# Learning in the Repeated Secretary Problem

DANIEL G. GOLDSTEIN, Microsoft Research
R. PRESTON MCAFEE, Microsoft Corporation
SIDDHARTH SURI, Microsoft Research
JAMES R. WRIGHT, Microsoft Research

In the classical secretary problem, one attempts to find the maximum of an unknown and unlearnable distribution through sequential search. In many real-world searches, however, distributions are not entirely unknown and can be learned through experience. To investigate learning in such a repeated secretary problem we conduct a large-scale behavioral experiment in which people search repeatedly from fixed distributions. In contrast to prior investigations that find no evidence for learning in the classical scenario, in the repeated setting we observe substantial learning resulting in near-optimal stopping behavior. We conduct a Bayesian comparison of multiple behavioral models which shows that participants' behavior is best described by a class of threshold-based models that contains the theoretically optimal strategy. In fact, fitting such a threshold-based model to data reveals players' estimated thresholds to be surprisingly close to the optimal thresholds after only a small number of games.

## 1 INTRODUCTION

This paper concerns a popular optimal stopping problem and whether people improve at it with experience. Consider the following scenario:

> You have been captured by an evil dictator. He forces you to play a game. There are 15 boxes. Each box has a different amount of money in it. You can open any number of boxes in any order. After opening each box, you can decide to open another box or stop. If you stop right after opening the box with the most money in it (of the 15 boxes), then you win. However, if you stop at any other time, you lose and the evil dictator will kill you.

Readers may recognize this as a variant of the "secretary problem", a compact question with an optimal stopping rule as its answer. As mathematical puzzles go, the secretary problem is a great success. Since its early published appearances in the mid-twentieth century, the problem has been modified and analyzed in hundreds of academic articles. Freeman [1983] provides an early review that itself has garnered several hundred citations. Even 25 years ago a review article noted that the secretary problem, "constitutes its own 'field' of study within mathematics-probability-optimization" [Ferguson, 1989].

The intense academic interest in the secretary problem may have to do with its similarity to real-life search problems such as choosing a mate [Todd, 1997], choosing an apartment [Zwick et al., 2003] or hiring, for example, a secretary. It may have to do with the way the problem exemplifies the concerns of core branches of economics and operations research that deal with search costs. Lastly, the secretary problem may have endured because of curiosity about its fascinating solution. In the classic version of the problem, the optimal strategy is to ascertain the maximum of the first $1/e$ boxes and then stop after the next box that exceeds it. Interestingly, this $1/e$ stopping rule wins about $1/e$ of the time in the limit [Gilbert and Mosteller, 1966]. The curious solution to the secretary problem only holds when the values in the boxes are drawn from an unknown distribution. To make this point clear, in some empirical studies of the problem, participants only get to learn the rankings of the boxes instead of the values (e.g., [Seale and Rapoport, 1997]). But is it realistic to assume that people cannot learn about the distributions in which they are searching?

In many real-world searches, people can learn about the distribution of the quality of candidates as they search. The first time a manager hires someone, she may have only a vague guess as to the quality of the candidates that will come through the door. By the fiftieth hire, however, she'll have hundreds of interviews behind her and know the distribution rather well. This should cause her accuracy in a real-life secretary problem to increase with experience.

While people seemingly *should* be able improve at the secretary problem with experience, surprisingly, prior academic research does not find evidence that they do. For example, Campbell and Lee [2006] attempted to get participants to learn by offering enriched feedback and even financial rewards in a repeated secretary problem, but concluded "there is no evidence people learn to perform better in any condition". Similarly, Lee [2006] found no evidence of learning, nor did Seale and Rapoport [1997].

In contrast, by way of a randomized experiment with thousands of players, we find that performance improves dramatically over a few trials and soon approaches optimal levels. We will show that players steadily increase their probability of winning the game with more experience, eventually getting to within 5 percentage points of the optimal win rate. Then we show that the improved win rates are due to players learning to make better decisions on a box-by-box basis and not just due to aggregating over boxes. Furthermore we will show that the learning we observe occurs in a noisy environment where the feedback they get, i.e. win or lose, may be unhelpful. After showing various types of learning in our data we turn our attention to modeling the players behavior. Using a Bayesian comparison framework we show that players' behavior is best described by a family of threshold-based models which include the optimal strategy. Moreover, the estimated thresholds are surprisingly close to the optimal thresholds after only a small number of games.

## 2 RELATED WORK

While the total number of articles on the secretary problem is large [Freeman, 1983], our concern with empirical, as opposed to purely theoretical, investigations reduces these to a much smaller set. We discuss here the most similar to our investigation. Ferguson [1989] usefully defines a "standard" version of the secretary problem as follows:

1. There is one secretarial position available.
2. The number $n$ of applicants is known.
3. The applicants are interviewed sequentially in random order, each order being equally likely.
4. It is assumed that you can rank all the applicants from best to worst without ties. The decision to accept or reject an applicant must be based *only* on the relative ranks of those applicants interviewed so far.
5. An applicant once rejected cannot later be recalled.
6. You are very particular and will be satisfied with nothing but the very best.

The one point on which we deviated from the standard problem is the fourth. To follow this fourth assumption strictly, instead of presenting people with raw quality values, some authors (e.g., Seale and Rapoport [1997]) present only the ranks of the candidates, updating the ranks each time a new candidate is inspected. This prevents people from learning about the distribution. However, because the purpose of this work is to test for improvement when distributions are learnable, we presented participants with actual values instead of ranks.

Others properties of the classical secretary problem could have been changed. For example, there exist alternate versions in which there is a payout for choosing candidates other than the best. These "cardinal" and "rank-dependent" payoff variants [Bearden, 2006] violate the sixth property above. We performed a literature search and found fewer than 100 papers on these variants, while finding over 2,000 papers on the standard variant. Our design preserves the sixth property for two reasons. First, by preserving it, our results will be directly comparable to the the greatest number of existing theoretical and empirical analyses. Second, changing more than one variable at a time is undesirable because it makes it difficult to identify which variable change is responsible for changes in outcomes.

While prior investigations, listed next, have looked at people's performance on the secretary problem, none have exactly isolated the condition of making the distributions learnable. Across several articles, Lee and colleagues [Campbell and Lee, 2006, Lee, 2006, Lee et al., 2004] conducted experiments in which participants were shown values one at a time and were told to try to stop at the maximum. Across these papers, the number of candidates (or boxes or secretaries) ranged from 5 to 50 and participants played from 40 to 120 times each. In all these studies, participants knew that the values were drawn from a uniform distribution between 0 and 100. For instance, Lee [2006] states, "It was emphasized that … the values were uniformly and randomly distributed between 0.00 and 100.00". With such an instruction, players can immediately and exactly infer the percentiles of the values presented to them, which helps them calculate the probability that unexplored values may exceed what they have seen. As participants were told about the distribution, these experiments do not involve learning distribution from experience, which is our concern. Information about the distribution was also conveyed to participants in a study by Rapoport and Tversky [1970], in which seven individual participants viewed an impressive 15,600 draws from probability distributions over several weeks before playing secretary problem games with values drawn from the same distributions. These investigations are similar to ours in that they both involve repeated play and that they present players with actual values instead of ranks. That is, they depart from the fourth feature of the standard secretary problem listed above. These studies, however, differ from ours in that they give participants information about the distribution from which the values are drawn before they begin to play. In contrast, in our version of the game, participants are told no information about the distribution, see no samples from it before playing, and do not know what the minimum or maximum values could be. This key difference between the settings may have had a great impact. For instance, in the studies by Lee and colleagues, the authors did not find evidence of learning or players becoming better with experience. In contrast, we find profound learning and improvement with repeated play.

Corbin et al. [1975] ran an experiment in which people played repeated secretary problems, with a key difference that these authors manipulated the values presented to subjects with each trial. For instance, the authors varied the support of the distribution from which values were drawn, and manipulated the ratio and ranking of early values relative to later ones. The manipulations were done in an attempt to prevent participants from learning about the distribution and thus make each trial like the "standard" secretary problem with an unknown distribution. Similarly, Palley and Kremer [2014] provide participants with ranks for all but the selected option to hinder learning about the distribution. In contrast, because our objective is to investigate learning, we draw random numbers without any manipulation.

Finally, in a study by Kahan et al. [1967], groups of 22 participants were shown up to 200 numbers chosen from either a left skewed, right skewed or uniform distribution. In this study, as well as ours, participants were presented with actual values instead of ranks. Also like our study, distributions of varying skew were used as stimuli. However, in Kahan et al. [1967], participants played the game just one time and thus were not able to learn about the distribution to improve at the game.

