Explaining Human Decision Making in Optimal Stopping Tasks

Christiane Baumann (c.baumann@psychologie.uzh.ch)
Department of Psychology, University of Zürich, Binzmühlestrasse 14/19
8050 Zurich, Switzerland

Henrik Singmann (singmann@psychologie.uzh.ch)
Department of Psychology, University of Zürich, Binzmühlestrasse 14/19
8050 Zurich, Switzerland

Vassilios E. Kaxiras (vassilioskaxiras@gmail.com)
Belmont High School, 221 Concord Ave
Belmont, MA 02478

Samuel J. Gershman (gershman@fas.harvard.edu)
Department of Psychology, Harvard University, 52 Oxford Street,
Cambridge, MA 02138

Bettina von Helversen (b.vonhelversen@psychologie.uzh.ch)
Department of Psychology, University of Zürich, Binzmühlestrasse 14/19
8050 Zurich, Switzerland

Abstract

In an optimal stopping problem, people encounter a sequence of options and are tasked with choosing the best one; once an option is rejected, it is no longer available. Recent studies of optimal stopping suggest that people compare the current option with an internal threshold and accept it when the option exceeds the threshold. In contrast, we propose that humans decide to accept or reject an option based on an estimate of the probability that a better option will be observed in the future. We develop a computational model that formalizes this idea, and compare the model to the optimal policy in two experiments. Our model provides a better account of the data than the optimal model. In particular, our model explains how the distributional structure of option values affects stopping behavior, providing a step towards a more complete psychological theory of optimal stopping.

Keywords: optimal stopping; cognitive modeling; sequential decision making; probabilistic choice behavior

Introduction

Choosing the best option from a sequence is a common problem in everyday life: whenever we search for a job, an apartment or a partner, we may decide to accept or reject the present option without knowing whether future options will be more attractive. In economics, this class of decision problems is referred to as “optimal stopping” or “secretary problems” (Gilbert & Mosteller, 1966; Seale & Rapoport, 1997). Decisions in such problems involve a trade-off between accepting a subprime option prematurely and the danger of rejecting the best offer out of false hopes for better options in the future. Importantly, the mathematically optimal solutions for different versions of the secretary problems are generally known (Gilbert & Mosteller, 1966).

A number of studies have examined two versions of this problem: rank order and full information. On the rank-order version of the problem (e.g., Seale & Rapoport, 1997; Bearden, Rapoport, & Murphy, 2006) only the rank of the option relative to those already seen is shown. In the full-information version of the task, the actual value of the option is presented (e.g., Lee, 2006; Guan & Lee, in press). In this manuscript we focus on the full information version of the optimal stopping task. As an illustration, imagine a person who wants to find the cheapest airplane ticket, and offers vary in price from day to day. The person is checking the actual price every day and has to decide to accept or reject the ticket without having the option to go back to a previously rejected offer. Moreover, the search is limited in time because the ticket has to be bought before the beginning of the trip. For this version, the optimal solution is based on calculating the *expected reward* of the remaining outcomes. From this expected reward, a threshold is calculated for each option in the sequence. This threshold is monotonically increasing when finding the minimum (as in our example) or monotonically decreasing when finding the maximum. If the value of the current option goes below (for mimimum) or exceeds (for maximum) the threshold, the option should be chosen.

Previous studies (Lee, 2006; Guan, Lee, & Vandekerckhove, 2015; Guan & Lee, in press; von Helversen & Mata, 2012) found evidence that people use thresholds that change over the position in an optimal stopping problem to make decisions. Based on this finding, Guan and Lee (in press) constructed a descriptive model in which the thresholds on each position were inferred on the level of individuals in order to test whether people use different thresholds in changing environments. One important finding from this line of research is that the empirical thresholds indeed decrease monotonically as the sequence progresses, as prescribed by the optimal solution. However, many participants were systematically biased away from the optimal threshold. Finally, Guan and Lee (in press) showed that the threshold depends on the nature of the environment: participants were generally more accurate when rewards were generated from a distribution skewed towards large rewards, consistent with a model in which thresholds are higher for large-reward environments.

