A BAYESIAN HIERARCHICAL MODEL OF LEXICOGRAPHIC CHOICE

A Thesis presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

SANGHYUK PARK

Dr. Clintin Davis-Stober, Thesis Supervisor

DECEMBER 2017

The undersigned, appointed by the Associate Vice Chancellor of the Office of Research and Graduate Studies, have examined the thesis entitled:

A BAYESIAN HIERARCHICAL MODEL
OF LEXICOGRAPHIC CHOICE

presented by Sanghyuk Park,
a candidate for the degree of Master of Arts in Psychology, and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Clintin Davis-Stober

Professor Jeffrey Rouder

Professor Christopher Wikle

ACKNOWLEDGMENTS

I would like to express my special thanks to my committee members, Dr. Clintin Davis-Stober, Dr. Jeffrey Rouder, and Dr. Christopher Wikle, for their insightful comments. Their comments made a lot of this work possible and their expertises in their respective fields helped me enlarge my understanding about my field.

I also thank my family for their love, support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

ABSTRACT

I present a lexicographic, threshold-based model of choice used to evaluate decision makers' preferences among risky alternatives. Using a hierarchical Bayesian framework, this model is able to account for observed individual differences by allowing for variable threshold values in attribute features, as well as the order that individuals consider attributes of the choice alternatives. Performance of the model is evaluated via a parameter recovery test using simulated data. I also apply the model to the choice data from a decision-making-under-risk experiment (Davis-Stober, Brown & Cavagnaro, 2015). Bayesian p-values are obtained to check the model fits for every individual, and sensitivity analysis is carried out to measure the degree to which choices of prior distributions affect the results. Finally, I discuss the implications of the Bayesian hierarchical model of lexicographic choice I present in this paper.

Chapter 1

Introduction

Multi-attribute problems require people to integrate multiple pieces of information into their decisions. Such problems require decision makers (DMs) to consider more than one aspect of each alternative. How DMs process this information to form preferences has been a fundamental issue in decision making research. Decision making researchers have proposed various models to address this issue, including rational models (e.g., Weighted additive model; Payne, Bettman, & Johnson, 1993) as well as heuristic models (e.g., Take the best model; Gigerenzer, Hoffrage, & Kleinbolting, 1991). Rational models are considered compensatory or optimal because they utilize all relevant pieces of information to arrive at a decision. Heuristic models, on the other hand, are usually considered non-compensatory, thus sub-optimal or even irrational, because they utilize just a part of the available information to arrive at a decision. For example, when deciding where to move to, you may consider only the quality of the neighborhood. When deciding what university to enter, you may decide solely on the basis of whether or not you received a scholarship. Many studies have shown that heuristic models often describe people's behavior better than rational models (Payne, Bettman, & Johnson, 1988; Payne et al., 1993; Gigerenzer, Todd, & the ABC Research Group, 1999).

Indeed, heuristic models have a number of advantages over rational models in accounting for human decision-making. As Shah and Oppenheimer (2008) argued, all heuristic models can be recast into an effort-reduction framework. While rational models require great mental effort, people only have a limited processing capacity. They argued that all heuristics are intended to help people reduce such mental effort by relying on one or more of the following methods: 1) examining fewer cues, 2) reducing the difficulty associated with retrieving and storing cue values, 3) simplifying the weighting principles for cues, 4) integrating less information, and 5) examining fewer alternatives. Many decision making researchers consider heuristics as an adaptive way for humans to make decisions (Goldstein & Gigerenzer, 2002).

One of the most well-known heuristics is a lexicographic heuristic (Tversky, 1969, 1972). Lexicographic heuristics describe a decision process that resembles the way that a dictionary orders words over letters. That is, a DM processes attributes of alternatives one at a time in order of importance, and stops considering subsequent attributes if the current attribute sufficiently distinguishes the alternatives. This simple, yet strong decision process, is supported by empirical evidence (Drolet & Luce, 2004; Buchanan, 1994; Ford, Schmitt, Schechtman, Hults, & Doherty, 1989), and many researchers in the past have mathematically modeled this process (Yee, Dahan, Hauser, & Orlin, 2007; Kohli & Jedidi, 2007). Most of the models were developed under the framework of maximizing the utility of alternatives, which does not describe actual DMs' decision-making processes. This paper thus aims to present a process model of lexicographic heuristics.

1.1 Lexicographic heuristic

As described above, lexicographic heuristics refer to the decision process that a DM considers only one attribute at a time. If there is a tie, she moves on to the second attribute, and so forth. Imagine a person who is buying a used car. There might be numerous factors that need to be considered, but suppose that only price and mileage of the car matter to this person. A lexicographic rule assumes that she processes only one piece of information at a time in her own order. If the price of the car came to her mind first, the mileage information wouldn't play any role while the price is under her consideration. Only when she finds the prices equal between the alternatives, the mileage starts to come into consideration. This example illustrates two aspects of the lexicographic heuristic as a non-compensatory strategy: 1) While one attribute is under consideration, the other attribute has no way to come into the agent's mind, and 2) if the first attribute of one alternative dominates that of the other alternative, there is no way for other attributes to compensate for it.

Due to the non-compensatory property, lexicographic heuristics can be utilized when DMs want to avoid tradeoffs. Typically, people don't want to tradeoff between complex choices, or too many choices. For example, Perry (1991) argues that the justices of the US Supreme Court select cases using lexicographic rules. In this case, the selection of cases offer too many choices, so they use lexicographic rules to simplify them. Also, Kohli and Jedidi (2007) and Yee and his colleagues (2007) report that a substantial portion of their participants used lexicographic rules to evaluate personal computers and smart phones, which have many attributes to consider. Even when tradeoffs are not difficult to make, there is much evidence that people use lexicographic rules. Tversky, Sattah, and Slovic (1988) argue that people use lexicographic rules on decision tasks where alternatives have two attributes. Slovic (1975) reports that people use lexicographic rules to break ties among equally-valued alternatives. And more evidence can be found in consumer research (see e.g., Colman & Stirk, 1999; Dhar & Nowlis, 1999; Roedder-John, 1999; Gonzalez-Vallejo, Bonazzi, & Shapiro, 1996).

There have been many attempts to mathematically model lexicographic heuristics. For example, Kohli and Jedidi (2007) propose an algorithm for identifying lexicographic rules. They characterize the necessary and sufficient conditions under which a linear model represents lexicographic preferences. They then derive a utility function that corresponds to lexicographic preferences under some constraints. With this algorithm, they were able to confirm that about two thirds of participants use any type of lexicographic rules to choose personal laptops. More recently, Davis-Stober (2012) suggests a lexicographic random preference model (see also, Davis-Stober, Brown, & Cavagnaro, 2015). This model allows DMs to change their preferences so long as they always use lexicographic rules in making decisions. Although the model is general in that it allows for any changes in lexicographic preferences, it is still parsimonious enough to be testable against empirical data. These previous approaches, however, do not need to address the actual processes of how DMs arrive at their decisions. Such as which order a particular DM used in her making decisions, and what threshold values this DM used to evaluate the various alternatives. This information is essential when we describe how DMs actually make decisions.

To this end, I present a lexicographic, threshold-based model. This model is a process model that naturally describes how lexicographic rules work in making decisions. Let's consider the car example above. When describing how the person arrives at her purchase on cars, it is necessary to know the following: 1) whether the person considers price or mileage the most important attribute and 2) the thresholds for a particular attribute that are similar enough to one another that the person would not be able to distinguish between the options on the basis of this attribute. For example, suppose that the person is considering three different cars: Car 1 costs $8,000 and has 60,000 miles, Car 2 costs $10,000 and has 30,000 miles, and Car 3 costs $8,500 and has 45,000 miles (see Figure 1). Assume that the person prioritizes price over mileage, and the cars that differ in price $1,000 or less are not significant enough to affect this person's decision. When price is under consideration, Car 2 is

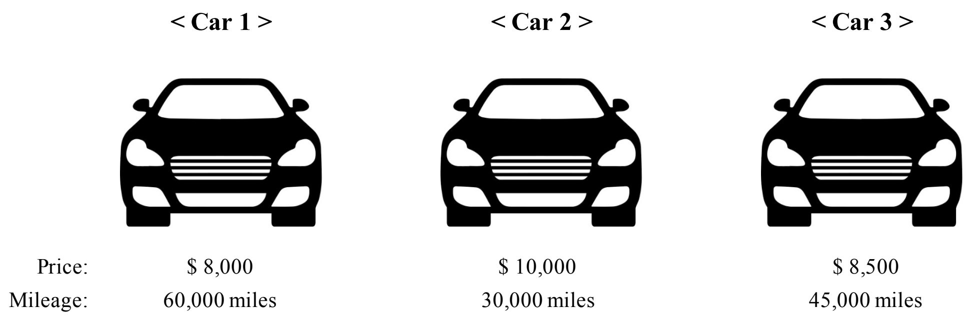

Figure 1.1: Car example

more expensive than Car 1 by more than $1,000. Thus, Car 2 would be eliminated from the consideration. Because Car 1 and Car 3 are indistinguishable in price, the person will now move on to mileage. Assume that cars that differ 10,000 miles or less are similar enough that the person would not distinguish between them. Considering the remaining cars (i.e., Car 1 vs. Car 3), Car 3 has less mileage than Car 1 does by more than 10,000 miles. Therefore, the person would end up choosing Car 3. This method allows the model to account for DMs' decision-making processes.

