http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–20
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Learning When to Stop Searching

Daniel G. Goldstein,[a] R. Preston McAfee,[b] Siddharth Suri,[a] James R. Wright[c]

[a] Microsoft Research, New York, New York 10011; [b] Microsoft Corporation, Redmond, Washington 98052; [c] Department of Computing Science, University of Alberta, Edmonton, Alberta T6G 2E8, Canada

Contact: dgg@microsoft.com, https://orcid.org/0000-0002-0970-5598 (DGG); preston@mcafee.cc, https://orcid.org/0000-0002-2587-3161 (RPM); suri@microsoft.com, https://orcid.org/0000-0002-1318-8140 (SS); james.wright@ualberta.ca, https://orcid.org/0000-0001-9622-5842 (JRW)

Received: August 4, 2017
Revised: April 20, 2018
Accepted: August 21, 2018
Published Online in Articles in Advance: August 7, 2019

https://doi.org/10.1287/mnsc.2018.3245

Abstract. 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 settings, we conduct a large-scale behavioral experiment in which people search repeatedly from fixed distributions in a "repeated secretary problem." 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. Fitting such a threshold-based model to data reveals players' estimated thresholds to be close to the optimal thresholds after only a small number of games.

History: Accepted by Yuval Rottenstreich, judgment and decision making.

Keywords: Bayesian model comparison • experiments • human behavior • learning • secretary problem

1. Introduction

How do people learn when to stop searching through candidates, balancing the risks of making a choice before or after the best one has been seen? In this article, we investigate this question with an empirical analysis of the secretary problem, a formal game in which an agent evaluates candidates one at a time in search of the best one, making an accept or reject decision after each evaluation. Only one candidate can be accepted, and after being rejected, a candidate can never be recalled. It is called the secretary problem because it resembles a hiring process in which secretaries are interviewed serially and, if rejected by one employer, are quickly hired by another.

Since its appearance in the mid-20th century, the secretary problem has enjoyed exceptional popularity (Freeman 1983). It is the prototypical optimal stopping problem, attracting so much interest from so many fields that one review article concluded that it "constitutes a 'field' of study" (Ferguson 1989, p. 282). In this century, analyses, extensions, and tests of the secretary problem have appeared in decision science, operations research, computer science, economics, statistics, and psychology as well as in the pages of this journal (Bearden et al. 2006, Palley and Kremer 2014, Alpern and Baston 2017).

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 that 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 proportion $1/e$ of the candidates and then stop after the next candidate that exceeds it. Interestingly, this $1/e$ stopping rule wins about $1/e$ of the time in the limit (Gilbert and Mosteller 1966). This curious solution to the secretary problem only holds when the decision maker has *no information* about the distribution from which the values in the candidates are drawn (e.g., Mahdian et al. 2008): that is, when only the rankings of values are revealed rather than the values themselves. The optimal solution when the decision maker has full information about the distribution is qualitatively different. However, 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 that a manager hires someone, she may have only a vague guess as to the quality of the candidates who will come through the door. By the 50th hire, however, she will 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.

Although people seemingly should be able improve at the secretary problem with experience, prior academic research surprisingly does not find evidence that they do. For example, Campbell and Lee (2006, p. 1068) attempted to get participants to learn in the full information condition by offering enriched feedback and even financial rewards in a repeated secretary problem, but they concluded that "there is no evidence people learn to perform better in any condition." Similarly, Lee (2006) and Guan and Lee (2017) found no evidence of learning in the full information version of the problem, and Seale and Rapoport (1997) found no evidence of learning in the no information (ranks-only) version.

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 coming close to the optimal win rate. Then, we show that the improved win rates are due to players learning to make better decisions on a candidate-by-candidate basis and not just due to aggregating over candidates. Furthermore, we will show that the learning that we observe occurs in an environment in which the feedback can be unhelpful, pointing players in the wrong direction with respect to the optimal strategy.

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

Although 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 those most similar to our investigation. Ferguson (1989, p. 282) 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, with each order being equally likely.
4. It is assumed that you can rank all of 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 who is 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 that 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.

Other 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, whereas we found 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 with 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.

Although prior investigations, listed below, 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 et al. (2004), Campbell and Lee (2006), and Lee (2006) conducted experiments in which participants were shown values one at a time and told to try to stop at the maximum. Across these papers, the number of candidates ranged from 5 to 50, and participants played from 40 to 120 times each. In all of these studies, participants knew that the values were drawn from a uniform distribution between 0 and 100. For instance, Lee (2006, p. 5) 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. Because participants were told about the distribution, these experiments do not involve learning the 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. In a study where participants played secretary problem games on five values drawn from either a left-skewed or a right-skewed distribution, Guan and Lee (2017) provided information about the distribution by requiring participants to first play eight "practice problems" before beginning the main part of the experiment. These investigations are similar to our study 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 our study 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 given 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 et al. (2004), Campbell and Lee (2006), and Lee (2006), 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 a left-skewed, right-skewed, or uniform distribution. In this study as well as our study, 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.

