU (2021) to highlight the difference: MMU (2021) focus on online sequential approvals in which the rankings of the alternatives play a role, whereas our model focuses on authorization behaviour in which the rankings do not matter. The model in MMU (2021) is quite different from ours, as the domain comprises lists, which are different orders of the same menu. Also, MMU's model does not allow for variations of menus, as their menu is fixed at the universal set X. Our model does not require the observation of orders, and it is compatible for categorizations from different menus.

The remainder of the paper is structured as follows. Section 2 introduces the logit categorization function. Next, in Section 3, we provide a full characterization of the logit categorization function with a simple axiom. In Section 4, we derive the full-fledged version of the logit categorization function, and focus on the empirical applicability of the model. This section also provides an intuitive explanation of the source of stochasticity in our model. Finally, Section 5 concludes the paper.

2 The Model

We denote by X the universe of alternatives, and by $\mathcal{D}$ the domain of subsets (i.e., the menus) of X. We first define a random categorization function.

Definition 1. A random categorization function is a map: $p : X \times \mathcal{D} \to [0,1]$ such that $p(a, B) = 0$ for all $a \notin B$, for all $B \in \mathcal{D}$; and $p(a, B) \in (0,1)$ for all $a \in B$, for all $B \in \mathcal{D}$.

We assume categorization responses to be given by a map p that indicates the probability $p(a, B)$ alternative a is authorized from menu B, which is similar to choice probability in the work of Luce (1959), and Mcfadden (1974).[6] Definition 1 is very general, and only requires $p(a, B) = 0$ for all $a \notin B$, for all $B \in \mathcal{D}$ (i.e., only items in the menu can be authorized); and $p(a, B) \in (0, 1)$ for all $a \in B$, for all $B \in \mathcal{D}$ (i.e., every item in a menu has a positive probability of being authorized). Note that, unlike standard stochastic choice function, we do not impose the constraint $\sum_{b \in B} p(b, B) = 1$. The reason is straightforward: it is not required for an agent to authorize one and only one alternative from a menu as in choice making. Then, we define a new type of random categorization function that has a 'logit-like' expression.

Definition 2. A random categorization function is a logit categorization function if there exists a pair (V, W), where V is a map $V : X \to \mathbb{R}_{++}$, and W is a map $W : \mathcal{D} \to \mathbb{R}_{++}$, such that

$$p_{V,W}(a, B) = \frac{V(a)}{V(a) + W(B)}.$$

Here $V(a)$, $W(B)$ can be interpreted as the 'response strength' associated with alternative a and threshold of menu B.[7] The logit categorization function is structurally similar to the standard choice function $p(a, B) = \frac{V(a)}{\sum_{b \in B} V(b)}$ in Luce (1959). Indeed, $p_{V,W}(a, B)$ can be seen as the probability that alternative a is preferred to the threshold of menu B according to 'Luce rule'. Note that we map a menu B to a value $W(B)$ by the function W, so that we take account of the menu dependency of the threshold.[8] The randomness in the categorizations is simply provided by logit-like probabilities in authorizations, and it does not necessitate any menu-dependency of the threshold.

[6] Obviously, $1 - p(a, B)$ can be interpreted as the rejection probability of a from menu B.

[7] The term 'response strength' is borrowed from Gulliksen (1953), and Luce (1959).

[8] This mapping from a menu to a value is similar to the set-up in Manzini, Mariotti, and Tyson (2013) & (2016).

With menu independent thresholds $W(A) = W(B)$ for all $A, B \in \mathcal{D}$, the categorizations in our model still display stochasticity. Indeed, we provide an underlying justification for the presence of stochasticity (i.e., due to the unobserved parts of utilities and thresholds) in the discussion of the full-fledged logit categorization model in Section 4.