Hence, the current model emphasizes the following points: 1) The model is able to estimate the order and threshold for every individual DM, which allows us to have a complete profile for each DM with regard to lexicographic rules and thus leads to better understanding of how they arrive at their decisions, 2) By comparing one's profile to one another, we are able to tell individual differences in applying lexicographic rules, and 3) From the profiles for every DM, we are able to predict ones' future choices.

Chapter 2

Model

2.1 Psychological process of the model

Suppose that there are two alternatives to choose from and each has two attributes to consider. Let $\mathbf{A} = (a_1, a_2)$ be alternative A with the two attributes, a_1 and a_2, and let $\mathbf{B} = (b_1, b_2)$ be alternative B with the two attributes, b_1 and b_2. Imagine a decision maker (DM) who considers choosing one among the two alternatives and follows the lexicographic rule. This DM would then compare one attribute at a time in her own order. There are two orders in which she can compare attributes: the first attribute (i.e., a_1 vs. b_1) then the second (i.e., a_2 vs. b_2) or the second attribute then the first. Once the order is decided, she would start comparing the attributes in that order. Let's for now assume that she prioritizes the first attribute (i.e., a_1 for A and b_1 for B) over the second (i.e., a_2 for A and b_2 for B). She would then compare a_1 to b_1 first. She moves on to the next attribute only if she finds alternatives equal with regard to the first attribute. Let τ_1 denote her threshold for the first attribute that distinguishes between alternatives. If $|a_1 - b_1| > \tau_1$, then she prefers the one that favors the first attribute. Otherwise, she would move on to the second attribute and start comparing it. Suppose that the first attribute doesn't help her distinguish alternatives and that now the second attribute comes to her consideration. If τ_2 is her threshold for the second attribute, then she prefers the one that favors the second attribute only if $|a_2 - b_2| > \tau_2$. She would end up being indifferent between alternatives otherwise, because there are only two attributes to consider. The decision process described here explains how the lexicographic rule works in making a preference in binary choice tasks. It can be formally represented as follows:

$$L(\mathbf{A}, \mathbf{B} \mid \boldsymbol{\tau}, O) = \begin{cases} 1, & \text{if A is preferred} \\ 2, & \text{if B is preferred} \\ 3, & \text{if indifferent} \end{cases}$$

where L denotes a lexicographic preference, which refers to a preference made by the lexicographic rule. The lexicographic preference is a function of attributes, $\mathbf{A}$ and $\mathbf{B}$, conditional on the threshold parameter, $\boldsymbol{\tau} = (\tau_1, \tau_2)$, and the order parameter, O, in which a DM compares attributes. The current model assumes that the order parameter is a random variable that has the following probability distribution:

$$O \sim \text{Bernoulli}(\pi),$$

where π corresponds to the probability that a DM chooses the first order (i.e., the first attribute then the second), which implies the probability of choosing the second order (i.e., the second attribute then the first) equals $1 - \pi$.

Even though she came to a preference between alternatives using the lexicographic rule, however, there is still a chance that she chooses different alternative than the one indicated by the lexicographic preference. She might want to change her mind and use a different decision rule that leads to a different choice. Or she simply made a mistake on executing the lexicographic rule correctly. The model thus assumes uncertainty on the behavior level, where she has a chance to choose different alternative. It is formalized into the model as error parameter, ε, that accounts for all kinds of uncertainty that could occur on the behavior level. By including this parameter in the model, the lexicographic rule can be tested against empirical data. Note that the lexicographic preference, L, itself is a deterministic function. Given O and $\boldsymbol{\tau}$, there is no variability in L. The error parameter, ε, is the one that makes the lexicographic preference testable. More specifically, observed choices are assumed to arise from a multinomial distribution with probabilities disturbed by the error parameter, ε:

$$y \mid L \sim \text{Multinomial}(\mathbf{P}),$$

$$\mathbf{P} = \begin{cases} (1-\varepsilon, \frac{\varepsilon}{2}, \frac{\varepsilon}{2}), & \text{if } L = 1 \\ (\frac{\varepsilon}{2}, 1-\varepsilon, \frac{\varepsilon}{2}), & \text{if } L = 2 \\ (\frac{\varepsilon}{2}, \frac{\varepsilon}{2}, 1-\varepsilon), & \text{if } L = 3 \end{cases}$$

where y is an observed choice taking values of either 1, 2, or 3, corresponding to A, B, or indifference between the two alternatives, respectively, conditional on the lexicographic preference, L.

Hence, the goal of the model specification described above is to estimate latent parameters responsible for observed choices: the probability of choosing the first order, π, and threshold parameters, $\boldsymbol{\tau}$. To estimate them, I will employ Bayesian approaches. Bayesian approaches have recently received significant amount of attention because of the computational advances in Markov chain Monte Carlo (MCMC). This approach needs to specify prior distributions to estimate the model, however. In the following section, I first extend the model into more general cases and then suggest prior distributions for the extended model.

2.2 Model extension

In the previous section, the model considered only about a simple case, where one DM makes a decision on one task. Now we extend the model into more general case in which there are multiple DMs and multiple tasks to make a decision on. The number of alternatives under comparison and the number of attributes, however, will still remain same throughout the paper as it is not trivial to generalize them beyond two. Hence, the current model can only deal with ternary choices (i.e., A, B, and indifference between two alternatives) and two attributes.

Suppose that there are n DMs, $i = 1, \ldots, n$, and m tasks, $j = 1, \ldots, m$. Then each DM is supposed to make a series of decisions on m tasks. In the course of making decisions, the order in which the DM considers attributes and the thresholds for each attribute are fixed same. That is, a DM is assumed to apply the same order and thresholds to every task she has to make a decision on. The estimate of the probability of choosing the first order, π, will then reflect how much the order parameter vary by across DMs. If O_i denotes the order for the i^{th} subject, then we have:

$$O_i \sim \text{Bernoulli}(\pi), \quad i = 1, \ldots, n,$$

As explained above, thresholds are also fixed the same across m tasks, and every DM has their own thresholds for each attribute. Let $\boldsymbol{\tau}_i$ be a vector of thresholds for the i^{th} DM, which consists of the i^{th} DM's thresholds for the first attribute, $\tau_{1,i}$, and the second attribute, $\tau_{2,i}$ (i.e., $\boldsymbol{\tau}_i = (\tau_{1,i}, \tau_{2,i})$). Putting the order parameter and thresholds together, we have:

$$L(\mathbf{A}_j, \mathbf{B}_j \mid \boldsymbol{\tau}_i, O_i) = \begin{cases} 1, & \text{if A is preferred} \\ 2, & \text{if B is preferred} \\ 3, & \text{if indifferent} \end{cases}$$

where $\mathbf{A}_j$ and $\mathbf{B}_j$ correspond to a pair of alternatives presented on the j^{th} task.

Once the lexicographic preference is decided, a DM would choose alternative indicated by the lexicographic preference. But the model assumes uncertainty on the behavioral level as described in the previous section. Let ε_i be error that could occur on the behavioral level of the i^{th} subject. If y_{ij} denotes the choice made by the i^{th} subject on the j^{th} task, then we have:

$$y_{ij} \mid L \sim \text{Multinomial}(\mathbf{P}_{ij}),$$
$$\mathbf{P}_{ij} = \begin{cases} (1-\varepsilon_i, \frac{\varepsilon_i}{2}, \frac{\varepsilon_i}{2}), & \text{if } L = 1 \\ (\frac{\varepsilon_i}{2}, 1-\varepsilon_i, \frac{\varepsilon_i}{2}), & \text{if } L = 2 \\ (\frac{\varepsilon_i}{2}, \frac{\varepsilon_i}{2}, 1-\varepsilon_i), & \text{if } L = 3 \end{cases}$$