In summation, for various reasons, prior empirical investigations of the secretary problem have not been designed to study learning about the distribution of values. These studies either informed participants about the parameters of the distribution before the experiment, allowed participants to sample from the distribution before the experiment, replaced values from the distribution with ranks, manipulated values to prevent learning, or ran single-shot games in which the effects of learning could not be applied to future games. Our investigation concerns a repeated secretary problem in which players can observe values drawn from distributions that are held constant for each player from game to game.

## 3 EXPERIMENTAL SETUP

To collect behavioral data on the repeated secretary problem with learnable distributions of values, we created an online experiment. The experiment was promoted as a contest on several web logs and attracted 5,220 players who played the game at least one time. A total of 40,754 games were played on the site. As users arrived at the game's landing page, they were cookied and their browser URL was automatically modified to include an identifier. These two steps were taken to assign all plays on the same browser to the same user id and condition, and to track person-to-person sharing of the game URL. Any user determined to arrive at the site via a shared URL (i.e., a non-cookied user entering via a modified URL) was excluded from analysis and is not counted in the 5,220 we analyze. We note that including these users makes little difference to our results and that we only exclude them to obtain a set of players that were randomly assigned to conditions by the website. Users saw the following instructions. Blanks stand in the place of the number of boxes, which was randomly assigned and will be described later.

> You have been captured by an evil dictator. He forces you to play a game. There are ___ boxes. Each box has a different amount of money in it. You can open any number of boxes in any order. After opening each box, you can decide to open another box or you can stop by clicking the stop sign. If you hit stop right after opening the box with the most money in it (of the ___ boxes), then you win. However, if you hit stop at any other time, you lose and the evil dictator will kill you. Try playing a few times and see if you improve with practice.

Immediately beneath the instructions was an icon of a traffic stop sign and the message "When you are done opening boxes, click here to find out if you win". Beneath this on the page were hyperlinks stating "Click here to open the first box","Click here to open the second box", and so on. As each link was clicked, an AJAX call retrieved a box value from the server, recorded it in a database and presented it to the user. If the value in the box was the highest seen thus far, it was marked as such on the screen. See Figure 9 in the Appendix for a screenshot. Every click and box value was recorded, providing a record of every box value seen by every player, as well as every stopping point. If a participant tried to stop at a box that was dominated by (i.e., less than) an already opened box, a pop-up explained that doing so would necessarily result in the player losing. After clicking on the stop icon or reaching the last box in the sequence, participants were redirected to a page that told them whether they won or lost, and showed them the contents of all the boxes, where they stopped, where the maximum value was, and by how many dollars (if any) they were short of the maximum value. To increase the amount of data submitted per person, players were told "Please play at least six times so we can calculate your stats".

### 3.1 Experimental Conditions

To allow for robust conclusions that are not tied to the particularities of one variant of the game, we randomly varied two parameters of the game: the distributions and the number of boxes. Each player was tied to these randomly assigned conditions so that their immediate repeat plays, if any, would be in the same conditions.

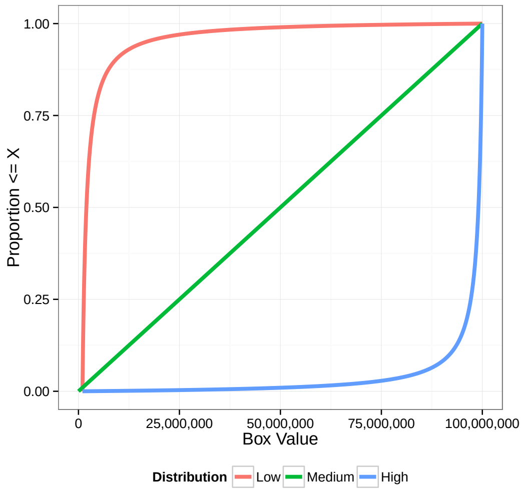


Fig. 1. Cumulative distribution functions (CDFs) of the three distributions from which box values were randomly drawn in the experiment. As probability density functions, the "low" distribution is strongly positively skewed, the "medium" distribution is a uniform distribution, and the "high" distribution is strongly negatively skewed.

*3.1.1 Random assignment to distributions.* The box values were randomly drawn from one of three probability distributions, as pictured in Figure 1. The maximum box value was 100 million, though this was not known by the participants. The "low" condition was strongly negatively skewed. Random draws from it tend to be less than 10 million, and the maximum value tends to be notably different than the next highest value. For instance, among 15 boxes drawn from this distribution, the highest box value is, on average, about 14.5 million dollars higher than the second highest value. In the "medium" condition numbers were randomly drawn from a uniform distribution ranging from 0 to 100 million. The maximum box values in 15 box games are on average 6.2 million dollars higher than the next highest values. Finally, in the "high" condition, boxes values were strongly negatively skewed, and bunched up near 100 million. In this condition, most of the box values tend to look quite similar (typically eight-digit numbers greater than 98 million). Among 15 boxes, the average difference between the maximum value and the next highest is rather small at only about 80,000 dollars. Note that the distributions are merely window dressing and are irrelevant for playing the game. Players only need to attend to percentiles of the distribution to make optimal stopping decisions. However, the varying distributions leads to more generalizable results than an analysis of a single, arbitrary setting.

*3.1.2 Random assignment to number of boxes.* The second level of random assignment concerned the number of boxes, which was either 7 or 15. While one would think this approximate doubling in the number of boxes would make the game quite a bit harder, it only affects theoretically optimal win rates by about 2 percentage points, as will be shown. Like with the distributions, varying the number of boxes leads to more generalizable results.

With either 7 or 15 boxes and three possible distributions, the experiment had a 2×3 design. In the 7 box condition, 902, 850, and 875 participants were randomly assigned to the low, medium, and high distributions, respectively, and in the 15 box condition, the counts were 877, 900, 816, respectively. The differences in cell counts were non significant ($p = .3$, by chi-square test), consistent with successful random assignment. We next turn to describing how an optimal agent would go about playing the game.

### 3.2 Optimal play

Before we begin to analyze the behavioral data gathered from these experiments we first discuss how one would play this game optimally. Assume values $X_t$ are drawn in an independently and identically distributed fashion from a cumulative distribution function $F$. One period consists of a player opening a box with a realization $x_t$ of $X_t$ in box $t$. Periods are numbered in reverse order starting at $T$, so $t = T, \dots, 1$. Periods are thus numbered in the reverse order of the boxes in the game, that is, opening the first box implies being in the seventh period (of a seven box game). An action is to select or reject. For example at time $t$, select box $t$, otherwise reject box $t$. Let

$$h_t = \max \ \{x_t, ..., x_T\}.$$

The history summary $h_{t+1}$ is visible to the player at time $t$. The payoff of the player who selects box $t$ is

$$\begin{cases} 1 & h_t = h_1 \\ 0 & h_t < h_1 \end{cases}.$$

Thus, the player only wins when they select the highest value. The problem is nontrivial because they are forced to choose without knowing future realizations of the $X_i$.

Optimal players will adopt a threshold rule, which says, possibly as a function of the history, accept the current value if it is greater than a critical dollar value $c_t$. It is a dominant strategy to reject any realization worse than the best historically observed, except for the last box which must be accepted if opened. In addition, in our game, a pop-up warning prevented players from choosing dominated boxes.

With a known distribution independently distributed across periods, the critical dollar value will be the maximum of the historically best value and a critical dollar value that does not depend on the history. The reason that the critical dollar value does not depend on the history is that there is nothing to learn about the future (known distribution) so it is either better to accept the current value than wait for a better future value or not; the point of indifference is exactly our critical dollar value. Thus, the threshold comes in the form $\max\{c_t, h_t\}$.

In addition, $c_t$ is non-decreasing in $t$. Suppose, for the sake of contradiction, that $c_t > c_{t+1}$. Taking any candidate in the interval $(c_{t+1}, c_t)$ entails accepting a candidate and then immediately regretting it, because as soon as the candidate is accepted, the candidate is no longer acceptable, being worse than $c_t$.

Let $i = T - t$, and $z_i = F(c_t)$. We refer to the $z_t$ as the critical values, which are are the probabilities of observing a value less than the critical dollar values $c_t$. Let $p_t(h)$ be the probability of a win given a history $h$. Table 1 provides the critical values $z_t$ and the probability of winning given a zero history $p_t(0)$ for fifteen periods. Derivations of these figures are found in section A.1 in the Appendix.

The relevant entries for our study are the games of 7 and 15 periods. These calculations, which coincide with those found in Gilbert and Mosteller [1966], who did not provide Equation (7) (see the Appendix), show that experienced players, who know the distribution, can hope to win at best 62.2% of the games for 7 period games and just under 60% of the time for 15 period games. Note that these numbers compare favorably with the usual secretary result, which are lesser for all game lengths, converging to the famous $1/e$ as the length diverges. Thus there is substantial value in knowing the distribution.