3 Characterization

In this section, we characterize a logit categorization function from a simple condition on observed authorization frequencies. The axiom is intended for all $A, B \in \mathcal{D}$, for all $a, b \in A \cap B$. Our axiom puts a constraint on the ratio

$$\gamma(a, B) = \frac{p(a, B)}{1 - p(a, B)}, \tag{1}$$

which is the authorization probability of alternative $a \in B$ from menu $B \in \mathcal{D}$ over the respective rejection probability. We name $\gamma(a, B) = \frac{p(a,B)}{1-p(a,B)}$ the Authorization to Rejection ratio (AR ratio) for alternative a from menu B. Given definition 1, we have $p(a, B), 1 - p(a, B) \in (0, 1)$ for all $a \in B$, for all $B \in \mathcal{D}$. Therefore, the AR ratio γ is positive and well defined for all $a \in B$, for all $B \in \mathcal{D}$. We have an independence of irrelevant alternatives (IIA) axiom on the AR ratios.

Axiom 1. *(γ-IIA)* $\frac{\gamma(a,A)}{\gamma(a,B)} = \frac{\gamma(b,A)}{\gamma(b,B)}$

By γ-IIA, $\frac{\gamma(a,A)}{\gamma(a,B)}$ (i.e., the AR ratio for alternative a from menu A, over the AR ratio for alternative a from menu B) is not dependent on the alternative a at stake, but only on the two menus A and B that have been considered.

Our first main result is the following.

Theorem 1. *A random categorization function p is a logit categorization function $p_{V,W}$ if and only if p satisfies γ-IIA.*

Proof. Sufficiency: As discussed at the beginning of this section, the AR ratio γ is positive and well defined. For all $a \in B$, and $B \in \mathcal{D}$, define the two maps in $p_{V,W}$ such that $V(a) = \gamma(a, X)$, and $W(B) = \frac{\gamma(a,X)}{\gamma(a,B)}$. By γ-IIA, $\frac{\gamma(a,X)}{\gamma(a,B)} = \frac{\gamma(b,X)}{\gamma(b,B)}$ for all $a, b \in B$, and $B \in \mathcal{D}$. So, $W(B) = \frac{\gamma(a,X)}{\gamma(a,B)} = \frac{\gamma(b,X)}{\gamma(b,B)}$ is fixed for a menu $B \in \mathcal{D}$, with $a, b \in B$, and is well defined. Then

$$p_{V,W}(a, B) = \frac{\gamma(a, X)}{\gamma(a, X) + \frac{\gamma(a,X)}{\gamma(a,B)}} \tag{2}$$

$$= \frac{1}{1 + \frac{1}{\gamma(a,B)}} = \frac{\gamma(a, B)}{\gamma(a, B) + 1} \tag{3}$$

$$= \frac{p(a, B)}{p(a, B) + 1 - p(a, B)} = p(a, B) \tag{4}$$

Necessity: Note that $V(a)$, and $W(B)$ are strictly positive, so $p_{V,W}(a, B) = \frac{V(a)}{V(a)+W(B)} \in (0, 1)$; and we have

$$\frac{\gamma(a, A)}{\gamma(a, B)} = \frac{p(a, A)}{1 - p(a, A)} \frac{1 - p(a, B)}{p(a, B)} \tag{5}$$

$$= \frac{V(a)}{W(A)} \frac{W(B)}{V(a)} = \frac{V(b)}{W(A)} \frac{W(B)}{V(b)} \tag{6}$$

$$= \frac{p(b, A)}{1 - p(b, A)} \frac{1 - p(b, B)}{p(b, B)} = \frac{\gamma(b, A)}{\gamma(b, B)} \tag{7}$$

□

The standard IIA property ($\frac{p(a,A)}{p(a,B)} = \frac{p(b,A)}{p(b,B)}$) in Luce (1959) and Mcfadden (1974) is criticized for imposing a restrictive structure on substitution patterns. The well known blue bus/red bus example in Debreu (1960) shows that IIA is inappropriate for menus containing close substitutes.[9] A logit

[9]Gul, Natenzon, and Pesendorfer (2010), Manzini and Mariotti (2014) develop different

categorization function is fully characterized by γ-IIA instead of Luce's IIA, which immediately makes it compatible for violations of IIA. However, the problem of IIA in a choice does not necessarily arise in a categorization. This is because an alternative does not have to 'make room' for its close substitute in the dimension of probabilities, since the authorization probabilities of alternatives in a menu do not have to add up to one. Instead, we discuss the limitation of γ-IIA, the axiom that characterizes our model, among categorizations behaviour.