These descriptive models have provided important insights into the dynamics of sequential choices but have neglected the underlying psychological processes that guide human decisions in such tasks. The goal of this study is to develop a new approximately normative model for optimal stopping that provides a psychological explanation for choice behavior given the distribution from which the options are sampled, the current value, and the number of remaining choices. This model is loosely based on the optimal solution, which utilizes the same information. However, in contrast to the optimal solution it is not based on expected *reward*, but on expected *rank*. We assume that individuals decide to accept or reject each option based on an estimate of how likely it is that a better option will appear in the remaining choices. Unlike most earlier models, our rank-based model does not employ a threshold on reward.

Optimal Model: Expected Reward

We consider a decision maker who encounters a sequence of options with rewards denoted by $x_1, \ldots, x_N$ and she wants to find the minimum value in the sequence. If the decision maker accepts option i, then the sequence terminates and she receives x_i; otherwise, she continues to the next option. When the last option N is reached, it must be accepted. The optimal policy is to choose option i when it goes below a position-dependent threshold t_i (which can be obtained by backward induction, as described below). As shown by Gilbert and Mosteller (1966), the optimal policy depends only on the distribution of rewards and the number of remaining choices. In order to model possibly stochastic decisions, we relax this deterministic model uses a logistic sigmoid policy with sensitivity parameter θ:

$$\pi_i = \frac{1}{1+\exp\{\theta(x_i - t_i)\}}, \tag{1}$$

where π_i denotes the probability of accepting option i. Small values of θ produce more stochasticity in decisions, whereas the policy approaches optimality in the limit $\theta \to \infty$.[1]

The optimal threshold t_i is calculated in the following manner: The threshold of the final item is 0, because the rules of the task stipulate that the final item must be accepted if no earlier item has been chosen. The thresholds for the previous items are determined by working backward from the final item, using conditional expectations (Gilbert & Mosteller, 1966). First, we calculate the expected value of the final item. This is the expectation of the overall probability distribution from which the options are sampled, because the threshold of this item is 0. To maximize expected reward, one's policy should be to accept a particular option if it is better than the expected reward if one continues under the optimal policy. The second-to-last item should be accepted if its value is greater than the expected value of the final item. This means that the threshold of the second-to-last item is the expected value of the last item.

The expected value of the second-to-last item is the expected value of the part of the probability distribution that is better (in our case smaller) than the threshold for the second-to-last item. The probability of this expected value is the area under the probability distribution that is better than this threshold. The overall expected reward at the second-to-last position (and therefore the threshold for the third-to-last item) is calculated as follows: we multiply the expected value for the second-to-last item with its probability plus the expected value of the last item multiplied with its probability (which is equal to 1 minus the probability of the second-to-last item). The remaining thresholds are calculated in the same way.

New Model: Expected Probability

Unfortunately, the optimal threshold calculation is computationally expensive due to its recursive structure. In fact, human behavior cannot be described adequately by the optimal solution (e.g., Guan et al., 2015).

We propose to replace the threshold with a different mechanism. We argue that the decision maker estimates the probability τ for observing a better option in the future. This estimate is used as the basis for a decision variable δ_i. Specifically, the choice function for the new model is formulated as follows:

$$\pi_i = \frac{1}{1+\exp\{\theta \cdot (\delta_i - 0.5)\}}. \tag{2}$$

When $\delta_i > 0.5$, the agent believes that it is likely to encounter a better option and therefore tends to reject the option. When $\delta_i < 0.5$ the agent believes that it is unlikely to encounter a better option and therefore tends to accept the option.

In order to calculate δ_i, we first have to compute τ_i for position i. As mentioned above, we are using a task where the goal is to find the smallest value. Therefore we compute τ_i as the probability to encounter at least one value that is better than the current option for the future positions as follows:

$$\tau_i = P_i(\exists x_j : x_j < x_i) = 1 - \left(1 - \Phi\left(\frac{x_i - \mu}{\sigma}\right)\right)^{N-i}, \tag{3}$$

where N is the number of the total tickets (in our study 10 tickets), and j stands for the positions of the remaining $N-i$ options. $\Phi(\bullet)$ is the cumulative distribution function (CDF) for a standard normal (i.e., Gaussian) distribution and μ and σ are the mean and standard deviation of the distribution from which the options are drawn (in our experiment $\mu = 180$ and $\sigma = 20$).