2.3 Priors

Bayesian approaches need priors to estimate a model. The model above involves free parameters to be estimated: π, and $\tau_{1,i}, \tau_{2,i}$, and ε_i for the i^{th} person, $i = 1, \ldots, n$. So we need to specify prior distributions for each of them. First, for π, a conjugate prior was utilized to make computation efficient:

$$\pi \sim \text{Uniform}(0, 1),$$

For $\tau_{1,i}$ and $\tau_{2,i}$, a hierarchical structure of priors has been employed. What determines a hierarchical structure is whether priors have its own priors. By having a hierarchical structure of priors, the model is able to estimate the parameters for each individual participant that reflects group-level information as well. A hierarchical structure of priors thus accounts for both similarities and differences between individuals (e.g., Lee, 2006; Rouder & Lu, 2005; Nilsson, Rieskamp, & Wagenmakers, 2011). It also makes the estimates more reliable. In this model, $\tau_{1,i}$ and $\tau_{2,i}$ independently arise from separate prior distributions as follows:

$$\tau_{1,i} \sim \text{Normal}^+(\mu_{\tau_1}, \sigma^2_{\tau_1}), \qquad \tau_{2,i} \sim \text{Normal}^+(\mu_{\tau_2}, \sigma^2_{\tau_2}),$$
$$\mu_{\tau_1} \sim \text{Normal}^+(0, 10000), \qquad \mu_{\tau_2} \sim \text{Normal}^+(0, 10000),$$
$$\sigma_{\tau_1} \sim \text{Uniform}(0, 100), \qquad \sigma_{\tau_2} \sim \text{Uniform}(0, 100),$$

where $X \sim \text{Normal}^+(\mu, \sigma^2)$ denotes that X has a truncated Normal distribution with mean of μ and variance of σ^2, for $X \in (0, \infty)$.

Note that the priors for $\pi, \tau_{1,i}$, and $\tau_{2,i}$ are all set to be as diffuse as possible. Given that a posterior distribution can be viewed as compromise between data and prior information, diffusive priors make estimating parameters rely more on data than on priors.

Finally, a prior distribution for the error parameter is assumed to have a uniform distribution, Uniform$(0, 0.5)$. This specification of error represents that a DM is assumed to use a lexicographic rule to make decisions at least a half of times. Since the error parameter indicates the probability that the lexicographic preference doesn't lead to observed choices, it can be interpreted as a rough measure of the goodness-of-fit of the model.

Thus far, we have extended the model into more general cases and specified priors for extend parameters. The extended model is now able to deal with more than one DM and multiple tasks to decide on. Before moving on, I would like to introduce a graphical model (see e.g., Koller, Friedman, Getoor, & Taskar, 2007). Graphical models are especially useful for illustrating how the model assumes the parameters generate behavioral data and how the parameters are related to one another. Figure 2 shows a graphical model for the current model specification. The same notation is used as in Lee (2008). In Figure 2, variables of interest are represented by nodes, and dependencies between the variables are indicated by the graph structure; children

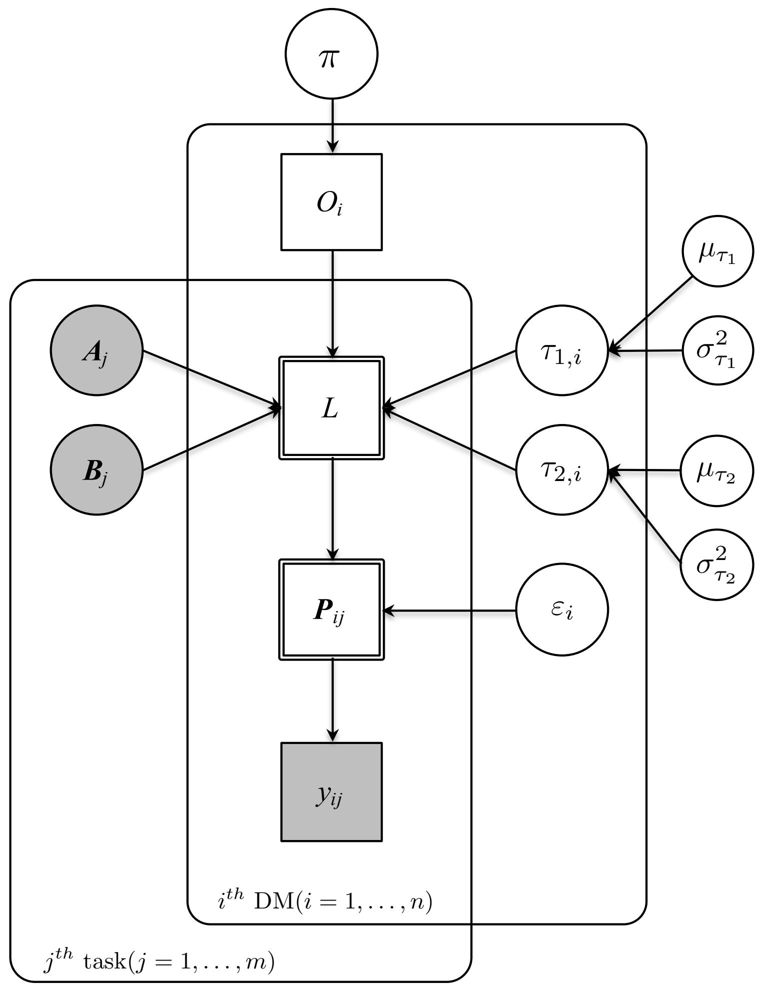

Figure 2.1: Graphical model for lexicographic choices

(ones to which arrows are pointing) depends on their parents (ones from which arrows are stretching out). Circular nodes represent continuous variables, and square nodes represent discrete variables. Observed variables are shown with shading and unobserved variables are shown without shading. Stochastic and deterministic unobserved variables are distinguished by using single and double borders, respectively. Plate notations, enclosing with square boundaries subsets of the graph that have independent replications in the model, is also used.

Chapter 3

Data set (Davis-Stober, Brown, & Cavagnaro, 2015)

Davis-Stober and his colleagues (2015) carried out a set of experiments to tease apart two theories regarding the algebraic structure of preferences. One is the strict weak-order (i.e., utility) representation theory. This theory is formulated as a mixture model in their analysis, weak order mixture model (WOMM), which allows people to switch their preference freely per their decision so long as preference stays under a strict weak order (i.e., a ranking with ties). The other is the lexicographic semiorder representation theory. This theory is also formulated as a mixture model, lexicographic semiorder mixture model (LSMM), whereby people's choice preference is allowed to switch so long as it is consistent with a lexicographic semiorder. By differentiating those two models from one another, the main purpose of the paper was to examine whether subjects' preferences on the given alternatives are consistent with the axiom of transitivity.

The design of the experiment followed a traditional Tversky (1969) gamble paradigm. Subjects were presented with a pair of gambles and asked to indicate which one they prefer to play or that those gambles are indifferent to them. The gambles in the

Set 1

Gamble	A	B	C	D	E
Payoff	\$25.43	\$24.16	\$22.89	\$21.62	\$20.35
Prob.	7/24	8/24	9/24	10/24	11/24

Set 2

Gamble	A	B	C	D	E
Payoff	\$31.99	\$27.03	\$22.89	\$19.32	\$16.19
Prob.	7/24	8/24	9/24	10/24	11/24

Table 3.1: Gamble stimulus sets, Set 1 and 2, from Davis-Stober, Brown and Cavagnaro (2015). Payoff value and probability of winning are given for each gamble.

experiment differed in payoff and probability of winning — the greater the probability of winning the gamble has, the smaller the payoff. There is no case for losing money. People thus either win the corresponding amount of money or win nothing with respective probabilities as a result of playing gambles. The experiment had five different gambles. On each trial, two gambles were randomly chosen out of five and showed up on the screen. Different pairs follow once the subjects make a decision on the current pair. Subjects had kept making decisions on a series of gamble pairs until they responded to all possible pairs of those five gambles 12 times for each, which equals $\binom{5}{2} \times 12 = 120$ trials total.

It is worth noting that the authors manipulated three experimental conditions to see if those manipulations (if any) could affect people's gamble choices. First, they manipulated the amount of payoff of the gambles used in the experiment. They used two different gamble sets, Set 1 and 2, each of which has five different gambles in it (see Table 3.1). Note that gambles in Set 2 has more variable payoff than those in Set 1. The authors also manipulated the display format in which a pair of gambles is presented. Gambles were presented either in a circle format or in a bar format (see Figure 3.1). Finally, they manipulated time within which subjects had to respond to

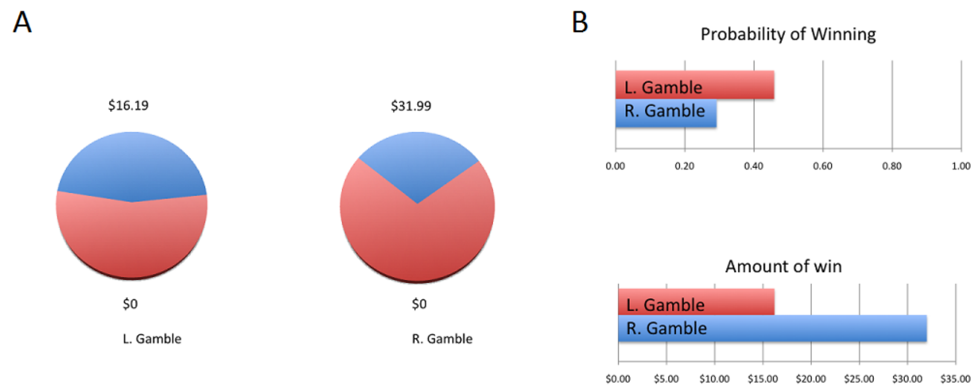

Figure 3.1: Gamble display format: a circle format (A) and a bar format (B)

the given pair — the "Timed" and "Un-timed" conditions. In "Timed" condition, subjects had only 4 seconds to make a decision on each trial. If they didn't respond within that time, the current trial ended with a message indicating that they ran out of time, and the experiment moved on to the next trial. These trials were dropped when encoded in the data set. In "Un-timed" condition, on the other hand, subjects were able to stay on each trial as long as they wanted when making a decision.