For a logit categorization function $p_{V,W}$, the odds $\frac{\gamma(a,A)}{\gamma(a,B)} = \frac{\gamma(b,A)}{\gamma(b,B)} = \frac{W(B)}{W(A)}$, can be interpreted as the ratio of menu B's threshold to menu A's threshold. Here, γ-IIA states that this ratio is independent on the alternatives being considered. When this ratio is observed to be dependent on the alternatives at stake, the γ-IIA property is violated. Here is an example.

Example 1. (Violation of γ-IIA) A user shares post(s) on a social media website. First assume two posts a, b by influencer 1 (denoted by a_1, b_1) are on the website (menu $C = \{a_1, b_1\}$), and let $p(a_1, C) = p(b_1, C) = \frac{2}{3}$.[10] Next, assume another post b by influencer 2 (denoted by b_2) is added (new menu $D = \{a_1, b_1, b_2\}$). Here b_1 and b_2 are the same post, and the only difference is that they have been uploaded by different influencers. If the agent does not care about the influencer, one might expect the agent to share (authorize) post a_1 with same probability as before; and split the previous authorization probability equally for b_1 and b_2, as it makes little sense sharing both of them. Therefore, it is expected that $p(a_1, D) = p(a_1, C) = \frac{2}{3}$, and $p(b_1, D) = p(b_2, D) = \frac{p(b_1,C)}{2} = \frac{1}{3}$.

This is a violation of γ-IIA, as $\frac{\gamma(a_1,C)}{\gamma(a_1,D)} = 1$, but $\frac{\gamma(b_1,C)}{\gamma(b_1,D)} = 4$. Thus, it cannot be accommodated by our model. The thresholds of menus C and D for the authorization of alternative a_1 are the same, as $p(a_1, C) = p(a_1, D)$;

stochastic choice models to accommodate the violation of IIA in choice making.

[10] Again the authorization probabilities do not add up to one, as the agent is not committed to sharing only one post.

however, the thresholds of menus C and D for the authorization of b_1 are different, as $\frac{p(b_1,C)}{2} = p(b_1, D)$. The user applies the same thresholds for the authorizations of a_1 from menus C and D; but uses different thresholds for the authorizations of b_1 from menus C and D. In this case, the ratio of menu C's threshold to menu D's threshold becomes dependent on the alternatives being considered, which leads to the violation of γ-IIA.

Nevertheless, we ought to point out the 'narrowness' of this example, as we implicitly makes the user 'choose' between b_1 and b_2. In so doing, a problem similar to the close substitutes in blue bus/red bus example comes up again. If the user does not have to choose between b_1 and b_2 (i.e., the user shares any post he finds better than the threshold), we would expect $p(a_1, C) = p(b_1, C) = p(a_1, D) = p(a_1, C) = \frac{2}{3}$. Then, γ-IIA is still satisfied.[11]

4 Full-Fledged Logit Categorization Model

In this section, we derive a full-fledged version of logit categorization function from simple adaptations of the multinomial logit model in Mcfadden (1974). What is more, these adaptations are in line with the differences between choices and categorizations behaviour. In particular, we discuss this full-fledged model from the perspective of an econometrician, and show how it can be used for empirical analysis.

The full-fledged model assumes random utilities and random thresholds. Random utilities are commonly assumed in stochastic choice models.[12] However, random categorizations can be due to not only random utilities, but also random thresholds. For example, a parole judge can be stricter if hungry, a doctor can be more prudent if recalling a failed treatment, and a teacher's grading criteria can be subjective to his or her understanding of the marking

[11]For simplicity purpose, here we assume the thresholds of menus C and D to be the same. Note that γ-IIA can still hold if the thresholds of menus C and D are different.

[12]See, e.g., Block and Marschak (1960), Mcfadden (1974).

rubric. The thresholds appear to be stochastic if these factors (i.e., hungriness of the judge, the doctor's recall of a failure, and the teacher's understanding of the marking rubric) cannot be observed. Indeed, we capture the randomness of thresholds by using an error term as in Mcfadden (1974).