To gain more modeling flexibility, we introduce a parameter α to account for individual differences in scaling the probability τ_i. Therefore, we calculate the rescaled probability as

[1] It should be noted that the usual formulation of Equation 1 is in terms of a policy to find the maximum value in a sequence. In this case, θ would have to be multiplied by -1. This also applies to Equation 2.

follows:

$$\delta_i = \begin{cases} 0 & , \quad \tau_i \cdot \alpha < 0 \\ \tau_i \cdot \alpha & , \quad 0 < \tau_i \cdot \alpha < 1 \\ 1 & , \quad \tau_i \cdot \alpha > 1 \end{cases} \tag{4}$$

We illustrate the effect of the parameter α in Figure 1. The dashed line for $\alpha = 1$ shows δ_i for each position, given that the actual value is 167 and the remaining options are sampled from a normal distribution with mean 180 and standard deviation 20. Individuals that underestimate this probability tend to stop earlier ($\alpha < 1$) and people that overestimate the probability ($\alpha > 1$) stop later in the sequence.

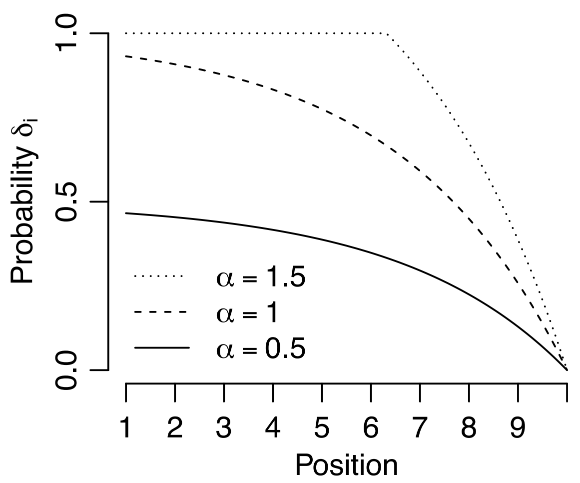

Figure 1: δ_i (rescaled probability of observing a better option) for $\alpha = 1$ (in this case, $\delta_i = \tau_i$), and for $\alpha = 1.5$ (dotted line) and $\alpha = 0.5$ on each position with current ticket value 167 and normal distribution with mean 180 and standard deviation 20.

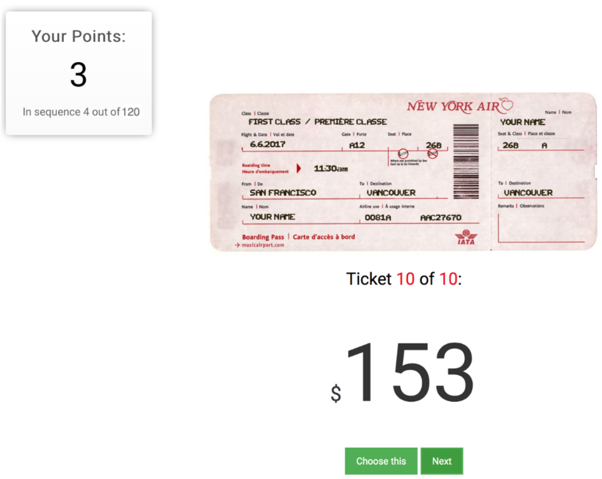

Figure 2: Screenshot of the ticket-shopping task.

Experiment 1

The goal of Experiment 1 was to obtain a data set which could be used to compare the expected probability model with the optimal threshold model within a full information optimal stopping task. One novel feature of our experimental design was that we included a within-subjects manipulation of position for a specific value within a fixed sequence of tickets. This allowed us to look at the probabilistic choice behavior for one and the same value across positions.