As is reasonably well known, the value of the classical secretary solution can be found by choosing a value $k$ to sample, and then setting the best value observed in the first $k$ periods as

Table 1. Critical values and probability of winning given a known distribution of values for up to 15 boxes.

| Boxes left, $t$ | Critical values $z_t$ | Pr(Win) $p_t(0)$ |
|---|---|---|
| 1 | 0 | 1 |
| 2 | 0.5 | 0.750 |
| 3 | 0.6899 | 0.684 |
| 4 | 0.7758 | 0.655 |
| 5 | 0.8246 | 0.639 |
| 6 | 0.8559 | 0.629 |
| 7 | 0.8778 | 0.622 |
| 8 | 0.8939 | 0.616 |
| 9 | 0.9063 | 0.612 |
| 10 | 0.9160 | 0.609 |
| 11 | 0.9240 | 0.606 |
| 12 | 0.9305 | 0.604 |
| 13 | 0.9361 | 0.602 |
| 14 | 0.9408 | 0.600 |
| 15 | 0.9448 | 0.599 |

a critical value. The distribution of the maximum of the first $k$ is $F(x)^k$. The probability that a better value is observed in round $m$ is $(1 - F(x))F(x)^{m-k-1}$. Suppose this value is $y$; then this value wins with probability $F(y)^{T-m}$. Thus the probability of winning for a fixed value of $k$ and $T$ periods is $\frac{k}{T}\sum_{m=k+1}^{T}\frac{1}{m-1}$. See the derivation in Section A.2 in the Appendix. The optimal value of $k$ maximizes $\frac{k}{T}\sum_{m=k+1}^{T}\frac{1}{m-1}$ and is readily computed to yield Table 2. Comparing the probability of winning shown in Tables 1 and 2 shows that making the distribution learnable allows for a much higher rate of winning.

How well can players do *learning* the distribution? To model this, we consider an idealized agent that plays the secretary problem repeatedly and learns from experience. The agent begins with the critical values and learns the percentiles of the distribution from experience; it will be referred to as the "LP" (learn percentiles) agent. The agent has a perfect memory, makes no mistakes, has derived the critical values in Table 1 correctly, and can re-estimate the percentiles of a distribution with each new value it observes. It is difficult to imagine a human player being able to learn at a faster rate than the LP agent. We thus include it as an unusually strong benchmark.

### 3.3 Learning Percentiles: The LP agent

The LP agent starts off knowing the critical values for a 7 or 15 box game in percentile terms. To be precise, these critical values are the first 7 or 15 rows under the heading $z_t$ in Table 1. (Despite the term "percentile", we use decimal notation instead of percentages for convenience.) The reason that the LP agent is not given the critical values as raw box values is that these would be unknowable because the distribution is unknown before the first play. However, it is possible to compute these critical values as percentiles from first principles, as we have done earlier in this section and in the Appendix. Armed with these critical values, the LP agent converts the box values it observes into percentiles in order to compare them to the critical values. The first box value the LP agent sees gets assigned an estimated percentile of .50. If the second observed box value is greater than the first, it estimates the second value's percentile to be .75 and re-estimates the first value's percentile to be .25. If the second value is smaller than the first, it assigns the estimate of .25 to the second value

Table 2. The probability of winning a game in the classical secretary problem (unknown distribution of values) for up to fifteen boxes

| Game Length (Periods) | Classical Secretary Problem Pr(Win) |
|---|---|
| 1 | 1 |
| 2 | 0.50 |
| 3 | 0.50 |
| 4 | 0.458 |
| 5 | 0.433 |
| 6 | 0.428 |
| 7 | 0.414 |
| 8 | 0.410 |
| 9 | 0.406 |
| 10 | 0.399 |
| 11 | 0.398 |
| 12 | 0.396 |
| 13 | 0.392 |
| 14 | 0.392 |
| 15 | 0.389 |

and .75 to the first value. It continues in this way, re-estimating percentiles for every subsequent box value encountered according to the percentile rank formula:

$$\frac{N_< + 0.5N_=}{N} \tag{1}$$

where $N_<$ is the number of values seen so far that are less than the given value, $N_=$ is the number of times the given value has occurred so far, and $N$ is the number of boxes opened so far.

After recomputing all of the percentiles, the agent compares the percentile of the box just opened to the relevant critical value and decides to stop if the percentile exceeds the critical value, or decides to continue searching if it falls beneath it, making sure never to stop on a dominated box unless it is in the last position and therefore has no choice. Recall that a dominated box is one that is less than the historical maximum in the current game. The encountered values are retained from game to game, meaning that the agent's estimates of the percentiles of the distribution will approach perfection and win rates will approach the optima in Table 1.

How well does the LP agent perform? Figure 2 shows its performance. Comparing its win rate on the first play to the 7 and 15 box entries in Table 2, we see that the LP agent matches the performance of the optimal player of the classic secretary problem in its first game. Performance increases steeply over the first three games and comes within a point of the theoretical maxima (black lines) in about a dozen games. In any given game a player can either stop when it sees the maximum value, in which case it wins, or the player could stop before or after the maximum value, in which case it loses. In addition to the win rates, Figure 2 also shows how often agents commit these two types of errors. Combined error is necessarily the complement of the win rate so the steep gain in one implies a steep drop the other. Both agents are more likely to stop before the maximum as opposed to after it, which we will see is also the case with human players.

The LP agent serves as strong benchmark against which human performance can be compared. It is useful to study its performance in simulation because the existing literature provides optimal win rates for many variations of the secretary problem, but is silent on how well an idealized agent

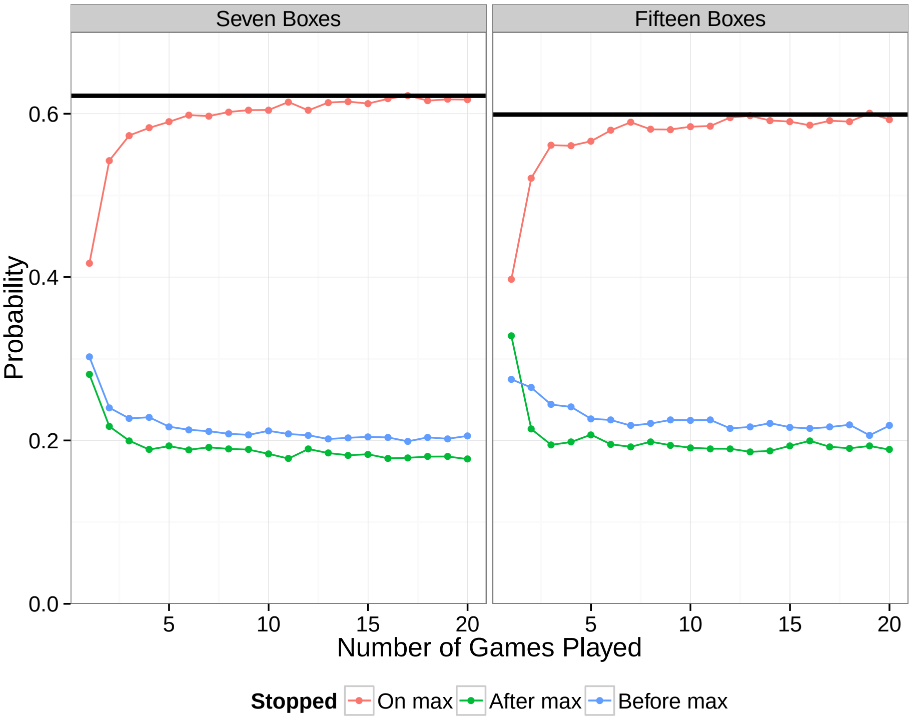


Fig. 2. Rates of winning (red lines), stopping too soon (blue lines) and stopping too late (green lines) for the LP agent. The theoretically maximal win rates for 7 and 15 boxes are given by the solid black lines.

would do when learning from scratch. In addition to win rates, these agents show the patterns of error that even idealized players would make on the path to optimality. In the next section, we will see how these idealized win and error rates compare to those of the human players in the experiment.

## 4 BEHAVIORAL RESULTS: LEARNING EFFECTS

As 40,754 games were played by 5,220 users, the average user played 7.81 games. Roughly half (51%) of users played 5 games or more, a quarter (24%) played 9 games or more, and a tenth played 16 games or more.

Prior research (e.g., Lee [2006]) has found no evidence of learning in repeated secretary problems with known distributions. What happens with unknown but learnable distribution? As shown in Figure 3, players rapidly improve in their first games and come within roughly five percentage points of theoretically maximal levels of performance. The leftmost point on each red curve indicates the how often first games are won. The next point to the right represents second games, and so on. The solid black lines at .622 and .599 show the maximal win rate attainable by an agent with perfect knowledge of the distribution. Note that these lines are not a fair comparison for early plays of the game in which knowledge of the distribution is imperfect or completely absent; in pursuit of a fair benchmark, we computed the win rates of the idealized LP agent shown in the dashed gray lines.