In empirical research, we are usually concerned with the categorizations made by individuals indexed by $i = 1, 2, ..., n$. Let $p(a|i, B)$ denote the probability that DM i from the population will authorize alternative $a \in B$, given that he or she faces menu $B \in \mathcal{D}$. The adaptations are naturally in line with the essential difference between a categorization problem and a choice problem: in denoting by $t(B, i)$ the threshold of menu B for DM i, we define $p(a|i, B)$ as $\Pr\{u(a, i) \geq t(B, i)\}$ (i.e., the probability that the utility of alternative a is (weakly) higher than the threshold of menu B); rather than the choice probability $p(a|i, B) = \Pr\{u(a, i) \geq u(b, i) : \forall b \in B\}$ in standard multinomial logit model in Mcfadden (1974).

As in Mcfadden (1974), we assume that the utility of an alternative is the sum of a representative utility and an error term. The utility of alternative a for DM i is,

$$u(a, i) = v(a, i) + \varepsilon(a, i), \tag{8}$$

where $v(a, i)$ is non-stochastic and is the 'representative utility' of alternative a for DM i, and $\varepsilon(a, i)$ is a stochastic error term, which reflects the unobserved utility of alternative a for DM i. What is more, we define

$$t(B, i) = w(B, i) + \varepsilon(B, i), \tag{9}$$

as the threshold of menu $B \in \mathcal{D}$ for DM i. Note that $w(B, i)$ is non-stochastic and is the 'representative threshold' of menu $B \in \mathcal{D}$ for DM i, and $\varepsilon(B, i)$ is the random error term that reflects the unobserved part of the threshold of menu B for DM i. Note here that the unobserved error term $\varepsilon(B,i)$ is not a function of $a \in B$, the alternative being considered. For DM i, the utilities of all the alternatives $a \in B$ in menu B, $u(a,i)$, are compared against the same threshold, $t(B,i)$. Namely, once $t(B,i)$ is generated, it is used for the categorizations of all $b \in B$ among menu B.

As seen from equation (8) and (9), the stochasticity of authorizations comes from the unobserved error term $\varepsilon(a,i)$ in a's utility for DM i, as well as the unobserved error term $\varepsilon(B,i)$ in menu B's threshold for DM i. We assume that the unobserved random part $\varepsilon(a,i)$ and $\varepsilon(B,i)$ are i.i.d. with type-1 extreme distribution.

The second main result of this paper is proving that, under the assumptions in Section 4, the authorization probability has a logit categorization function representation.

Theorem 2. *Suppose that authorization probability of alternative* $a \in B$ *from menu* $B \in \mathcal{D}$ *by DM* i *is given by the function* $p(a|i,B) = \Pr\{u(a,i) \geq t(B,i)\}$. *Also, suppose each member* i *of a population has a utility function* $u(a,i) = v(a,i) + \varepsilon(a,i)$, *and a threshold function* $t(B,i) = w(B,i) + \varepsilon(B,i)$, *where* v *is a non-stochastic function reflecting 'representative' tastes for the alternatives,* w *is a non-stochastic function reflecting 'representative' tastes for the menus, and* ε *is a function that varies randomly in the population, such that the values* $\varepsilon(a,i)$ *and* $\varepsilon(B,i)$ *are i.i.d. with type-1 extreme distribution. Then, the authorization probability has a logit categorization function representation.*

Proof. The probability that agent i with menu $B \in \mathcal{D}$, will authorize $a \in B$ equals

$$p(a|i,B) = \Pr[u(a,i) \geq t(B,i)] \tag{10}$$
$$= \Pr[v(a,i) + \varepsilon(a,i) \geq w(B,i) + \varepsilon(B,i)] \tag{11}$$
$$= \Pr[\varepsilon(a,i) \geq \varepsilon(B,i) + w(B,i) - v(a,i)]. \tag{12}$$

The values $\varepsilon(a,i)$ and $\varepsilon(B,i)$ are both i.i.d. with type-1 extreme distribution. The density is $f(\varepsilon) = e^{-\varepsilon}e^{-e^{-\varepsilon}}$, and the cumulative distribution is $F(\varepsilon) = e^{-e^{-\varepsilon}}$. If $\varepsilon(B,i)$ is given, we can write the conditional probability as $p(a|i,B)|\varepsilon(B,i) = 1 - e^{-e^{-(\varepsilon(B,i)+w(B,i)-v(a,i))}}$. However $\varepsilon(B,i)$ is not given, instead we get $p(a|i,B)$ from the integral of $p(a|i,B)|\varepsilon(B,i)$ over all values of $\varepsilon(B,i)$ weighted by its density:

$$p(a|i,B) = \int_{-\infty}^{\infty}\left(1 - e^{-e^{-(\varepsilon(B,i)+w(B,i)-v(a,i))}}\right)e^{-\varepsilon(B,i)}e^{-e^{-\varepsilon(B,i)}}d\varepsilon(B,i). \tag{13}$$