The experiment was within-subject design, where every subject had to complete all eight of the above conditions. 60 subjects participated in the experiment. Before they began, they were told that the experimenter will randomly choose one gamble among the ones they select during the experiment and play it for real money as an incentive at the end of the experiment. The authors assumed that this way of giving an incentive would encourage subjects to behave as if they were at real gambles. In the current analysis, only one condition of the data set will be analyzed for illustration. What can be gained by the model is similar regardless of the data set the model is fitting to. I first present a simulation study to demonstrate if the Bayesian estimation procedure is able to recover parameters of interest. I, then, analyze one condition of their data set and discuss what the model can contribute to the existing literature.

Chapter 4

Analyses

4.1 Study 1: Parameter recovery

A simulation study is carried out to confirm that the model is able to estimate parameters of interest accurately. One way to examine it is to see how well the model recovers data-generating parameters. If the data is generated under the lexicographic rule using the same parameters in the model, then the model should be able to recover those parameters. Gamble set 1 (See Table 3.1) was used as stimuli and the number of repetition was 12 for each gamble pair as in Davis-Stober et al. (2015).

The way people compare two gambles identifies the decision strategy they use. People with the lexicographic rule would consider one attribute at a time with their own threshold values. Since there are two attributes to consider (i.e., payoff value and probability of winning), people are identified with a lexicographic rule with either one of two orders and two different threshold values. In the given gamble stimulus set, however, multiple threshold values can lead to the same preference. For example, let's consider about gambles A and B. If a DM has a threshold of \$1.00 for payoff, then she would be able to distinguish the gambles with regard to payoff (i.e., $\$25.43 - \$24.16 >$

\$1.00). But this is also the case if the DM has a threshold of \$0.50. Indeed, one would be able to distinguish A from B with regard to payoff if she has any value of threshold for payoff less than \$1.27. The range of threshold values that results in the same preference is called a equivalence class.

Another issue can arise when determining thresholds together with the order of attributes to be considered. If a DM prioritizes payoff over probability of winning in deciding a gamble and has a threshold of \$0.10 for payoff, then she wouldn't ever consider about probability of winning. This is because the smallest difference in payoff between gambles is greater than her threshold \$0.10. She would thus pick whichever that has better payoff. In this case, her threshold for probability never plays a role in making a decision, thus the model is not able to estimate the threshold for probability.

In Study 1, all possible combinations of an order and threshold values have been considered, which gives exhaustive lexicographic preferences for the given gamble stimulus set (see Table 1). Note that ranges of thresholds in Table 1 specify equivalence classes, which implies any value in the same class lead to the same lexicographic preference. Some thresholds can not be estimated or only lower bound of thresholds can be estimated because of the reason explained above. Data is then simulated according to those preferences by adding various levels of error. Error used for this simulation study ranges from 0.01 to 0.50. To include all the levels of error as equally as possible, every level of error is picked up to be equally spaced from 0.01 to 0.50 and randomly assigned to each combination. The last column of Table 1 corresponds to error. For each combination, responses to all pairs of gambles are generated 12 times per pair. This gives the same number of trials in the study of Davis-Stober et al. (i.e., 120 trials per each DM). Then the Bayesian hierarchical model of lexicographic choices is fitted to the generated data.

DM	O	τ_1	τ_2	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE	ε
1	1	$0 \leq \tau_1 \leq 0.042$	(NA)	B	C	D	E	C	D	E	D	E	E	0.062
2	1	$0.042 < \tau_1 \leq 0.083$	$0 \leq \tau_2 \leq 1.27$	A	C	D	E	B	D	E	C	E	D	0.474
3	1	$0.042 < \tau_1 \leq 0.083$	$1.27 < \tau_2$	I	C	D	E	I	D	E	I	E	I	0.319
4	1	$0.083 < \tau_1 \leq 0.125$	$0 \leq \tau_2 \leq 1.27$	A	A	D	E	B	B	E	C	C	D	0.371
5	1	$0.083 < \tau_1 \leq 0.125$	$1.27 < \tau_2 \leq 2.54$	I	A	D	E	I	B	E	I	C	I	0.216
6	1	$0.083 < \tau_1 \leq 0.125$	$2.54 < \tau_2$	I	I	D	E	I	I	E	I	I	I	0.397
7	1	$0.125 < \tau_1 \leq 0.167$	$0 \leq \tau_2 \leq 1.27$	A	A	A	E	B	B	B	B	B	B	0.500
8	1	$0.125 < \tau_1 \leq 0.167$	$1.27 < \tau_2 \leq 2.54$	I	A	A	E	I	B	B	I	C	I	0.268
9	1	$0.125 < \tau_1 \leq 0.167$	$2.54 < \tau_2 \leq 3.81$	I	I	A	E	I	I	B	I	I	I	0.242
10	1	$0.125 < \tau_1 \leq 0.167$	$3.81 < \tau_2$	I	I	I	E	I	I	I	I	I	I	0.294
11	0	(NA)	$0 \leq \tau_2 \leq 1.27$	A	A	A	A	B	B	B	C	C	D	0.139
12	0	$0 \leq \tau_1 \leq 0.042$	$1.27 < \tau_2 \leq 2.54$	B	A	A	A	C	B	B	D	C	E	0.165
13	0	$0.042 < \tau_1$	$1.27 < \tau_2 \leq 2.54$	I	A	A	A	I	B	B	I	C	I	0.010
14	0	$0 \leq \tau_1 \leq 0.042$	$2.54 < \tau_2 \leq 3.81$	B	C	A	A	C	D	B	D	E	E	0.191
15	0	$0.042 < \tau_1 \leq 0.083$	$2.54 < \tau_2 \leq 3.81$	I	C	A	A	I	D	B	I	E	I	0.113
16	0	$0.083 < \tau_1$	$2.54 < \tau_2 \leq 3.81$	I	I	A	A	I	I	B	I	I	I	0.448
17	0	$0 \leq \tau_1 \leq 0.042$	$3.81 < \tau_2 \leq 5.08$	B	C	D	A	C	D	E	D	E	E	0.345
18	0	$0.042 < \tau_1 \leq 0.083$	$3.81 < \tau_2 \leq 5.08$	I	C	D	A	I	D	E	I	E	I	0.423
19	0	$0.083 < \tau_1 \leq 0.125$	$3.81 < \tau_2 \leq 5.08$	I	I	D	A	I	I	E	I	I	I	0.087
20	0	$0.125 < \tau_1$	$3.81 < \tau_2 \leq 5.08$	I	I	I	A	I	I	I	I	I	I	0.036

Table 4.1: All combinations of parameters and corresponding lexicographic preferences. O denotes the order parameter, where $O = 1$ if a DM prioritizes probability over payoff, $O = 0$ if a DM prioritizes payoff over probability. τ_1 denotes the threshold for probability, and τ_2 denotes the threshold for payoff. ε denotes error.

4.1.1 Identification constraints and model fitting

In fitting the model to the generated data, an issue with model identification has arisen. In other words, there exist multiple ways to get to the same lexicographic preference. Imagine a DM who has thresholds of 0 for probability and \$20.00 for payoff. This DM would then choose whichever gamble that has higher probability of winning no matter what order she has. If she prioritizes probability of winning over payoff, she would choose B between gambles A and B. If she prioritizes payoff over probability of winning, she would choose B between A and B. This happens for all pairs of gambles, thus the model is not able to estimate the order parameter correctly. To fix this problem, one constraint is placed on threshold parameters. That is, the threshold for the attribute should be less than the maximum difference of the gamble set in that attribute if this attribute is to be the first attribute to consider. For example, the DM's threshold for payoff, \$20.00, shouldn't be the first attribute, because it exceeds the maximum difference the given gamble set can make in payoff (i.e., $\$25.43 - \$20.35 = \$5.08 < \$20.00$). If it is the first attribute for her to consider, payoff value wouldn't help her distinguish between gambles throughout entire trials. By imposing such constraint, the model now becomes identifiable.