We asked participants to solve an optimal stopping problem in the form of a computer-based ticket-shopping task. Participants were told that they are planning a plane trip and need to buy a ticket. Ticket values were presented sequentially and the goal was to find the cheapest ticket. Once they rejected a ticket, they could not return to this option. The interface provided feedback about the rank of the chosen ticket and a cumulative count of the collected points that had been made for all of the completed trials (see Figure 2). Before participants had to perform the shopping task, they learned the distribution of values from which the tickets were drawn.

Materials and Methods

Participants We recruited 69 participants (mean age: 31 years, range: 19-62) on Amazon Mechanical Turk to participate in the experiment. Participants gave informed consent, and the Harvard Committee on the Use of Human Subjects approved the experiment. Participants were excluded from analysis if they accepted the first option in the sequence in more than 95% of the trials since this would show that they did not search at all. After applying these criteria, we included data from 60 participants in the subsequent analysis.

Procedure In the first phase of the experiment, participants experienced the distribution of the values. The procedure was as follows: Participants encountered sequentially 50 ticket values drawn from a normal distribution with mean 180 and standard deviation 20. After every ten tickets participants had to guess the value of the next ticket. This question was added to ensure that participants learn the distribution and the correct answer to this question was the mean of the previous ten tickets. After each guess participants were told the correct response. At the end of the learning phase participants were asked to complete a histogram (by dragging the bars) for additional 100 tickets that could be drawn from the same distribution. Participants received feedback by observing the correct distribution superimposed over their estimate (Goldstein, 2014). Visual inspection of the performance in the histogram task suggested that participants learned the target distribution fairly well.

In the second phase of the experiment participants performed the ticket-shopping task. It started with a practice trial followed by 120 test trials. In each trial participants searched through a sequence of ten ticket values. For each ticket, they could decide to accept or reject it at their own speed. People were aware that they could see up to 10 tickets in each trial and they were always informed about the actual position and the number of remaining tickets (see Figure 2 for a screen shot). It was not possible to go back to an earlier option after it was initially declined. Once they reached the last ticket (10^{th}) they were forced to choose this ticket. When participants accepted the ticket, they received explicit feedback about its rank and the points earned. Then participant moved to next sequence of tickets.

Participants were paid according to their performance. They received 3 points if they chose the best ticket (Rank 1) and 1 point for the second, third, and forth-best ticket (Rank 2, 3, 4). Participants received a base payment of $4 and earned between $0 and $6 additionally depending on their performance in the task.

Within-Subjects Manipulation We manipulated 80 out of 120 sequences as follows: The rank order of the sequence was kept constant and also the values were almost identical within a range of ± 1 point (see Table 1). Then we inserted the value of 167 on position 3 (for 20 sequences), on position 5 (for 20 sequences), on position 7 (for 20 sequence) and on position 9 (for 20 sequences) (see Table 1). In the following we restrict our analysis to the responses to the tickets with value 167 in the manipulated sequences. The remaining 40 sequences were chosen randomly to distract the participants from the regularity of the 80 manipulated sequences.

Results

Manipulation Check We asked participants at the end of the experiment if they noticed anything special about the sequences. Only four participants reported to have noted something, but it was not related to our within-subjects manipulation.

Observable behavior Figure 3 shows the choice probabilities for each individual when the value 167 was on position 3, 5, 7, and 9. The data suggest that participants decided probabilistically on each position, and the probability of choosing the same value increased with position.

To test this hypothesis quantitatively, we used a binomial generalized linear mixed model for the individual choices with fixed-effect for position (as categorical variable) and by-participant random intercepts, random slopes for the fixed-effect, and correlation among random terms. The model revealed a significant effect of position, $\chi^2(3) = 83.44$, $p < .0001$. The estimated marginal means (EMM) for the different positions exhibited a clearly increasing pattern: EMM_3 = 53%, 95%-CI [43%, 63%], EMM_5 = 69%, [61%, 76%], EMM_7 = 88%, [84%, 91%], and EMM_9 = 99%, [97%, 99%].