Substitutes $\varepsilon(B,i)$ by s, we have

$$p(a|i,B) = \int_{-\infty}^{\infty}\left(1 - e^{-e^{-(s+w(B,i)-v(a,i))}}\right)e^{-s}e^{-e^{-s}}ds \tag{14}$$
$$= \int_{-\infty}^{\infty} e^{-s}e^{-e^{-s}}ds - \int_{-\infty}^{\infty} e^{-e^{-s}\left(e^{-(w(B,i)-v(a,i))}+1\right)}e^{-s}ds. \tag{15}$$

Define $x = e^{-s}$, we get

$$p(a|i,B) = \int_0^\infty e^{-x}dx - \int_0^\infty e^{-x\left(e^{-(w(B,i)-v(a,i))}+1\right)}dx \tag{16}$$

$$= \left[-e^{-x}\right]_0^\infty - \left[\frac{e^{-x\left(e^{-(w(B,i)-v(a,i))}+1\right)}}{e^{-(w(B,i)-v(a,i))}+1}\right]_0^\infty \tag{17}$$

$$= \frac{e^{v(a,i)}}{e^{v(a,i)} + e^{w(B,i)}}. \tag{18}$$

Since $e^{v(a,i)}$, and $e^{w(B,i)}$ are strictly positive, we can define $e^{v(a,i)} = V(a,i)$, and $e^{w(B,i)} = W(B,i)$, so that $p(a|i,B) = \frac{V(a,i)}{V(a,i)+W(B,i)}$.[13] □

We name the newly derived categorization function $p(a|i,B) = \frac{e^{v(a,i)}}{e^{v(a,i)}+e^{w(B,i)}}$ the *full-fledged logit categorization model*. Most applied work define the representative utility of alternatives as linear in the attributes of alternatives $v(a,i) = \beta_i x_a$, where x_a is the vector of the observed attributes of alternative a, and β_i is the coefficients vector of agent i. We further extend this and define the representative threshold as $w(B,i) = \alpha_i z_B$, where z_B is a vector of the observed attributes of menu B. We can write the full-fledged logit categorization model with linear-in-parameters as,

$$p(a|i,B) = \frac{e^{\beta_i x_a}}{e^{\beta_i x_a} + e^{\alpha_i z_B}}. \tag{19}$$

By making use of the observed characteristics of alternatives and menus, the full-fledged model can be useful in empirical analysis. The authorization probability has a closed form, so that the traditional maximum-likelihood method can be applied. Assume an observer of categorizations from menu B by DMs indexed by i is obtained for the purpose of estimation. In a standard

[13] Here we write $V(a,i)$, $W(B,i)$ instead of $V(a)$, $W(B)$, as we are expressing the authorization probability of alternative a from menu B by DM i.

logit model, the element used for a maximum-likelihood analysis is[14]

$$\prod_{i}\prod_{a\in B}p\left(a|i,B\right)^{y(a|i,B)}, \tag{20}$$

where $y\left(a|i,B\right)=1$ if DM i chose a from menu B, and $y\left(a|i,B\right)=0$ otherwise. However, this has to be modified, as a decrease in $w\left(B,i\right)$ strictly increases $p\left(a|i,B\right)$, so that the estimation by likelihood function $\prod_{i}\prod_{a\in B}p\left(a|i,B\right)^{y(a|i,B)}$ leads to the lowest possible estimation of the threshold, which is usually the lower bound set in the statistical program. Instead, we maximize the probability of the DMs authorizing the alternatives that were actually observed to be authorized and rejecting the alternatives that were actually observed to be rejected. In a full-fledged logit categorization model, the element used for a maximum-likelihood analysis is

$$\prod_{i}\prod_{a\in B}p\left(a|i,B\right)^{y(a|i,B)}\left(1-p\left(a|i,B\right)\right)^{1-y(a|i,B)}, \tag{21}$$

where $y\left(a|i,B\right)=1$ if a is authorized by DM i from menu B, and $y\left(a|i,B\right)=0$ if rejected.