Performance in the first games, in which players have very little knowledge of the distribution is quite a bit lower than would be expected by optimal play in the classic secretary problem with 7 (optimal win rate .41) or 15 boxes (optimal win rate .39). Thus, some of the learning has to do with

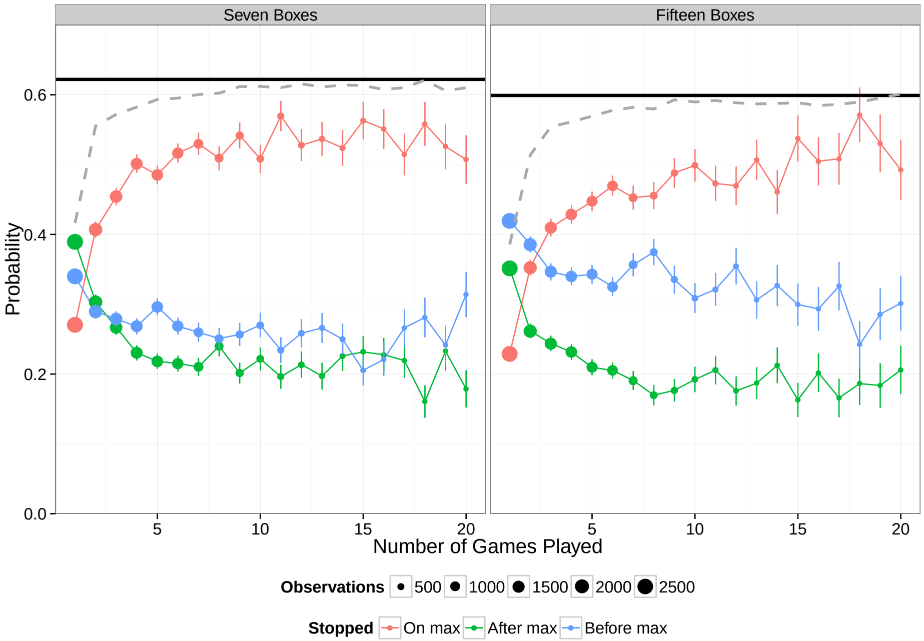


Fig. 3. Solid lines indicate the rates of winning the game and committing errors for human players with varying levels of experience. Error bars indicate ±1 standard error; when they are not visible they are smaller than the points. The area of each point is proportional to the number of players in the average. The graph is cut at 20 games as less than 1% of games played were beyond a user's 20th. The dashed gray line is the rate of winning the game for the LP agent.

starting from a low base. However, the classic version's optima are reached by about the second game and improvement continues another 10 to 15 percentage points beyond the classic optima.

One could argue that the apparent learning we observe is not learning at all but a selection effect. By this logic, a common cause (e.g., higher intelligence) is responsible both for players persisting longer at the game and winning more often. To check this, we created Figure 10, in the Appendix, which is a similar plot except it restricts to players who played at least 7 games. Because we see very similar results with and without this restriction, we conclude that Figure 3 reflects learning and not selection effects.

Having established that players do learn from experience, we turn our attention to what is being learned. One overarching trend is that soon after their first game, people learn to search less. As seen in Figure 4, in the first five games, the depth of search decreases by about a third of one box. Players can lose by stopping too early or too late. These search depth results suggest that concern with stopping too late is the primary concern that participants address early in their sequence of games. This is also reflected in the rate of decrease in the "stopping after max" errors in Figure 3. In both panels, rates of stopping after the maximum decrease most rapidly.

### 4.1 Optimality of box-by-box decisions

Do players' decisions become more optimal with experience? Recall that when the distribution is known one can make an optimal decision about when to stop search by comparing the percentile of an observed box value to the relevant critical value in Table 1. If the observed value exceeds the critical value, it is optimal to stop, otherwise it is optimal to continue search. In Figure 5, the

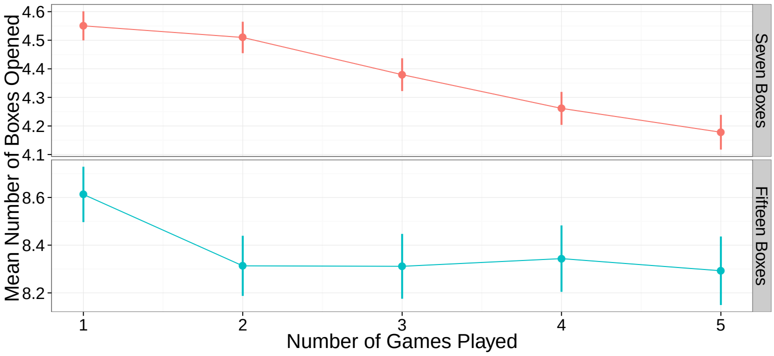
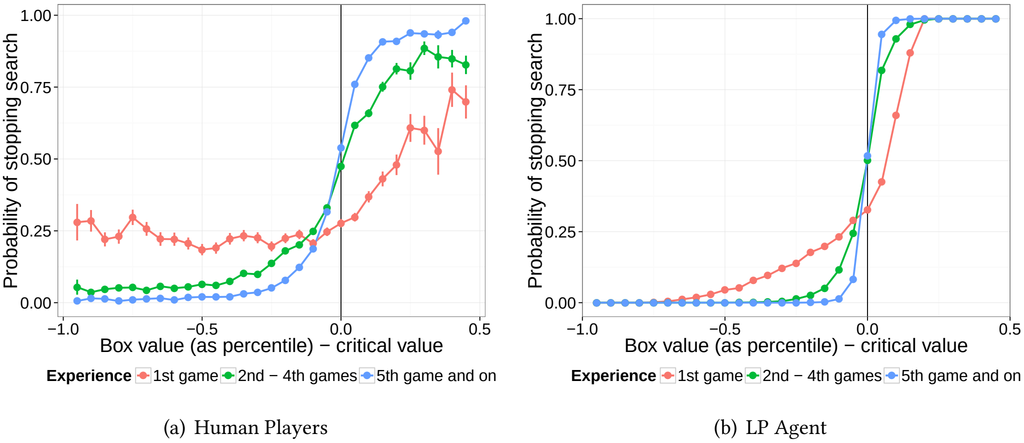


Fig. 4. Search depth for players in their first games measured by the number of boxes opened. Error bars indicate ±1 standard error



(a) Human Players

(b) LP Agent

Fig. 5. Left panel: Empirical rates of stopping search for box values above and below the critical values. Right panel: Version of 5(a) with data from simulated agents instead of human players. Only non-dominated boxes are included in this analysis.

horizontal axis shows the difference between observed box values (as percentiles) and the critical values given in Table 1. The vertical axis shows the probability of stopping search when values above or below the critical values are encountered. The data in the left panel are from human players and reflect all box-by-box decisions.

An optimal player who knows the exact percentile of any box value, as well as the critical values, would always keep searching (stop with probability 0) when encountering a value whose percentile is below the critical value. Similarly, such an optimal player would always stop searching (stop with probability 1) when encountering a value whose percentile exceeds the critical value. Together these two behaviors would lead to a step function: stopping with probability 0 to the left of the critical value and stopping with probability 1 above it.

Figure 5(a) shows that on first games (in red), players tend to both under-search (stopping about 25% of the time when below the critical value) and to over-search (stopping at a maximum of 75% of the time instead of 100% of the time when above the critical value). In a player's second through fourth games (in green) performance is much improved, and the probability of stopping search is close to the ideal .5 at the critical value. The blue curve, showing performance in later games, approaches ideal step function. To address possible selection effects in this analysis, Figure 11 in the Appendix is similar to Figure 5 except it restricts to the games of those who played a substantial number of games. Because there are fewer observations, the error bars are larger but the overall trends are the same suggesting again that these results are due to learning as opposed to selection bias.

Attaining ideal step-function performance is not realistic when learning about the distribution from experience. Comparison to the LP agent provides a baseline of how well one could ever hope to do. Figure 5(b) shows that in early games, even the LP agent both stops and continues when it should not. Failing to obey the optimal critical values is a necessary consequence of learning about a distribution from experience. Compared to the human players, however, the LP agent approaches optimality more rapidly. Furthermore, on the first game, it is less likely to make large-magnitude errors. While the human players never reach the ideal stopping rates of 0 and 1 on the first game, the LP agent does so when the observed values are sufficiently far from the critical vales.

Figure 5(a) shows that stopping decisions stay surprisingly close to optimal thresholds in aggregate. Recall that the optimal thresholds depend on how many boxes are left to be opened (see Table 1). Because early boxes are encountered more often than late ones, this analysis could be dominated by decisions on the early boxes. To address this, in what follows we estimate the threshold of each box individually.