I estimate the model in JAGS (Plummer, 2013).The model is sampled for 9,000 iterations following an burn-in of 6,000 iterations. Model convergence is assessed via time series plots.

4.1.2 Results

Table 3 summarizes the main results. Median of the posterior distributions is used as a summary statistic. The left side of Table 3 shows true values of parameters used to simulate the data, and the right side shows posterior mean of parameters of interest. Even when the data is generated out of the considerable amount of error, the

	True values of data-generating parameters				Posterior median of parameters			
DM	O	τ_1	τ_2	ε	O	τ_1	τ_2	ε
1	1	$0 \leq \tau_1 \leq 0.042$	(NA)	0.062	1	0.024	(NA)	0.038
2	1	$0.042 < \tau_1 \leq 0.083$	$0 \leq \tau_2 \leq 1.27$	0.474	1	0.066	0.783	0.436
3	1	$0.042 < \tau_1 \leq 0.083$	$1.27 < \tau_2$	0.319	1	0.066	3.072	0.311
4	1	$0.083 < \tau_1 \leq 0.125$	$0 \leq \tau_2 \leq 1.27$	0.371	1	0.108	0.783	0.359
5	1	$0.083 < \tau_1 \leq 0.125$	$1.27 < \tau_2 \leq 2.54$	0.216	1	0.108	1.97	0.203
6	1	$0.083 < \tau_1 \leq 0.125$	$2.54 < \tau_2$	0.397	1	0.108	3.704	0.361
7	1	$0.125 < \tau_1 \leq 0.167$	$0 \leq \tau_2 \leq 1.27$	0.500	1	0.149	0.788	0.467
8	1	$0.125 < \tau_1 \leq 0.167$	$1.27 < \tau_2 \leq 2.54$	0.268	1	0.149	1.962	0.293
9	1	$0.125 < \tau_1 \leq 0.167$	$2.54 < \tau_2 \leq 3.81$	0.242	1	0.149	3.125	0.227
10	1	$0.125 < \tau_1 \leq 0.167$	$3.81 < \tau_2$	0.294	1	0.150	4.562	0.277
11	0	(NA)	$0 \leq \tau_2 \leq 1.27$	0.139	0	(NA)	0.755	0.130
12	0	$0 \leq \tau_1 \leq 0.042$	$1.27 < \tau_2 \leq 2.54$	0.165	0	0.036	1.868	0.184
13	0	$0.042 < \tau_1$	$1.27 < \tau_2 \leq 2.54$	0.010	0	0.155	2.019	0.030
14	0	$0 \leq \tau_1 \leq 0.042$	$2.54 < \tau_2 \leq 3.81$	0.191	0	0.025	3.244	0.223
15	0	$0.042 < \tau_1 \leq 0.083$	$2.54 < \tau_2 \leq 3.81$	0.113	0	0.064	3.289	0.122
16	0	$0.083 < \tau_1$	$2.54 < \tau_2 \leq 3.81$	0.448	0	0.161	3.279	0.416
17	0	$0 \leq \tau_1 \leq 0.042$	$3.81 < \tau_2 \leq 5.08$	0.345	0	0.025	4.334	0.309
18	0	$0.042 < \tau_1 \leq 0.083$	$3.81 < \tau_2 \leq 5.08$	0.423	0	0.064	4.394	0.371
19	0	$0.083 < \tau_1 \leq 0.125$	$3.81 < \tau_2 \leq 5.08$	0.087	0	0.104	4.461	0.192
20	0	$0.125 < \tau_1$	$3.81 < \tau_2 \leq 5.08$	0.036	0	0.187	4.554	0.039

Table 4.2: Results from the simulation study

model is able to recover the data-generating parameters including error. Recall that the ranges of τ_1 and τ_2 in the left side represent equivalence classes. The recovered values of τ_1 and τ_2 are considered okay if they fall within the ranges. Regardless of the error rate, all threshold parameters are perfectly recovered. From this simulation study, I conclude that the model is able to recover the data-generating parameter values accurately. In the next section, the model is fitted to the empirical data from Davis-Stober et al. (2015).

4.2 Study 2: Analysis of Davis-Stober et al. (2015)

In this section, I analyze the empirical data (Davis-Stober et al., 2015) using the same model as in Study 1. The primary goal of the original study was to classify subjects according to the WOMM or LSMM. People who made decisions using a lexicographic rule were classified into LSMM in the original analysis. As the current model is to identify people's lexicographic rule with a search order and thresholds, I chose one condition out of their data set based upon the number of people classified into the LSMM. In the original analysis, the condition "Set 1/Bar/Un-timed" had the greatest number of people classified into LSMM among all of 8 conditions. 32 out of 60 subjects in this condition have Bayes factors greater than 3.14 in favor of the LSMM versus an unconstrained model, and 9 subjects were best fitted by LSMM among all the models they considered. Therefore the current analysis aims at this condition of the data set in the rest of the paper.

As in Study 1, posterior distributions for parameters of interest are computed via the MCMC simulation using JAGS (Plummer, 2013). The current simulation initiates 3 chains with different starting values for each parameter. This way of simulation allows researchers to see if those chains converge to the posterior distribution regardless of different starting values. Each chain draws 5,000 samples from posterior distributions after 5,000 burn-in samples. The convergence of the model is judged through time series plots and autocorrelation function plots.

4.2.1 Results

The current model has 4 free parameters, $\varepsilon, \tau_1, \tau_2$, and O, to be estimated, and those parameters provide sufficient information on one's lexicographic rule. Hence, the analysis from here on focuses on estimating and interpreting those parameters. Median of posterior distributions is used as a summary statistic as in Study 1.

The estimated search order (O) and thresholds for probability of winning (τ_1) and payoff (τ_2) are presented along with error (ε) in Table 4.3 and 4.4 for each individual. It is easy to notice that majority of subjects prioritized probability of winning over payoff (i.e., $O = 1$; $O = 0$ vice versa) when considering which gamble to play. Only 6 subjects considered payoff first. This may be due to the design of the experiment. Davis-Stober and his colleagues (2015) used two different gamble sets as stimuli for the experiment. The stimulus Set 1 originally came from the Tversky (1969) gambles, where the expected value of the gambles increases as the probability of winning increases. Hence, the strategy of choosing the gamble based solely on the probability of winning, in fact, maximizes the expected utility. Plus, at the end of the experiment, participants were allowed to play one of the gambles they chose during the experiment for real money as an incentive. To increase the chance to win money, it would be reasonable to choose the gamble with higher probability of winning. All being considered, it is natural for people to prioritize probability of winning over payoff. One thing I found interesting is, however, about 10% of people still preferred payoff over probability of winning. Given the reasons mentioned here, choosing gambles based primarily on payoff would be an inferior strategy to use in this condition. It might make people consider different decision-making strategies if payoff of the gambles matters to them. Which would be why the model tends to induce more error when accounting for the data of those who prioritize payoff over probability of winning. They may have used different strategies or constantly made mistakes on their choices.

It is important to note that some thresholds may not be identifiable as pointed out in Study 1. A search order combined with a low threshold for the first attribute would keep one from considering the second attribute. This person would then always choose whatever that favors the first attribute, and the threshold for the second attribute for this person remains unidentifiable. It is not possible for the model to estimate it that never affects one's behaviors. The same logic applies to the threshold that goes beyond the range of the attributes. Once the threshold gets greater than the range of the attributes of the stimuli, any value of the threshold has the same effects on his or her choices. No matter how big the threshold is, every possible pair of gambles wouldn't be distinguishable under this threshold. For these reasons, Table 4.3 and 4.4 include numerous "(NA)"s (i.e., unidentifiable) and some arbitrary large values

Subject	Search Order (O)	Threshold for probability of winning (τ_1)	Threshold for payoff (τ_2)	Error (ε)
1	1	.024	(NA)	.031
2	1	.025	(NA)	.079
3	1	.024	(NA)	.006
4	1	.089	.863	.330
5	1	.025	(NA)	.063
6	1	.025	(NA)	.039
7	1	.025	(NA)	.137
8	1	.025	(NA)	.055
9	1	.054	.862	.129
10	1	.025	(NA)	.039
11	1	.025	(NA)	.014
12	1	.025	(NA)	.129
13	1	.025	(NA)	.204
14	0	(NA)	.924	.088
15	1	.025	(NA)	.137
16	1	.025	(NA)	.137
17	0	(NA)	.920	.436
18	1	.024	(NA)	.089
19	1	.025	(NA)	.039
20	1	.053	2.081	.335
21	0	107.770	2.829	.374
22	1	.025	(NA)	.072
23	1	.025	(NA)	.030
24	1	.025	(NA)	.038
25	1	.025	(NA)	.030
26	1	.025	(NA)	.022
27	0	(NA)	.920	.327
28	1	.025	(NA)	.377
29	1	.025	(NA)	.006
30	1	.025	(NA)	.014