The log-likelihood function becomes

$$LL=\prod_{i}\prod_{a\in B}y\left(a|i,B\right)\ln p\left(a|i,B\right)+\prod_{i}\prod_{a\in B}\left(1-y\left(a|i,B\right)\right)\ln\left(1-p\left(a|i,B\right)\right). \tag{22}$$

A nice feature of a logit categorization function is that it allows for the exploration of how the thresholds are determined by estimating α (i.e., the coefficients of the characteristics of the menus). We use two simple exam-

[14]See the discussion on the maximum-likelihood analysis in discrete choice models in Train (2008).

ples to illustrate how our model can bring forth new insights and research possibilities.

Example 2. (Choice Overload) A consumer decides whether to include/exclude a jar of jam on the shelf into her consideration set (i.e., the set of products she actually considers). The consumer considers a jar of jam (authorization) if she thinks the jam is preferred to the threshold of the menu, and ignores a jar of jam (rejection) otherwise.[15] The experiment in Iyengar and Lepper (2000) finds that, the consumers were less likely to buy a jar of jam if the sizes of the menus were larger.

Choice overload can happen if the thresholds are dependent on the size of menu. For instance, consumers may use a higher threshold with a larger menu to pre-empt regret or dissatisfaction. In this case, the consumers will be less likely to consider an alternative if there are more options.[16] We can define menu B as the set of jam jars on the shelf, and define $w(B, i) = \alpha_i s_B + c$, where s_B is the cardinality of menu B, and c is a constant. Then, we can estimate α_i to see if the decreased demand in jams was due to higher thresholds, which was caused by more options in the menu.[17]

Example 3. (Availability Bias) A doctor decides if the patients have bacteremia. The doctor diagnoses a patient as having bacteremia (authorization) if the symptoms of the patient are more severe than the threshold; conversely, the patient is diagnosed as not having it (rejection) otherwise. The doctor's decisions are subject to availability bias. In other words, a doctor is more likely to diagnose the patient as having a disease if he can recall more relevant examples, and vice versa. The experiment in Poses and Anthony (1991) finds that doctors were more likely to diagnose a patient as having bacteremia if they could recall more bacteremic patients in the past month.

[15] This formation of consideration set is in line with the compensation rule in Hauser (2014).

[16] For a review on choice overload, see Chernev, Böckenholt, and Goodman (2015).

[17] A higher threshold leads to a lower probability for a jar of jam to be considered, thus decreasing the probability of buying a jar of jam.

This decision process is similar to the 'threshold approach' to clinical decision making by Pauker and Kassirer (1975&1980). Availability bias can be detrimental in diagnoses, as it results in rare diseases being underdiagnosed and common diagnoses being overdiagnosed. Our model can be used to illuminate the nature of the availability bias. We can define menu B as the set of patients the doctor has diagnosed in the previous month, and define $w(B,i) = \alpha_i n_B + c$, where n_B is the number of patients who were diagnosed as positive. The estimation of the coefficient α_i can reveal whether the doctor's threshold changes with the ease of recalling diagnosed patients who had bacteremia.

Notice that the full-fledged logit categorization model is not only an empirical view, but also provides an interpretation of the stochasticity in categorizations. Namely, the DM compares the utility of an alternative to the threshold of the menu, and since both utilities and thresholds are random, the DM categorizes stochastically. With the randomness of a specific form (i.i.d with type-1 extreme distribution), the logit-like representation of the authorization probabilities emerges. The simple characterization axiom in Section 3 is an intuitive property of the logit categorization model, which provides insight into the categorization patterns. This full-fledged model provides a proper explanation of the source of stochasticity in categorizations.