Table 1: Manipulated Sequences in Experiment 1

1	2	3	4	5	6	7	8	9	10
174	186	**167**	211	178	181	187	195	209	157
175	186	192	210	**167**	180	188	193	210	158
175	185	191	211	179	181	**167**	194	209	157
174	187	191	210	178	180	187	195	**167**	158

Note. Values in bold were manipulated across sequences.

Table 2: Modeling Results of Experiment 1

Model	Param	Estimate	BIC
Optimal	θ	$0.04\ (0.002, 0.13)$	4279
Expected Probability	θ	$15.6\ (8.7 - 16.3)$	3200
	α	$0.59\ (0.56 - 0.66)$	

Note. The value in the estimate column is the median across participants, values in parentheses give the interquartile range. Param = Parameter.

However, Figure 3 also shows that there was considerable inter-individual variability in the slope of the increase.

Modeling Results Next we fit the two models to the data using maximum-likelihood estimation. The optimal model had one free parameter, θ, and the expected probability model had two free parameters, θ and α. We applied the model to the individual-level data with 80 data points per participant. Figure 3 shows the model predictions derived from the individual-level maximum-likelihood parameter estimates of both models.

Inspection of Figure 3 reveals a close fit between observed and predicted data for the expected options model. In contrast, the predictions for the optimal model are in many cases quite far from the observed data. Furthermore, in some cases the optimal model even makes the clearly false prediction that the probability to accept 167 *decreases* along positions. This occurred when θ took on negative values. Overall, the Figure 3 shows that the participants' behavior is clearly not optimal and that the expected probability model provides a considerably better account than the optimal model.

The impression that the expected probability model provides a better account than the optimal model was corroborated by an analysis using the Bayesian information criterion (BIC) presented in Table 2: the difference in BIC was over 1000. Table 2 also shows the median parameter estimates for both models. For the expected probability model, we observed a median $\alpha = 0.59$, which shows that participants discounted the probability for observing a better option.

Experiment 2

The goal of Experiment 2 was to collect a data set that would allow us to test the out-of-sample predictive ability of the two models. To this end we ran an experiment that was almost identical to Experiment 1 on a new set of participants with the only difference that we replaced the ticket value that was repeated across sequences from 167 to 170. The mean and standard deviation of the generating distribution remained at 180 and 20, respectively.

Materials and Methods

Participants 70 participants (mean age: 33 years, range: 20–71) were recruited on Amazon Mechanical Turk. Using the same criteria as in Experiment 1 we excluded two participants and the following analysis is based on the remaining 68