### 4.2 Effects of unhelpful feedback

One may view winning or losing the game as a type of feedback for the player to indicate if the strategy used needs adjusting. Taking this view, consider a player's first game. Say this player over-searched in the first game, that is, they saw a value greater than the critical value but did not stop on it. Assume further that this player won this game. This player did not play the optimal strategy but won anyway, so their feedback was unhelpful. The middle panel of Figure 6(a) shows the errors made during a second game after over-searching and either winning or losing during their first game. The red curve tends to be above the blue curve, meaning that players who stopped too late but didn't get punished (blue) are less likely to stop on most box values in the next game, compared to players who stopped too late and got punished (red).

Similarly, the bottom panel shows the blue curve to be above the red curve, meaning that players who stopped too early but didn't get punished (blue) are more likely to stop on most box values in the next game, compared to players who stopped too early and got punished (red).

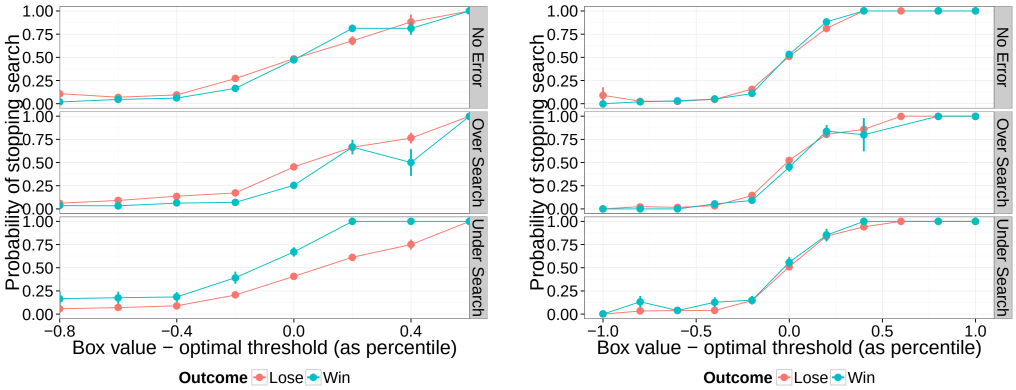
This finding makes the results in Figures 3 and 5 even more striking as it is a reminder that the participants are learning in an environment where the feedback they receive is noisy. Figure 6(b) shows the errors in the fifth game given the feedback from the first game. Even a quick glance shows that the curves are essentially on top of each other. Thus, those who received unhelpful feedback in the first game were able to recover—and perform just as well as those who received helpful feedback—by the fifth game.

## 5 MODELING PLAYER DECISIONS

In this section we explore the predictive performance of several models of human behavior in the repeated secretary problem with learnable distributions. We begin by describing our framework for evaluating predictive models. We then describe the models, and compare their performance.

### 5.1 Evaluation and comparison

Our goal in this section is to compare several psychologically plausible models in terms of how well they capture human behavior in the repeated secretary game to give us some insight as to



(a)

(b)

Fig. 6. Errors in the second (left) and fifth (right) games given whether the first game was won (red curves) or lost (blue curves). Vertically arranged panels indicate what type of error, if any, was made on the first game.

how people are learning to play the game. Since our goal is to compare how likely each model is given the data the humans generated we use a Bayesian model comparison framework. The models we compare, defined in Section 5.2, are probabilistic, allowing them to express differing degrees of confidence in any given prediction. This also allows them to capture heterogeneity between players. In contexts where players' actions are relatively homogeneous, their actions can be predicted with a high degree of confidence, whereas in contexts where players' actions differ, the model can assign probability to each action.

After opening each box, a player makes a binary decision about whether or not to stop. Our dataset consists of a set of *stopping decisions* $y_t^g \in \{0, 1\}$ that the player made in game $g$ after seeing non-dominated box $t$. If the player stopped at box $t$ in game $g$, then $y_t^g = 1$; otherwise, $y_t^g = 0$. Our dataset also contains the *history* $x_{T:t}^g = \left(x_T^g, x_{T-1}^g, \ldots, x_t^g\right)$ of box values that the player had seen until each stopping decision. We represent the full dataset by the notation $\mathcal{D}$.

In our setting, a probabilistic model $f$ maps from a history $x_{T:t}^g$ to a probability that the agent will stop. (This fully characterizes the agent's binary stopping decision.) Each model may take a vector $\theta$ of parameters as input. We assume that every decision is independent of the others, given the context. Hence, given a a model and a vector of parameters, the likelihood of our dataset is the product of the probabilities of its decisions; that is,

$$p\left(\mathcal{D} \mid h, \theta\right) = \prod_{(x_{T:t}^g, y_t^g) \in \mathcal{D}} \left[ f\left(x_{T:t}^g \,\middle|\, \theta\right) y_t^g + \left(1 - f\left(x_{T:t}^g \,\middle|\, \theta\right)\right)\left(1 - x_{T:t}^g\right)\right].$$

We compare models by how probable they are given the data. That is, we say that a model $f^1$ has better predictive performance than model $f^2$ if $p\left(f^1 \,\middle|\, \mathcal{D}\right) > p\left(f^2 \,\middle|\, \mathcal{D}\right)$, where

$$p\left(f \mid \mathcal{D}\right) = \frac{p(f) p\left(\mathcal{D} \mid f\right)}{p(\mathcal{D})}. \tag{2}$$

As we have no reason a priori to prefer any specific model, we assign them equal prior model probabilities $p(f)$. Comparing the model probabilities defined in Equation (2) is thus equivalent to comparing the models' *model evidence*, defined as

$$p(\mathcal{D} \mid f) = \int_{\Theta} p(\mathcal{D} \mid f, \theta)\, p(\theta) d\theta. \tag{3}$$

The ratio of model evidences $p\left(\mathcal{D} \,\middle|\, f^1\right) \big/ p\left(\mathcal{D} \,\middle|\, f^2\right)$ is called the *Bayes factor* [e.g., see Kruschke, 2015]. The larger the Bayes factor, the stronger the evidence in favor of $f^1$ versus $f^2$.

This probabilistic approach has several advantages. First, the Bayes factor between two models has a direct interpretation: it is the ratio of probabilities of one model's being the true generating model, conditional on one of the models under consideration being the true model. Second, it allows models to quantify the confidence of their predictions. This quantification allows us to distinguish between models that are almost correct and those that are far from correct in a way that is impossible for coarser-grained comparisons such as predictive accuracy.

Finally, the Bayes factor contains a built-in compensation for overfitting. Models with a higher dimensional parameter space are penalized, due to the fact that the integral in Equation (3) must average over a larger space. The more flexible the model, the more of this space will have low likelihood, and hence the better the fit must be in the high-probability regions in order to attain the same evidence as a lower-parameter model. In particular, this means that when one model generalizes another but has equivalent (or even insufficiently better) fit at its best-fitting parameters, the more restricted model will have a high Bayes factor relative to the generalized model. We use uninformative prior distributions for all of our parameters, which gives especially strong protection against preferring overfitted models.

The integral in Equation (3) is analytically intractable, so we followed the standard practice of approximating it using Markov chain Monte Carlo sampling. Specifically, we used the PyMC software package's implementation [Salvatier et al., 2016] of the Slice sampler [Neal, 2003] to generate 25000 samples from each posterior distribution of interest, discarding the first 5000 as a "burn in" period. We then used the Gelfand-Dey method [Gelfand and Dey, 1994] to estimate Equation (3) based on this posterior sample.[1]

## 5.2 Models

We start by defining our candidate models, each of which assumes that an agent decides at each non-dominated box whether to stop or continue, based on the history of play until that point. For notational convenience, we represent a history of play by a tuple containing the number of boxes seen $i$, the number of non-dominated boxes seen $i^*$, and the percentile of the the current box $q_i$ as estimated using Equation 1. Formally, each model is a function $f : \mathbb{N} \times \mathbb{N} \times [0,1] \to [0,1]$ that maps from a tuple $(i, i^*, q_i)$ to a probability of stopping at the current box.

*Definition 5.1 (Value Oblivious).* In the Value Oblivious model, agents do not attend to the specific box values. Instead, conditional upon reaching a non-dominated box $i$, an agent stops with a fixed probability $p_i$.

$$f^{\text{value-oblivious}}\left(i, i^*, q_i \,\middle|\, \left\{p_j\right\}_{j=1}^{T-1}\right) = p_i.$$

*Definition 5.2 (Viable k).* The Viable k model stops on the $k$th non-dominated box.

$$f^{\text{viablek}}(i, i^*, q_i \mid k, \epsilon) = \begin{cases} \epsilon & \text{if } i^* < k, \\ 1 - \epsilon & \text{otherwise.} \end{cases}$$

[1] This is nontrivial because the integral is with respect to the prior, not the posterior. However, most of the contribution to the integral's total comes from high-posterior regions of the parameter space, so simply sampling from the prior would produce a very noisy estimate.