5 Concluding Remarks[18]

5.1 Categorization Precision

The categorization function in our full-fledged model clearly shows that the authorization probability is determined by the representative threshold. We model menu dependency by defining the representative threshold as a function of the menu. A natural companion of our model is that, instead of

[18]I thank the referees for suggesting most of the topics in this section.

the representative threshold, the precision of the comparison is affected by the menu. For instance, the authorization probability can be defined as $p(a,B) = \frac{V(a)^{\sigma(B)}}{V(a)^{\sigma(B)}+W^{\sigma(B)}}$, in which $\sigma(B) \in \mathbb{R}$ captures the precision of the categorizations. This expression is similar to the stochastic choice model in Rehbeck (2021), which uses an extended Luce choice rule where the alternative values are exponentiated by a parameter to accommodate choice overload. This is not a logit categorization model, as it does not satisfy γ-IIA. In this model, the representative threshold W is menu independent, which can be taken as an 'ideal' threshold that is applied over every menu.[19] Nevertheless, the authorization probability of alternative a from menu B still changes with the menu, as the categorization precision is determined by the menu. This version of categorization function can be used to clarify the understanding of decision precision.

5.2 Stochastic Representative Threshold

In the full-fledged logit categorization model, we analyse the categorizations behaviour in which the representative threshold $w(B,i)$ is fixed across all decisions. However, sometimes $w(B,i)$ itself can be stochastic, reflecting random 'tastes' for the thresholds of menus. For example, different doctors can be affected by availability bias to different extents. If the representative thresholds follow random distribution $f(w(B,i))$, the authorization probability of alternative a from menu B becomes

$$p(a|i,B) = \int \frac{e^{v(a,i)}}{e^{v(a,i)}+e^{w(B,i)}} f(w(B,i))\, dw(B,i).$$

The full-fledged model is a special case where the distribution of $w(B,i)$

[19]Effects from menus on the representative threshold and on the categorization precision can both be present. In that case, we can define the authorization probability as $p(a,B) = \frac{V(a)^{\sigma(B)}}{V(a)^{\sigma(B)}+W(B)^{\sigma(B)}}$.

is degenerate at fixed parameter $w^*(B,i)$: $f(w(B,i)) = 1$ for $w(B,i) = w^*(B,i)$; and 0 for $w(B,i) \neq w^*(B,i)$, where $w^*(B,i)$ is the (fixed) intrinsic representative threshold for DM i. With parametric assumptions (e.g., assume $w(B,i) \sim N(\mu,\sigma^2)$), we are able to estimate the corresponding parameters through simulation and identify the randomness in $w(B,i)$. This is structurally similar to the mixed logit choice probability, as we essentially use the methodology from the mixed logit model.

5.3 A Richer Data Set

A different type of data set would be given by a stochastic authorization correspondence $p(C|i,B)$, which is the probability that all the alternatives in set $C \subseteq B$ are authorized by DM i, while all the other alternatives in set $B \setminus C$ are rejected. This type of data incorporate richer information on the correlations between authorizations of different alternatives. While we allow for the thresholds to be menu dependent, the categorizations in our model are done by judging every alternative independently on its own merits. This independence property leads to

$$p(C|i,B) = \prod_{b\in C} p(b|i,B) \prod_{c\in B\setminus C} (1 - p(c|i,B)),$$

for all $C \subseteq B$, for all $B \in \mathcal{D}$. If the equation above holds, then our models can be used. If the above equation does not hold, we should expect correlations between authorizations of alternatives, which makes our models incompatible. A potential solution is imposing parametric structure on the utilities of the subsets $C \subseteq B$ (the utility of the bundle of options).[20]

[20]Hendel (1999) applies similar methodology in his empirical analysis on the demand of computers.

Funding Statement

This work was not supported by specific funding. Royal Economic Society COVID Academic Support fund is gratefully acknowledged.

Acknowledgements

This paper is based on a chapter of my Ph.D. dissertation. I thank the Associate Editor Alan Beggs and two anonymous referees for their useful comments and suggestions, which greatly improved the paper. I am very grateful to Paola Manzini, Pawel Dziewulski for highly valuable and insightful discussions. I would also like to thank Giordano Mion, Matthew Embrey, Tom Potoms, Miguel Costa-Gomes, and seminar participants at the University of Sussex and the University of St Andrews for helpful comments. All remaining errors are my own.

References

[1] Block, H.D., and Marschak, J. (1960) Random Orderings and Stochastic Theories of Responses, *Contributions to Probability and Statistics*, ed. Olkin, I., Gurye, S.G., Hoeffding, W., Madow, W.G., and Mann, H.B. Stanford, CA: Stanford University Press.