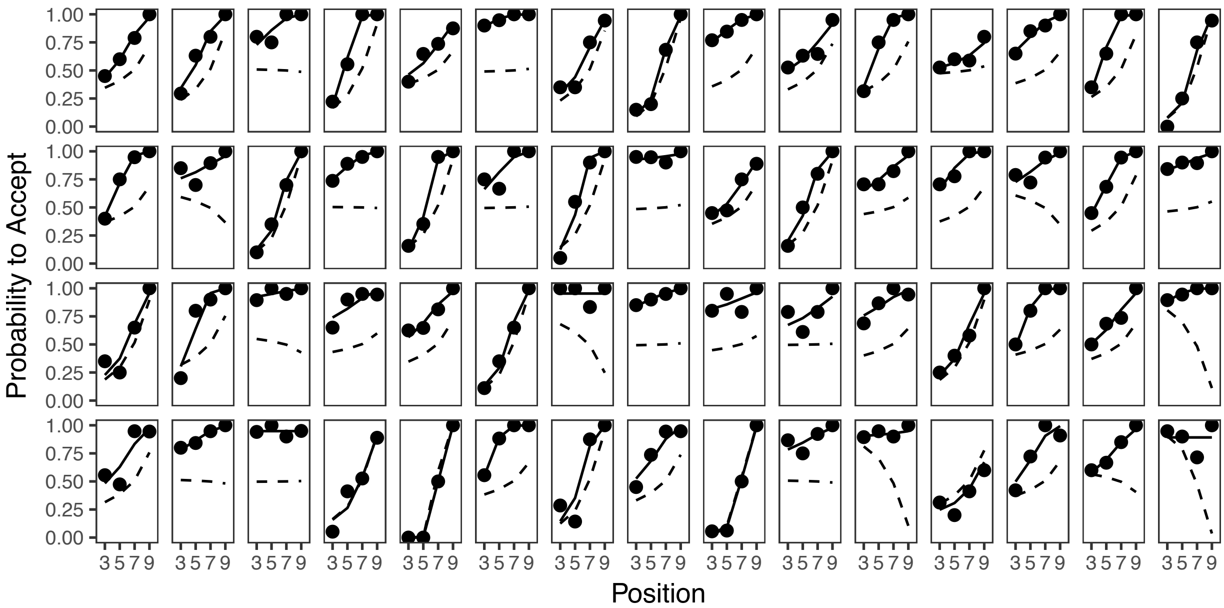

Figure 3: Individual choice probabilities and model fits for Experiment 1. Each panel shows the data from one participant. The points show the choice probability on position 3, 5, 7, and 9 for a ticket with a value of 167. The solid lines show the predictions of the expected probability model. The dashed lines show the predictions of the optimal model.

participants. As in Experiment 1, the visual inspection of the performance in the histogram task showed no gross violations from the target distribution.

Results

Manipulation Check 6 participants reported to have noted something, but as in Experiment 1 reported irrelevant details unrelated to our within-subjects manipulation.

Modeling Results To assess the predictive ability of the models we used the median parameter estimates obtained in Experiment 1 (see Table 2) to generate predictions for the value 170 at positions 3, 5, 7, and 9 from both models. These predictions as well as the choice probabilities aggregated across participants are shown in Figure 4. As can be seen, the expected probability model again provides a relatively accurate account, despite the fact that the parameter values were derived from a different group of participants and from options with a slightly different value. For position 3 the prediction is spot on. For positions 5, 7, and 9 the model predicts a larger acceptance rate than observed with the difference increasing across positions. In contrast, the predictions of the optimal model are again relatively far off, with the only exception being position 3.

Discussion

The primary goal of this work is to understand the psychological processes that are involved in the choice behavior in optimal stopping tasks. We suggest that the decision to accept or reject is governed by the probability of observing a better option in the future, and presented a computational model that translates this probability into choice probabilities. Our model predictions closely matched human choice probabilities in two experiments. Importantly, the predictions of the second experiment were generated using parameter values obtained in the first experiment showing that our model is able to perform true out-of-sample prediction on the aggregated level.

The important idea of our model is that we re-framed the processes underlying behavior in optimal stopping tasks in completely probabilistic terms. Instead of relying on a threshold, individuals are assumed to estimate the probability of better future options and base their decision on this probability. It is a question for future research how this probability is obtained. One obvious candidate would be sampling from memory (e.g., Stewart, Chater, & Brown, 2006).

In our study participants learn the distribution of the ticket values in the first phase of the experiment. We make this assumption based on the idea that humans facing an optimal stopping task (e.g. search for cheapest ticket, search for apartment or partner) are usually familiar with the range of the option's values they will encounter. In this manner we can also minimize learning during the task, since tickets encountered in the training phase are sampled from the same distribution as in the testing phase. We verified the lack of learning during the task by comparing the performance in the first half and the second half of the problems.

Previous studies have found that in the original optimal stopping problem, people often sample less than what an

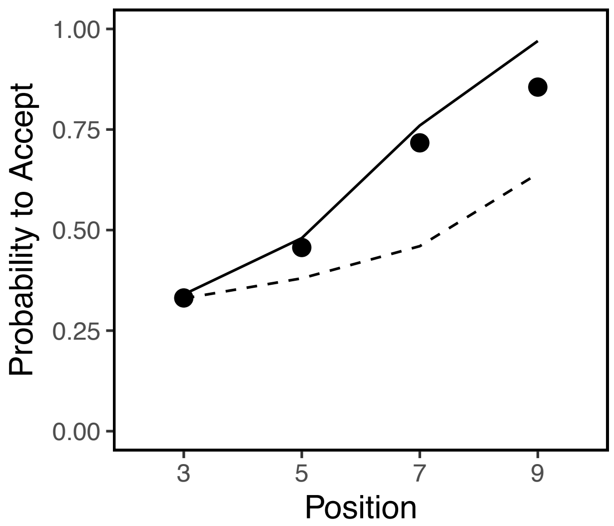

Figure 4: Out-of-sample predictive ability of the expected probability model (solid line) and optimal model (dashed line) against the aggregated choice data from Experiment 2. Parameter values generating the predictions were taken from Experiment 1 (see Table 2).