Table 4.3: Search order (O) and thresholds for probability of winning (τ_1) and payoff (τ_2) for the first 30 subjects

Subject	Search Order (O)	Threshold for probability of winning (τ_1)	Threshold for payoff (τ_2)	Error (ε)
31	1	.025	(NA)	.121
32	1	.053	.864	.030
33	1	.089	2.919	.194
34	1	.025	(NA)	.006
35	1	.025	(NA)	.014
36	1	.025	(NA)	.064
37	1	.025	(NA)	.038
38	1	.025	(NA)	.047
39	0	(NA)	.923	.088
40	1	.025	(NA)	.038
41	1	.053	.857	.196
42	1	.025	(NA)	.030
43	1	.025	(NA)	.096
44	1	.025	(NA)	.228
45	1	.025	(NA)	.138
46	1	.025	(NA)	.261
47	1	.089	2.927	.277
48	0	112.127	1.787	.294
49	1	.025	(NA)	.343
50	1	.025	(NA)	.269
51	1	.025	(NA)	.054
52	1	.054	.862	.336
53	1	.025	(NA)	.063
54	1	.025	(NA)	.294
55	1	.053	.865	.252
56	1	.025	(NA)	.055
57	1	.025	(NA)	.079
58	1	.025	(NA)	.310
59	1	.054	.846	.137
60	1	.025	(NA)	.145

Table 4.4: Search order (O) and thresholds for probability of winning (τ_1) and payoff (τ_2) for the rest 30 subjects

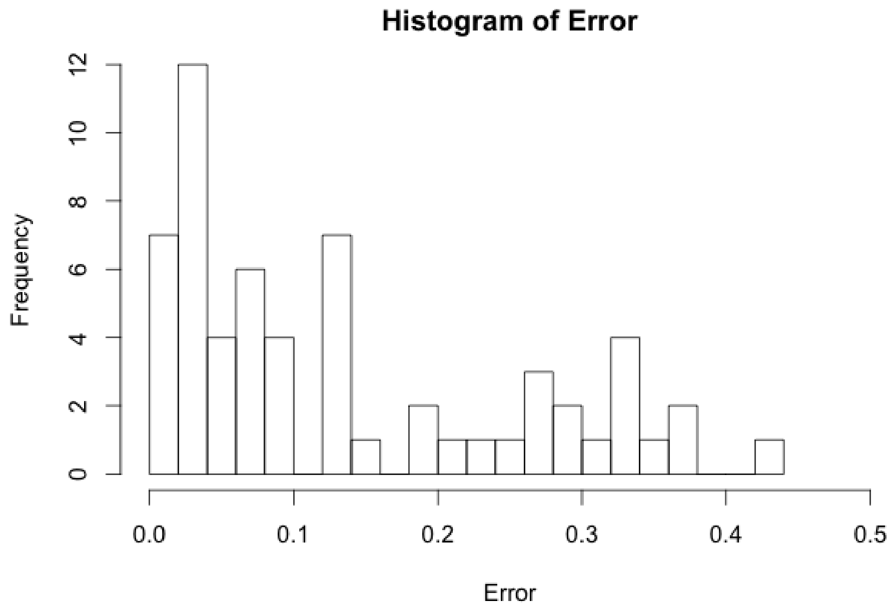

Figure 4.1: Histogram of Error.

for thresholds.

Under the design of the current experiment, estimated thresholds can be thought of as discrete values rather than continuous values. This is mainly because the experiment has only 5 different gambles, which yields only 10 unique pairs. In other words, there are only 10 different values for thresholds to choose from. Also, the increment and decrement in probability of winning and payoff are about equal between adjacent gambles, which makes variability of thresholds much smaller. Under this setting, the total number of different values thresholds for each attribute can have is 5. This is why the same estimates for thresholds are repeatedly shown for different subjects in Table 4.3 and 4.4. The results show that most thresholds fall within the smallest category. That is, people tend to choose whichever wins the attribute they prioritize most.

Finally, the estimated error (ε) is summarized in Figure 4.1 and Table 4.5. By inspecting both the histogram and the cumulative frequency table for error, one can learn about how many people utilized a lexicographic rule in this condition as well as how often they used a lexicographic rule. In general, error (ε) in this condition is quite

Error (ε)	Proportion
$\varepsilon < .05$	33.33%
$\varepsilon < .10$	55.00%
$\varepsilon < .20$	71.67%
$\varepsilon < .30$	85.00%
$\varepsilon < .40$	98.33%
$\varepsilon < .50$	100.00%

Table 4.5: Cumulative frequency table for Error.

low. One-third of subjects made choices consistent with a lexicographic rule with less than the error rate of .05. Over half of people followed a lexicographic rule with less than .10 error. In other words, more than half of subjects in this condition used a lexicographic rule 9 out of 10 times to choose gambles to play. They are generally willing to use the lexicographic rule to make their decisions in this condition.

4.2.2 Comparison with the original results

The main goal of the original study was to test WOMM and LSMM against empirical data to differentiate one from another. Recall that LSMM is a mixture model of all possible lexicographic semiorders on the given alternatives. People under a LSMM are supposed to make choices consistent with lexicographic semiorders at every trial. The same logic applies to the current model of this paper as well. Lexicographic semiorders are the binary relations that account for lexicographic decision-making processes. An important property of lexicographic semiorders is that decision makers are assumed to have a threshold, above which his or her preferences change. This threshold is, in fact, one of parameters the current model is trying to estimate. A major claim of the current model is that together with one's own search order, estimated thresholds for each attribute of the alternatives give a sufficient information to identify his or her lexicographic rule. Hence, it is important to compare the current results with the original results. Subjects well fitted by the current model should be classified according to LSMM in the original results. This will serve as a converging evidence, or external validity, of the current model.

Before comparing the current results with the original ones, it should be noted that some types of lexicographic preferences can also be classified into WOMM. This is due to the model mimicry, where one model generates a set of data that is also well accounted for by the other model. For example, if one has very low threshold for the attribute to be considered first, this person would always choose whichever favors that attribute. This way of making decisions on gamble pairs allows him or her to rank his or her preferences over the five gambles in the order of that attribute, thus satisfying transitivity of preferences. This type of data is, of course, well supported by WOMM, and this makes the data being classified according to WOMM than LSMM. For this reason, in this section, I will consider only the subjects whose thresholds for probability of winning, or payoff, or both are greater than the minimum differences (i.e., .042 for probability of winning, \$1.27 for payoff).

Table 4.6 shows the subjects whose thresholds are greater than the minimum differences along with the original results. In the original paper, the authors employed the Klugkist and Hoijtink (2007) Bayes factor methodology to classify subjects. Importantly, the authors computed the Bayes factors which compared the model of interest to the unconstrained model. Thus, the Bayes factors for each model indicate how much more likely the observed data is generated from the model of interest than from the unconstrained model. The original analysis divided LSMM further into LSMM1 and LSMM2. LSMM1 represents a mixture model of all lexicographic semiorders on probability of winning; LSMM2 on payoff.

The authors followed the standard Jeffreys (1998) interpretation scale to classify the subjects: a Bayes factor between 1 and 3 is weak evidence for the model, a Bayes factor between 3 and 10 is strong evidence, and a Bayes factor greater than 100 is

Sub.	Search order (O)	Threshold for Prob. (τ_1)	Threshold for Payoff (τ_2)	Bayes factors WOMM	LSMM1	LSMM2
4	1	.089	.863	390.293	16.667	$< .001$
9	1	.054	.862	.310	30.702	$< .001$
20	1	.053	2.081	14.448	16578.95	$< .001$
21	0	107.770	2.829	1.397	30.702	4222.807
32	1	.053	.864	$< .001$	19.298	$< .001$
33	1	.089	2.919	$< .001$	76400.88	2824.561
41	1	.053	.857	124.810	43.860	$< .001$
47	1	.089	2.927	.034	19327.19	14.035
48	0	112.127	1.787	$< .001$	236.842	7339.474
52	1	.054	.862	859.259	.877	$< .001$
55	1	.053	.865	1723.828	34.211	$< .001$
59	1	.054	.846	21.431	24.561	$< .001$

Table 4.6: The Bayes factors for WOMM, LSMM1, and LSMM2. Only those whose thresholds are greater than the minimum differences are included.

decisive evidence. As shown in Table 4.6, all subjects, except Subject 52, show strong or decisive evidence for LSMM1 or LSMM2 against the unconstrained model. More than half of the subjects is fitted best by LSMM1 or 2 among all three models. The Bayes factors in the original results agree with the estimated search order as well. The subjects who appeared to consider payoff first in the current analysis also gave the highest Bayes factor to LSMM2 in the original analysis. Overall, the results from both analyses mostly agreed even though the two models take the whole different forms of analysis. The comparison in this section provides a converging evidence for the current model.