[2] Chernev, A., Böckenholt, U., and Goodman, J. (2015) Choice overload: A conceptual review and meta-analysis, *Journal of Consumer Psychology*, 25, 333–58.

[3] Debreu, G. (1960) Review of "Individual Choice Behaviour" by R.D. Luce, *American Economic Review*, 50, 186–88.

[4] Danziger, S., Levav, J., and Avnaim-Pesso, L. (2011) Extraneous factors in judicial decisions, *Proceedings of the National Academy of Sciences of the United States of America*, 108, 6889–92.

[5] Elstein A.S. (1988) Cognitive processes in clinical inference and decision making, *Reasoning, Inference, and Judgment in Clinical Psychology*, eds. Turk, D., and Salovey, P, 17–50. Free Press, New York.

[6] Gulliksen, H. (1953) A generalization of Thurstone's learning function, Psychometrika, 18, 297–307.

[7] Goeree, M.S. (2008) Limited information and advertising in the US personal computer industry, *Econometrica,* 76, 1017–74.

[8] Gul, F., Natenzon, P., and Pesendorfer, W. (2010) Random Choice as Behavioral Optimization, *Econometrica*, 82, 1873–1912.

[9] Gibbons, S., and Marshall, B. (2010) Assessing English: a trial collaborative standardised marking project, *English Teaching: Practice and Critique*, 9, 26–39.

[10] Hendel, I. (1999) Estimating Multiple-Discrete Choice Models: An Application to Computerizzation Returns, *Review of Economic Studies,* 66, 423–46.

[11] Hauser, J. (2014) Consideration-set heuristics, *Journal of Business Research*, 67, 1688–99.

[12] Iyengar, S.S., and Lepper, M.R. (2000) When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79, 995–1006.

[13] Luce, R.D. (1959) Individual Choice Behavior: A Theoretical Analysis. New York: Wiley.

[14] McFadden, D.L. (1974) Conditional logit analysis of qualitative choice behavior. in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York: 105–42.

[15] Manski, C., Sherman, L. (1980) An Empirical Analysis of Household Choice Among Motor Vehicles, *Transportation Research*, 14A, 349–66.

[16] McFadden, D.L. (2000) Economic Choices, *American Economic Review*, 91, 351–78.

[17] Manzini, P., Mariotti, M. and Tyson, C.J. (2013) Two-stage Threshold Representations, *Theoretical Economics*, 8, 875–882.

[18] Manzini, P., Mariotti, M. (2014) Stochastic choice and consideration sets, *Econometrica,* 82, 1153–76.

[19] Manzini, P., Mariotti, M. and Tyson, C.J. (2016) Partial Knowledge Restrictions on the Two-Stage Threshold Model of Choice, *Journal of Mathematical Economics*, 64, 41–47.

[20] Manzini, P., Mariotti, M., and Ulku, L. (2021) Sequential Approval: A Model of "Likes", Paper Downloads and Other Forms of Click Behaviour, *working paper*, The University of Bristol. https://www.dropbox.com/s/bsofnkx6tugls5u/DepthAndLove_241220.pdf?dl=0.

[21] Pauker S.G., and Kassirer J.P. (1975) Therapeutic decision making: a cost-benefit analysis, *The New England Journal of Medicine,* 31, 293, 229–34.

[22] Pauker, S.G., and Kassirer J.P. (1980) The threshold approach to clinical decision making, *The New England Journal of Medicine,* 15, 302,1109–17.

[23] Poses R.M., and Anthony M. (1991) Availability, wishful thinking, and physicians' diagnostic judgments for patients with suspected bacteremia, *Medical Decision Making*, 11, 159–168.

[24] Richter, M.K. (1966) Revealed Preference Theory, *Econometrica*, 34, 635–45.

[25] Rehbeck, J. (2021) Choice Overload via a Menu Dependent Luce Model, *working paper*, The Ohio

State University. https://drive.google.com/file/d/1_-BL3A481OYawBnjMKc34ohO2plkl5ju/view.

[26] Samuelson, P.A. (1938) A Note on the Pure Theory of Consumer's Behaviour, *Economica*, 5, 61–71.

[27] Train, K. (2008) Discrete Choice Methods with Simulation, Cambridge University Press.