In this model and the next agents are assumed to err with probability $\epsilon$ on any given decision.

*Definition 5.3 (Sample k).* The Sample k model stops on the first non-dominated box that it encounters after having seen at least $k$ boxes, whether those boxes were dominated or not.

$$f^{\text{sample}}(i, i^*, q_i \,|\, k, \epsilon) = \begin{cases} \epsilon & \text{if } i < k, \\ 1 - \epsilon & \text{otherwise.} \end{cases}$$

When $k = \lceil T/e \rceil$ and $\epsilon = 0$, this corresponds to the optimal solution of the classical secretary problem in which the distribution is unknown.

*Definition 5.4 (Multiple Threshold).* The Multiple Threshold model stops at box $i$ with increasing probability as the box value increases. We use a logistic specification which yields a sigmoid function at each box $i$ such that at values equal to the threshold $\tau_i$ an agent stops with probability 0.5; an agent stops with greater (less) than 0.5 probability on values higher (lower) than $\tau_i$, with the probabilities becoming more certain as the value's distance from $\tau_i$ grows. We also learn a single parameter $\lambda$ across all boxes representing how quickly the probability changes as a box value becomes further from $\tau_i$. Intuitively, $\lambda$ controls the slope of the sigmoid[2].

$$f^{\text{thresholds}}\left(i, i^*, q_i \,\middle|\, \lambda, \left\{\tau_j\right\}_{j=1}^{T-1}\right) = \frac{1}{1 + \exp[\lambda(q_i - \tau_i)]}.$$

When the thresholds are set to the critical values of Table 1 so that $\tau_i = z_i$, this model corresponds to the optimal solution of the secretary problem with a known distribution.

*Definition 5.5 (Single Threshold).* The Single Threshold model is a simplified threshold model in which agents compare box values to a single threshold $\tau$ rather than box-specific thresholds.

$$f^{\text{single-threshold}}(i, i^*, q_i \,|\, \lambda, \tau) = \frac{1}{1 + \exp[\lambda(q_i - \tau)]}.$$

*Priors.* Each of the models described above has free parameters that must be estimated from the data. We used the following uninformative prior distributions for each parameter:

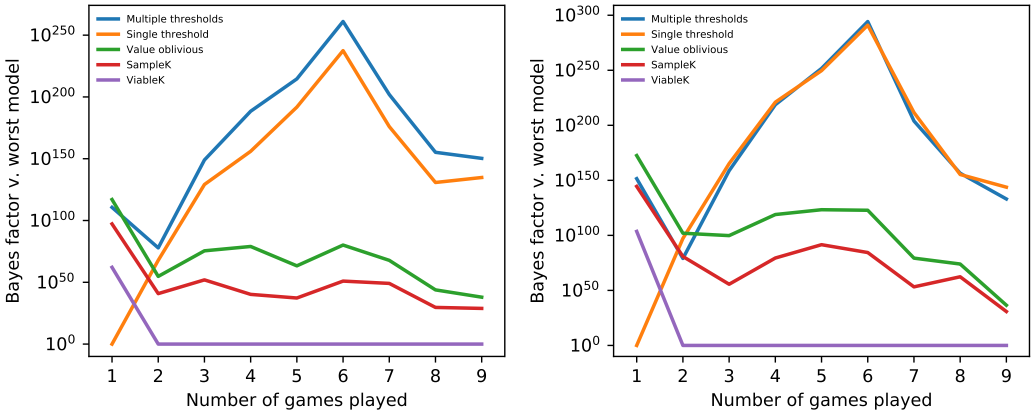
$$\begin{aligned} p_i &\sim \text{Uniform}[0, 1] & \tau, \tau_i &\sim \text{Uniform}[0, 1] \\ k &\sim \text{Uniform}\{1, 2, \ldots, T - 1\} & \lambda &\sim \text{Exponential}(\mu = 1000). \\ \epsilon &\sim \text{Uniform}[0, 0.5] \end{aligned}$$

The hyperparameter $\mu$ for precision parameters $\lambda$ was chosen manually to ensure good mixing of the sampler. Each parameter's prior is independent; e.g., in the single-threshold model a given pair $(\lambda, \tau)$ has prior probability $p(\lambda, \tau) = p(\lambda)p(\tau)$.

### 5.3 Model comparison results

Figure 7 gives the Bayes factors for each of the models of Section 5.2. The models were estimated separately for each number of games; that is, each model was estimated once on all the first games played by participants, again on all the second games, etc. This allows us to detect learning by comparing the estimated values of the parameters across games. The Bayes factor is defined as a ratio between two model evidences. Since we are instead comparing multiple models, we take the standard approach of expressing each Bayes factor with respect to the lowest-evidence model for a given number of games. These normalized Bayes factors are consistent, in the sense that if the normalized Bayes factor for model $h^1$ is $k$ times larger than the normalized Bayes factor for $h^2$, then the Bayes factor between $h^1$ and $h^2$ is $k$. As a concrete example, the Sample k model had the

[2] We considered models with one $\lambda$ per box but they did not perform appreciably better than the single $\lambda$ models.



(a) Seven Boxes

(b) Fifteen Boxes

Fig. 7. Bayes factors for various models, compared to the lowest-evidence model in each game.

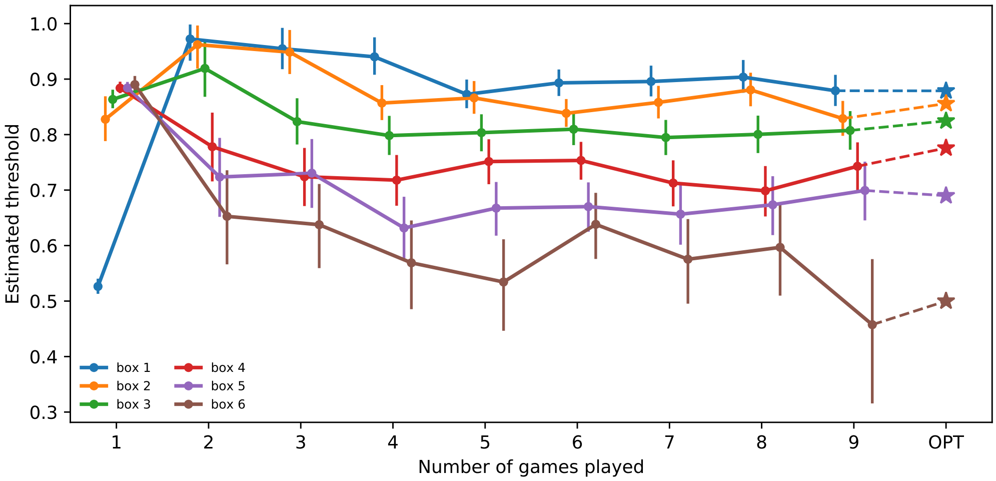


Fig. 8. Estimated thresholds in the seven box games. The rightmost set of points are the optimal thresholds. Error bars represent the 95% posterior credible interval.

lowest model evidence for participants' first games in the seven box condition; the model evidence for the Value Oblivious model was $10^{117}$ times greater than that of the Single Threshold model, and $10^{48}$ times greater than that of the Viable k model.

In first game played, in both the seven box and fifteen box conditions, the best performing model was Value Oblivious. In all subsequent games, a threshold model was the best performing model, and Viable k was the worst performing model. For the seven box condition, the multiple threshold model outperformed the single-threshold model; in the fifteen box condition, the single and multiple threshold models had approximately equivalent performance.

Evidently, players behaved consistently with the optimal class of model for the known distribution—multiple thresholds—as early as the second game. This is consistent with the observations of Section 4.1, in which players' outcomes improved with repeated play. In addition, it is consistent with the learning of optimal thresholds in Figure 5(a) but improves on that analysis because here the most common stopping points—the early boxes—do not dominate the average. Futhermore, players' estimated thresholds approached the theoretically optimal values remarkably quickly. Figure 8 shows the estimated thresholds for the seven box condition, along with their 95% posterior credible intervals. The estimated thresholds for the second and subsequent games are strictly decreasing in the number of boxes seen, like the optimal thresholds. Overall, the thresholds appear to more closely approximate their optimal values over time. After only four games, each threshold's credible interval contains the optimal threshold value.[3] Thus, workers learned to play according to the optimal family of models and learned the optimal threshold settings within that family of models.