4.2.3 Model checking

The last step of a Bayesian analysis is assessing adequacy of the model fit to the data. A Bayesian model assumes a probability model over a entire prior-to-posterior structure. If the model is poor, Bayesian inferences could be misleading. Therefore it is crucial to include some check of how well the model fit to the data. The current analysis presents two different ways of checking the model: 1) Posterior predictive checking and 2) Sensitivity analysis.

Posterior predictive checking

The posterior predictive checking is suggested by (Gelman, Carlin, Stern, & Rubin, 2014) to check the joint posterior predictive distribution $p(y^{rep}|y)$, where y^{rep} represents the data that could have been observed under the same setting as the current data set. If the model fits to the data well, posterior predictions of the model should be consistent with the observed data. In other words, it assesses how plausible the observed data is generated from the model. The joint posterior predictive distribution $p(y^{rep}|y)$ can be obtained by:

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)d\theta,$$

where θ represents a collection of all parameters in the model. Once y^{rep} is generated from the posterior predictive distribution $p(y^{rep}|y)$, one needs to decide how to measure discrepancy between model and data. Following Gelman et al.'s (2014) suggestion, I choose a test statistic denoted by $T(y)$ and use this statistic to summarize discrepancies between model and data. Note that it is a researcher responsible for deciding what kind of a test statistic will be used. By defining the test statistic as the researcher wants, he or she could investigate aspects of the data he or she wishes to check. In this paper, I choose three test statistics, $T_1(y), T_2(y)$, and $T_3(y)$, which represent the number of choosing the gamble with better payoff over the other, the number of choosing the gamble with higher probability of winning over the other, and the number of choosing indifference between them, respectively. The reason I chose them is that the main purpose of the paper is to identify one's lexicographic rule with the search order and thresholds for the attributes of the stimuli and to predict one's future behavior based upon the identified lexicographic rule. Thus, the observed data should at least be replicated under the model that assumes a lexicographic way of decision-making.

To this end, I employed the tail-area probability of the test statistic to measure lack of fit of the data with respect to the posterior predictive distribution (Gelman et al., 2014). This is called the Bayesian p-value, p_B, often considered as the Bayesian counterpart of the classical p-value:

$$p_B = \Pr(T(y^{rep}) \geq T(y)|y).$$

As indicated by the above equation, the Bayesian p-value is the probability that the test statistic of the replicated data is greater than that of the observed data (the sign of inequality changes based upon the nature of the test statistic). Small values of it imply that the model hardly generates the observed data, as measured by the test statistics. In many practical applications, however, it is nearly impossible to compute the above equation analytically. Hence, an estimate of the Bayesian p-value is computed using posterior simulation of (θ, y^{rep}). The algorithm for posterior simulation Rubin (1984) suggested is as follows:

1. Generate N draws of the parameters $\theta^1, \theta^2, \ldots, \theta^N$ from the posterior distribution $p(\theta|y)$

2. Draws N replicated data sets $y^{rep,n}$ from the likelihood distribution $p(y^{rep,n}|\theta^n)$, $n = 1, \ldots, N$

3. Compute an estimate of the Bayesian p-value from the simulated data sets, $p_B \approx \frac{1}{N} \sum_{n=1}^{N} I_{T(y^{rep,n}) \geq T(y)}$,

where I represents an indicator function.

In this paper, I generated 10,000 (i.e., $N = 10,000$) data sets for each subject and

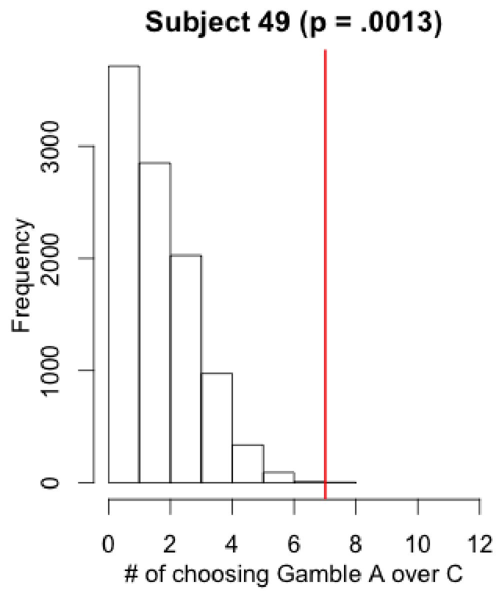

Figure 4.2: Histograms of 10,000 simulated data sets for Subjects 49 (left) and 10 (right) with red vertical lines representing their observed values.

calculated the Bayesian p-values for all of the three test statistics using the algorithm above. Since those three test statistics are dependent on each other (they must sum up to 12), I generated 10,000 data sets for each test statistics, which makes 30,000 simulated data sets total. The Bayesian p-values for each test statistic were then computed within the separated 10,000 data sets in trying to make them independent of each other. Figure 4.2 illustrates how the Bayesian p-value works. The left side of Figure 4.2 is the histogram of simulated data for Subject 49 with a red vertical line representing his or her observed value. This subject chose Gamble A over C 7 out of 12 times during the experiment, but most of 10,000 simulated data generated 0 or 1 out of 12 times on Gamble A for this subject. Given the simulation result, the model with the parameters for this person is highly unlikely to generate the observed data. Only 13 out of 10,000 simulated data did so, which yields the Bayesian p-value of .0013. The histogram in the right side of Figure 4.2, on the other hand, shows an adequate fit of the model to Subject 10's observed data. Subject 10 chose Gamble C over B 11 out of 12 times during the experiment. Most of the simulated data

Gamble pair	# of subjects with $p_{B1} < .05$	# of subjects with $p_{B2} < .05$	# of subjects with $p_{B3} < .05$
A vs. B	7	7	3
A vs. C	5	3	1
A vs. D	2	2	1
A vs. E	3	2	1
B vs. C	5	6	1
B vs. D	3	2	0
B vs. E	1	1	1
C vs. D	14	10	1
C vs. E	4	4	1
D vs. E	9	7	4

Table 4.7: The number of subjects whose Bayes p-value is less than .05 for all gamble pairs. The Bayes p-values were computed for all the three test statistics, $T_1(y), T_2(y)$, and $T_3(y)$

generated 11 or 12 times of choosing Gamble C for this person, which yields the Bayesian p-value of .3847.

Main results of the Bayesian p-values are summarized in Table 4.7. The Bayesian p-values for the test statistics, $T_1(y), T_2(y)$, and $T_3(y)$, are denoted by p_{B1}, p_{B2}, and p_{B3}, respectively. Cutoff point for deciding whether the model fits well to the data is set by .05 following the tradition of the classical p-value. Numbers in Table 4.7 represent the number of subjects whose Bayesian p-value is less than .05 for each gamble pair. Overall, the model provides an adequate fit to the given data. Total number of misfits is 111 out of 1,800. In other words, only 6.2% of the data set is not accounted for by the model. When investigating the data more closely, however, I found some regularities in such misfits. First, the model appears better at accounting for indifference. Numbers in the third column of Table 4.7 are remarkably smaller than those in the other two columns. This might be because there were only a few subjects who choose indifference between two gambles. Predicting the alternative never chosen by subjects would be straightforward for the model as the model simply allocates it zero probability mass. Secondly, there are some gamble pairs that tend to

1st set of prior distributions	2nd set of prior distributions
$\pi = \Phi(z)$	$\pi \sim \text{Exp}_{[0,1]}(10)$
$z \sim \text{Normal}(0, 1)$	
$\tau_{1,i} \sim \text{Normal}^+(\mu_{\tau_1}, \sigma^2_{\tau_1})$	$\tau_{1,i} \sim \text{Normal}^+(\mu_{\tau_1}, \sigma^2_{\tau_1})$
$\mu_{\tau_1} \sim \text{Normal}^+(0, 1)$	$\mu_{\tau_1} \sim \text{Exp}(10)$
$\sigma^2_{\tau_1} \sim \text{Uniform}(0, 10000)$	$\sigma^2_{\tau_1} \sim \text{Exp}(1)$
$\tau_{2,i} \sim \text{Normal}^+(\mu_{\tau_2}, \sigma^2_{\tau_2})$	$\tau_{2,i} \sim \text{Normal}^+(\mu_{\tau_2}, \sigma^2_{\tau_2})$
$\mu_{\tau_2} \sim \text{Normal}^+(0, 1)$	$\mu_{\tau_2} \sim \text{Exp}(1)$
$\sigma^2_{\tau_2} \sim \text{Uniform}(0, 10000)$	$\sigma^2_{\tau_2} \sim \text{Exp}(1)$
$\varepsilon_i \sim \text{Normal}_{[0,.5]}(0, 1)$	$\varepsilon_i \sim \text{Exp}_{[0,.5]}(10)$

Table 4.8: Different sets of prior distributions used to assess the sensitivity of the results

induce more misfits. The gamble pair of C and D, for example, induced the greatest number of misfits among all gamble pairs. In general, adjacent gamble pairs (i.e., A vs. B, B vs. C, C vs. D, and D vs. E) have more misfits than other gamble pairs that are not adjacent. Similar attributes may have caused subjects to switch their choices over time even for the same gamble pair. Finally, the model shows roughly the same performance on the test statistics $T_1(y)$ and $T_2(y)$. This indicates that no matter whether subjects choose gambles based upon probability of winning or upon payoff, the model fits to both types of people about equally.