optimal strategy would advise depending on whether or not search costs are assumed (Zwick, Rapoport, Lo, & Muthukrishnan, 2003). This conclusion is consistent with our finding that participants underestimate the probability for observing a better option in the future (estimated parameter $\alpha < 1$ for all participants, see Table 2 and Figure 1) and therefore accept too early. Moreover, individual differences in this tendency to underestimate the probability of finding a better option could be related to individual differences reported in related fields such as risk aversion, ambiguity aversion, or delay discounting.

Much of the previous research on optimal stopping problems has proposed formal models of the decision-making process (e.g., Bearden et al., 2006; Gilbert & Mosteller, 1966; Seale & Rapoport, 1997), although sometimes their evaluation has taken the form of simulation than making inferences from human data. The threshold model proposed from Guan and Lee (in press) can describe the individual thresholds for a problem with a specific length and distribution but makes strong theoretical assumptions in order to make the deterministic threshold model probabilistic. Although these models give a good insight into the dynamics of the decision making in optimal stopping problems, they do not allow predictions for different variants of the problem as e.g. different length or different distributions. The main difference of our model from most existing threshold models is that it uses the information of the sampling environment and current option and thus can be used to generate predictions for new environments and new options such as we did for Experiment 2. Therefore, our model provides a framework that allows a more principled approach for studying the effect of various external variables on behavior in optimal stopping tasks. For example, one important finding is that explicit search costs lead to earlier stopping (e.g., Seale & Rapoport, 1997). One possibility is that such cost are reflected in a reduced α parameter. Alternatively, they could require the inclusion of an additional parameter or processes explicitly accounting for this.

Acknowledgments

Christiane Baumann was supported by Swiss National Science Foundation Grant P0ZHP1_168889.

References

Bearden, J. N., Rapoport, A., & Murphy, R. O. (2006). Experimental studies of sequential selection and assignment with relative ranks. *Journal of Behavioral Decision Making*, *19*(3), 229–250.

Gilbert, J. P., & Mosteller, F. (1966). Recognizing the maximum of a sequence. *Journal of the American Statistical Association*, *61*(313), 35-73.

Goldstein, D. G. (2014). Lay understanding of probability distributions. , *9*(1), 1–14.

Guan, M., & Lee, M. (in press). The effect of goals and environments on human performance in optimal stopping problems. *Decision*.

Guan, M., Lee, M., & Vandekerckhove, J. (2015). A Hierarchical Cognitive Threshold Model of Human Decision Making on Different Length Optimal Stopping Problems. In *Proceedings of the 37th annual conference of the cognitive science society* (pp. 824–829). Austin, TX: Cognitive Science Society.

Lee, M. (2006). A hierarchical bayesian model of human decision-making on an optimal stopping problem. *Cognitive Science*, *30*(3), 1–26.

Seale, D. A., & Rapoport, A. (1997). Sequential Decision Making with Relative Ranks : An Experimental Investigation of the " Secretary Problem ". *Organizational Behavior and Human Decision Making*, *69*(3), 221–236.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*, 1–26.

von Helversen, B., & Mata, R. (2012). Losing a dime with a satisfied mind: positive affect predicts less search in sequential decision making. *Psychology and Aging*, *27*(4), 825–39.

Zwick, R., Rapoport, A., Lo, A. K. C., & Muthukrishnan, a. V. (2003). Consumer Sequential Search: Not Enough or Too Much? *Marketing Science*, *22*(4), 503–519.