The success of the Value Oblivious model in the first game suggests that neither of the threshold-based models fully capture players' decision making in their initial game. This is further supported by the best-estimates of thresholds for the first game: unlike subsequent games which have thresholds that strictly decrease in number of boxes seen, in the first game the estimated thresholds are strictly *increasing* in number of boxes seen. This is consistent with players using a Value Oblivious model. If players who stop on later boxes do so for reasons independent of the box's value, then they will tend to stop on higher values merely due to the selection effect from only stopping on non-dominated boxes.

In sum, the switch from increasing to decreasing thresholds in Figure 8 is consistent with moving from a value-oblivious strategy, which generalizes the optimal solution for the classical problem, to a threshold strategy, which generalizes the optimal strategy for known distributions.

## 6 CONCLUSION: BEHAVIORAL INSIGHTS

The main research question we addressed in this work is whether people improve at the secretary problem through repeated play. In contrast to prior research (Campbell and Lee [2006], Lee [2006], Seale and Rapoport [1997]), across thousands of players and tens of thousands of games, we document fast and steep learning effects. Rates of winning increase by about 25 percentage points over the course of the first ten games (Figure 3).

From the results in this article, it seems as if players not only improve, but also learn to play in a way that approaches optimality in several respects, which we list here. Rates of winning come within about five percentage points of the maximum win rates possible, and this average is taken without cleaning the data of players who were obviously not trying. In looking at box-by-box decision making, player's probabilities of stopping came to approximate an optimal step function after a handful of games (Figure 5). And similar deviations from the optimal pattern were also observed in a very idealized agent that learns from data, suggesting that some initial deviation from optimality is inevitable. Perhaps even more remarkably, they were able to do this with no prior knowledge of the distribution and, consequentially, sometimes unhelpful feedback (Figure 6).

In the first game, player behavior was relatively well fit by the Value Oblivious model which had a fixed probability of stopping at each box, independent of the values of the boxes. In later plays, threshold-based decision making—the optimal strategy for known distributions—fit the data best (Figure 7). Further analyses uncovered that players' implicit thresholds were close to the optimal critical values (Figure 8), which is surprising given the small likelihood that players actually would, or could, solve for these values.

A few points of difference could explain the apparent departure from prior empirical results. First, to our knowledge, ours is the first study to begin with an unknown distribution that players can learn over time. Seemingly small differences in instructions to participants could have a large effect. As mentioned, other studies have informed participants about the distribution, for example its minimum, maximum, and shape. Second, some prior experimental designs have presented ranks

[3] In games 5–8, either one or two credible intervals no longer contain the corresponding optimal value; by game 9 all thresholds' credible intervals again contain their optimal values.

or manipulated values that made it difficult to impossible for participants to learn the distributions. Third, past studies have used relatively few participants, making it difficult to detect learning effects. For example, Campbell and Lee [2006] have 12 to 14 participants per condition and assess learning by binning the first 40, second 40, and third 40 games played. In contrast, with over 5,000 participants, we can examine success rates at every number of games played beneath 20 with large sample sizes. This turns out to be important for testing learning, as most of it happens in the first 10 games. While our setting is different than prior ones, the change of focus seems merited because many real-world search problems (such as hiring employees in a city) involve repeated searches from learnable distributions.

A promising direction for future research would be to propose and test a unified model of search behavior that can capture several properties observed here such as: the effects on unhelpful feedback (Figure 6), the transition from value-oblivious to threshold-based decision making (Figure 7), and the learning of near-optimal thresholds (Figure 8). Having established that people learn to approximate optimal stopping in repeated searches through distributions of candidates, the next challenge is to model how individual strategies evolve with experience.

## REFERENCES

J Neil Bearden. 2006. A new secretary problem with rank-based selection and cardinal payoffs. *Journal of Mathematical Psychology* 50, 1 (2006), 58–59.

James Campbell and Michael D. Lee. 2006. The effect of feedback and financial reward on human performance solving secretaryproblems. In *Proceedings of the 28th annual conference of the cognitive science society*. 1068–1073.

Ruth M Corbin, Chester L Olson, and Mona Abbondanza. 1975. Context effects in optional stopping decisions. *Organizational Behavior and Human Performance* 14, 2 (1975), 207–216.

Thomas S Ferguson. 1989. Who solved the secretary problem? *Statistical science* (1989), 282–289.

P. R. Freeman. 1983. The secretary problem and its extensions: A review. *International Statistical Review* 51, 2 (1983), 189–206.

Alan E Gelfand and Dipak K Dey. 1994. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)* (1994), 501–514.

John P Gilbert and Frederick Mosteller. 1966. Recognizing the Maximum of a Sequence. *J. Amer. Statist. Assoc.* 61, 313 (1966), 35–73.

James P. Kahan, Amnon Rapoport, and Lyle V. Jones. 1967. Decision making in a sequential search task. *Perception & Psychophysics* 2, 8 (1967), 374–376.

John Kruschke. 2015. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2 ed.). Academic Press.

Michael D. Lee. 2006. A Hierarchical Bayesian Model of Human Decision-Making on an Optimal Stopping Problem. *Cognitive Science* 30, 3 (2006), 1–26.

Michael D. Lee, Tess A. OConnor, and Matthew B. Welsh. 2004. Decision-Making on the Full Information Secretary Problem. In *Proceedings of the 26th annual conference of the cognitive science society*. 819–824.

Radford M Neal. 2003. Slice sampling. *Annals of statistics* (2003), 705–741.

Asa B Palley and Mirko Kremer. 2014. Sequential search and learning from rank feedback: Theory and experimental evidence. *Management Science* 60, 10 (2014), 2525–2542.

Amnon Rapoport and Amos Tversky. 1970. Choice behavior in an optional stopping task. *Organizational Behavior and Human Performance* 5, 2 (1970), 105–120.

John Salvatier, Thomas V Wiecki, and Christopher Fonnesbeck. 2016. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* 2 (2016), e55.

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

Peter M Todd. 1997. Searching for the next best mate. In *Simulating social phenomena*. Springer, 419–436.

Rami Zwick, Amnon Rapoport, Alison King Chung Lo, and AV Muthukrishnan. 2003. Consumer sequential search: Not enough or too much? *Marketing Science* 22, 4 (2003), 503–519.

# A APPENDIX

## A.1 Computation of critical values and probability of winning in known distribution case

Note that

$$p_1(h) = 1 - F(h). \tag{4}$$

This is the probability of observing something on the last round that exceeds the best observation. Generally

$$\begin{aligned} p_t(h) &= \begin{cases} \int_h^{\infty} F(x)^{t-1} f(x)dx + F(h)p_{t-1}(h) & c_t \leq h \\ \int_{c_t}^{\infty} F(x)^{t-1} f(x)dx + \int_h^{c_t} p_{t-1}(x) f(x)dx + F(h)p_{t-1}(h) & c_t > h \end{cases} \\ &= \begin{cases} \frac{1-F(h)^t}{t} + F(h)p_{t-1}(h) & c_t \leq h \\ \frac{1-F(c_t)^t}{t} + \int_h^{c_t} p_{t-1}(x) f(x)dx + F(h)p_{t-1}(h) & c_t > h \end{cases} \end{aligned} \tag{5}$$

To understand (5), first note that if the critical value is less than the historically best observation, anything exceeding the historically best observation $h$ is acceptable. Thus, if something better, $x$, is observed, it is accepted, in which case the player wins if all the subsequent observations are worse, with probability $F(x)^{t-1}$. Otherwise, we inherit $h$ and have a probability of winning $p_{t-1}(h)$ in the next period.

If $c_t > h$, then an acceptance occurs only if the realization $x$ exceeds $c_t$, in which case the probability of winning remains $F(x)^{t-1}$. If the player experiences a value between $h$ and $c_t$, the historical maximum rises but is not accepted. Finally, if the observation is less than $h$, the historically best observation is not incremented and the player moves to period $t - 1$.