Sensitivity analysis

Sensitivity analysis can be used to examine the sensitivity of the results to the non-informative prior distributions. Typically, researchers choose different prior distributions other than the ones used to estimate the model and re-estimate the model with the different prior distributions to see if the results significantly change. Different sets of prior distributions I choose for the sensitivity analysis are in Table 4.8.

Parameter estimates under different prior distributions mostly agreed with the original results. Some subjects, however, show a slight disagreement in the parameters for search order and thresholds. Search orders of Subjects 33 and 47 are both estimated as being probability of winning first under the old and first priors in Table 4.8, while the second priors in Table 4.8 estimated them as being payoff first. Also, the second priors in Table 4.8 make the thresholds of large numbers (i.e., Subject 21 and 48's thresholds for probability of winning) shrink greatly toward less than 1. This is not surprising, because under such exponential distributions, large numbers are very unlikely to be generated. This might cause the slight disagreement in the estimates. Other than that, all results nearly agree with the original results. The agreement in parameter estimates indicates that the data dominate the priors.

Chapter 5

Discussion

The main purpose of the present paper is to present a hierarchical Bayesian model to account for the choice data from a perspective of a lexicographic decision-making process. The present paper has also shown how the models works on the empirical data and provides various evidence to ascertain the model's ability to estimate one's threshold and search order. First, the simulation study demonstrates that the model is able to recover all data-generating parameters even when the data is generated out of the high rate of error. Importantly, the model includes the parameter that estimates the error itself. This gives useful information on how well the observed data is consistent with a lexicographic rule, so researcher would know when to apply a lexicographic rule to one's choices. In the analysis of the empirical data (Davis-Stober et al., 2015), the model was fitted to one of their experimental conditions, "Set 1/Bar/Un-timed." The main results I found is that a majority of subjects relied on probability of winning to decide which gamble they prefer to play. Plus, most of them appeared to have thresholds for that attribute less than the minimum difference that the current gamble set possibly makes. In other words, most subjects chose whichever has higher probability of winning while completely ignoring payoff information of gambles. These results immediately enable researchers to predict these subjects' future choices — they are highly likely to choose the gamble which has higher probability of winning.

To validate the current results, I compared it with the original results. The way the current model describes how people make decisions can be thought of as a simple lexicographic semiorder (Davis-Stober, 2012). So, subjects shown to use a lexicographic rule to make decisions in the current analysis should be classified according to a LSMM in the original analysis. After accounting for the model mimicry (i.e., a certain type of lexicographic choices is consistent with WOMM as well), the current and original results both nearly agreed. Subject 52, however, showed an inconsistency between the two models. The current model says that this person is supposed to make decisions in a lexicographic way, but the original result says that this person isn't classified into LSMM. I would attribute it to the high rate of error, .336, this person showed in executing a lexicographic rule. This inconsistent result illustrates that one needs to apply the current results carefully given the error rate. For everyone else, both models showed converging evidence, which gives the current model an external validity.

Finally, I assessed the fit of the model to the data, or model checking. It is critical to check how well the model fits to the data because a Bayesian model typically assumes uncertainty over a whole prior-to-posterior structure. One mis-specification of the model could lead to poor performance. The present paper included two different types of model checking — Posterior predictive checking and sensitivity analysis. For the posterior predictive checking, I employed the Bayesian p-value to measure the discrepancy between the observed data and the model. One benefit from computing the Bayesian p-value is that it allows researchers to check various aspects of the model, which sheds light on the part the researcher needs to improve on. The current analysis shows that misfits of the model did not spread over gamble pairs randomly, but it instead seemed to have a systemic structure behind it. As pointed out in the previous section, adjacent gamble pairs tend to induce more misfits than pairs that are not adjacent. This may be because similar attributes had made people switch their choices back and forth across trials. The model is not good at accounting for this type of behaviors, since the model assumes lexicographic choices as deterministic. All uncertainties that arise at the response level are due to error according to the model. Thus, in future research, one needs to come up with another way to implement uncertainty into responses. Sensitivity analysis showed no significant changes in results when the model is estimated using different sets of priors. It implies that data dominates prior information. No matter what prior distributions are employed, the results will stay mostly the same.

In sum, a lexicographic rule can help people ease their burden of having to consider all relevant information when making decisions. Under a lexicographic rule, people simply choose one attribute that matters to them the most and then make decisions based upon it. Many studies have examined the lexicographic way of making decisions (Kohli & Jedidi, 2007; Yee et al., 2007). While they found a considerable portion of people actually use the lexicographic rule in a certain setting (Colman & Stirk, 1999; Perry, 1991; Tversky et al., 1988; Slovic, 1975), a complete picture of the lexicographic rule they employed has rarely been examined. In this regard, the current model has incorporated the threshold and search order parameters into the model to identify one's lexicographic rule more completely. Once those parameters are successfully estimated, one can have a sufficient information about people's lexicographic rules. Furthermore, this allows him or her to predict people's future decisions on similar problems.

References

Buchanan, J. M. (1994). Health care for an again population. In C. Hackler (Ed.), (pp. 57–68). Albany, NY: State University of New York Press.

Colman, A. M., & Stirk, J. A. (1999). Singleton bias and lexicographic preferences among equally valued alternatives. *Journal of Economic Behavior & Organization*, *40*, 337–351.

Davis-Stober, C. P. (2012). A lexicographic semiorder polytope and probabilistic representations of choice. *Journal of Mathematical Psychology*, *56*, 86–94.

Davis-Stober, C. P., Brown, N., & Cavagnaro, D. R. (2015). Individual differences in the algebraic structure of preference. *Journal of Mathematical Psychology*.

Dhar, R., & Nowlis, S. M. (1999). The effect of time pressure on consumer choice deferral. *Journal of Consumer Research*, *25*, 369–384.

Drolet, A., & Luce, M. F. (2004). Cognitive load and trade-off avoidance. *Journal of Consumer Research*, *31*, 63–77.

Ford, J. K., Schmitt, N., Schechtman, S. L., Hults, B. M., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, *43*, 75–117.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Chapman & Hall/CRC Boca Raton, FL, USA.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* Oxfort, England: Oxford University Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90.

Gonzalez-Vallejo, C., Bonazzi, A., & Shapiro, J. (1996). Effects of vague probabilities and of vague payoffs on preference: A model comparison analysis. *Journal of Mathematical Psychology*, *40*, 130–140.

Jeffreys, H. (1998). *The theory of probability.* OUP Oxford.

Klugkist, I., & Hoijtink, H. (2007). The bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis*, *51*, 6367–6379.

Kohli, R., & Jedidi, K. (2007). Representation and inference of lexicographic preference structures and their variants. *Marketing Science*, *26*, 380–399.

Koller, D., Friedman, N., Getoor, L., & Taskar, B. (2007). *Graphical models in a nutshell* (L. Getoor & B. Taskar, Eds.). Cambridge, MA: MIT Press.

Lee, M. D. (2006). A hierarchical bayesian model of human decision-making on an optimal stopping problem. *Cognitive Science*, *30*, 1–26.

Lee, M. D. (2008). Three case studies in the bayesian analysis of cognitive models. *Psychonomic Bulletin and Review*, *15*, 1–15.

Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, *55*, 84–93.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

Perry, H. W. (1991). *Deciding to decide: Agenda setting in the united states supreme* *court*. Cambridge, MA: Harvard University Press.

Plummer, M. (2013). rjags: Bayesian graphical models using mcmc. *R package version*, *3*.

Roedder-John, D. (1999). Consumer socialization of children: A retrospective look at twenty-five years of research. *Journal of Consumer Research*, *26*, 183–213.

Rouder, J. N., & Lu, J. (2005). An introduction to bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin and Review*, *12*, 573–604.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, *12*, 1151–1172.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*, 207–222.

Slovic, P. (1975). Choice between equally valued alternatives. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 280–287.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*, 31–48.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299.

Tversky, A., Sattah, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, *95*, 371–384.

Yee, M., Dahan, E., Hauser, J. R., & Orlin, J. (2007). Greedoid-based noncompensatory inference. *Marketing Science*, *26*, 532–549.