Let $\bar{p}_t$ be the value of $p_t$ arising when $c_t = h_t$. Then

Lemma A.1. $\bar{p}_t(h) = \sum_{j=1}^{t} \frac{F(h)^{j-1} - F(h)^t}{t+1-j}$

Proof. The proof is by induction on $t$. The lemma is trivially satisfied at $t = 1$. Suppose it is satisfied at $t - 1$. Then, from (5),

$$\begin{aligned} \bar{p}_t(h) &= \frac{1-F(h)^t}{t} + F(h)p_{t-1}(h) = \frac{1-F(h)^t}{t} + F(h)\sum_{j=1}^{t-1} \frac{F(h)^{j-1} - F(h)^{t-1}}{t-j} \\ &= \frac{1-F(h)^t}{t} + \sum_{j=1}^{t-1} \frac{F(h)^j - F(h)^t}{t-j} \\ &= \frac{1-F(h)^t}{t} + \sum_{j=2}^{t} \frac{F(h)^{j-1} - F(h)^t}{t+1-j} = \sum_{j=1}^{t} \frac{F(h)^{j-1} - F(h)^t}{t+1-j} \end{aligned}$$

□

Note that, at the value $x_t = c_t$, the searcher must be indifferent between accepting $x_t$ and rejecting, in which case the history becomes $c_t$. Therefore,

$$\left.\frac{\partial p_t}{\partial c_t}\right|_{h_t = c_t} = 0.$$

This gives

$$0 = -f(c_t)F(c_t)^{t-1} + f(c_t)\bar{p}_{t-1}(c_t),$$

or,

$$F(c_t)^{t-1} = \sum_{j=1}^{t-1} \frac{F(c_t)^{j-1} - F(c_t)^{t-1}}{t-j}. \tag{6}$$

Equation 6 is intuitive, in that it says $F(c_t)^{t-1} = \bar{p}_{t-1}(c_t)$, that is, the probability of winning given an acceptance of $c_t$, which is $F(c_t)^{t-1}$, equals the probability of winning given that $c_t$ is rejected and becomes the going-forward history, which would give a probability of winning of $\bar{p}_{t-1}(c_t)$. Thus,

$$p_t(h) = \begin{cases} \sum_{j=1}^{t} \frac{F(h)^{j-1} - F(h)^t}{t+1-j} & c_t \leq h \\ \frac{1-F(c_t)^t}{t} + \int_h^{c_t} p_{t-1}(x) f(x) dx + F(h) p_{t-1}(h) & c_t > h \end{cases}$$

Letting $i = T - t$, and $z_i = F(c_t)$ we can rewrite (6) to give

$$z_t^{t-1} = \sum_{j=1}^{t-1} \frac{z_t^{j-1} - z_t^{t-1}}{t-j}, \text{ or } z_t^t = \sum_{j=1}^{t-1} \frac{z_t^j - z_t^t}{t-j} \tag{7}$$

## A.2 Computation of probability of winning in unknown distribution case

The probability of winning for a fixed value of $k$ and $T$ periods is
Thus,

$$1 = \sum_{j=1}^{t-1} \frac{z_t^{j-t} - 1}{t-j} = \sum_{i=1}^{t-1} \frac{z_t^{-i} - 1}{i} \tag{8}$$

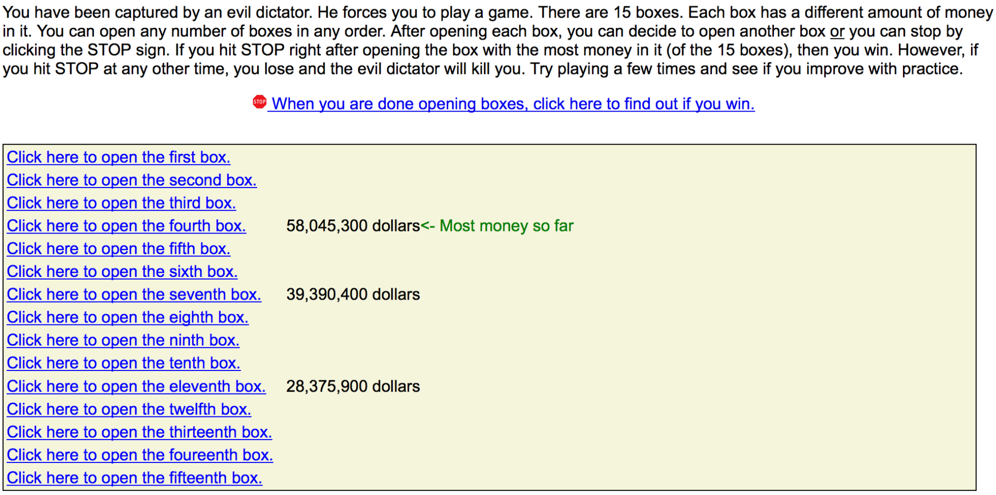
$$p_t(h) = \begin{cases} \sum_{j=1}^{t} \frac{h^{j-1} - h^t}{t+1-j} & z_t \leq h \\ \frac{1-z_t^t}{t} + \int_h^{z_t} p_{t-1}(x) dx + h p_{t-1}(h) & z_t > h \end{cases} \tag{9}$$

$$\int_0^\infty kF(x)^{k-1} f(x) \sum_{m=k+1}^{T} F(x)^{m-k-1} \int_x^\infty f(y) F(y)^{T-m} dy dx$$

$$= \sum_{m=k+1}^{T} \int_0^1 kx^{m-2} \int_x^1 y^{T-m} dy dx = \sum_{m=k+1}^{T} \int_0^1 kx^{m-2} \frac{1 - x^{T-m+1}}{T-m+1} dx$$

$$= \sum_{m=k+1}^{T} \frac{k}{T-m+1} \int_0^1 x^{m-2} - x^{T-1} dx = \sum_{m=k+1}^{T} \frac{k}{T-m+1} \left( \frac{1}{m-1} - \frac{1}{T} \right) = \frac{k}{T} \sum_{m=k+1}^{T} \frac{1}{m-1}$$

## A.3 Appendix Figures



You have been captured by an evil dictator. He forces you to play a game. There are 15 boxes. Each box has a different amount of money in it. You can open any number of boxes in any order. After opening each box, you can decide to open another box or you can stop by clicking the STOP sign. If you hit STOP right after opening the box with the most money in it (of the 15 boxes), then you win. However, if you hit STOP at any other time, you lose and the evil dictator will kill you. Try playing a few times and see if you improve with practice.

When you are done opening boxes, click here to find out if you win.

Click here to open the first box.
Click here to open the second box.
Click here to open the third box.
Click here to open the fourth box. 58,045,300 dollars<- Most money so far
Click here to open the fifth box.
Click here to open the sixth box.
Click here to open the seventh box. 39,390,400 dollars
Click here to open the eighth box.
Click here to open the ninth box.
Click here to open the tenth box.
Click here to open the eleventh box. 28,375,900 dollars
Click here to open the twelfth box.
Click here to open the thirteenth box.
Click here to open the foureenth box.
Click here to open the fifteenth box.

Fig. 9. Screenshot of the a 15 box treatment with 3 boxes opened.

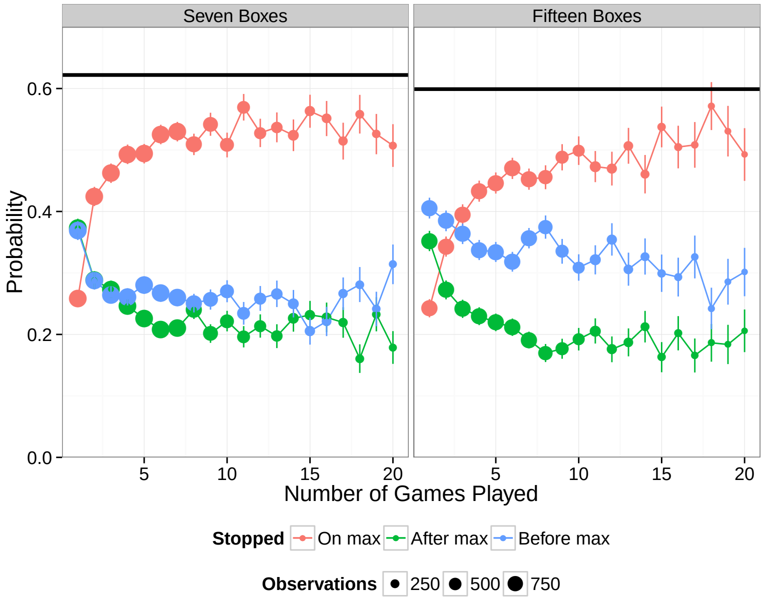


Fig. 10. Rates of winning the game for human players where each player played at least 7 games. Error bars indicated ±1 standard error, when they are not visible they are smaller than the points. The area of each point is proportional to the number of players in the average.

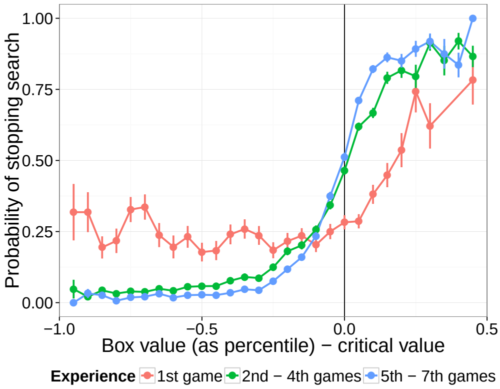


Fig. 11. In Figure 5(a) different players contribute to different curves. For example, a player who only played one time would only contribute to the red curve, while someone who played 10 times would contribute to all three curves. To address these selection effects, in this plot, we restrict to the first 7 games of those who played at least 7 games.