Other empirical studies have investigated the process by which people update their choices of strategy based on feedback in settings other than the secretary problem. Rieskamp and Otto (2006) propose a reinforcement learning-based framework in which participants choose from a strategy repertoire based on the strategies' historical payoffs. Worthy and Maddox (2014) compare a reinforcement learning-based framework with a "win-stay, lose-shift" model as well as a hybrid model. In this study, we focus on modeling the within-game strategies chosen by participants rather than the process by which those strategies evolve in response to feedback; however, modeling the feedback-response process is an important direction for future work.

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 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. We focus on understanding players' behavior within games, leaving the question of modeling the evolution in individuals' strategies for future work.

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 on a prominent blog, a newsletter of a behavioral economics consultancy, and the website of one of the authors. This type of data collection has the advantages of being inexpensive, leading to very large samples, and recruiting a more diverse population than the standard population of university undergraduates in an experimental economics laboratory (Rubinstein 2013). The experiment did not involve monetary incentives. Like Rubinstein (2013), who wrote about similar experiments that "the behavioral results are, in my judgment, not qualitatively different from those obtained by more conventional methods," we feel that the pros (collecting more data on a fixed budget and eliminating the transaction costs of having each participant register and provide payment credentials) outweigh the cons (the lack of a financial incentive to try). If incentives improve performance on this task, the results from this unincentivized setting might be taken as a lower bound for how well a similar population should be able to learn to play the repeated secretary problem.

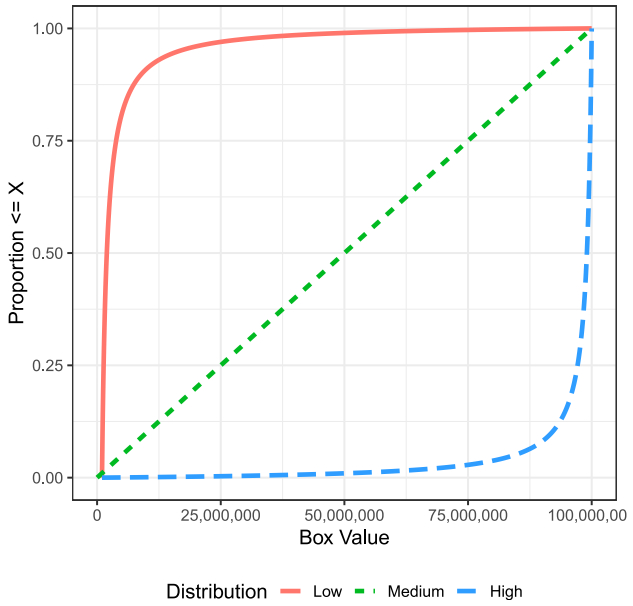
The experiment attracted 6,537 players who played the game at least one time. A total of 48,336 games were played on the site. As users arrived at the game's landing page, they were cookied, and their browser URL was automatically modified to include an identifier. These two steps were taken to assign all plays on the same browser to the same arbitrary user identifier string 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 noncookied user entering via a modified URL) was excluded from analysis and is not counted in the 6,537 that 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 who were randomly assigned to conditions by the website. As a precaution against cheating, we decided to exclude any player whose success rate was so high as to have a less than 1 in 10,000 chance of occurring under optimal play. One player won 46 of 54 games and was excluded with this criterion, although given over 6,000 players, it is not impossible that this player was both skillful and lucky. 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.

The secretary problem was lightly disguised as the "evil dictator game" to somewhat lessen the chances that a respondent would search for the problem online and discover the classical solution. The boxes correspond to the candidates, and the amount of money in each box represents the candidate's value.

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." Underneath 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, the corresponding box value was presented 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 A.1 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 of 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 number of observations submitted per person, players were told, "Please play at least six times so we can calculate your stats."

Figure 1. (Color online) Cumulative Distribution Functions of the Three Distributions from Which Box Values Were Randomly Drawn in the Experiment



Note. For probability density functions, the low distribution is strongly positively skewed (solid line), the medium distribution is a uniform distribution (dotted line), and the high distribution is strongly negatively skewed (dashed line).

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 via a browser cookie to these randomly assigned conditions so that their immediate repeat plays, if any, would be in the same conditions.

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, although 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 from the next highest value. For instance, among 15 boxes drawn from this distribution, the highest box value is, on average, about $14.5 million 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 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. Note that players only need to attend to the percentiles of the distribution to make optimal stopping decisions; the details of the distribution beyond the percentiles are irrelevant to *optimal* play. However, different distributions could potentially lead to differences in *actual human* play. Hence, varying the distribution presented to participants 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. Although one would think that 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 two percentage points, which will be shown. As 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, 1,145, 1,082, and 1,103 participants were randomly assigned to the low, medium, and high distributions, respectively, and in the 15-box condition, the counts were 1,065, 1,127, and 1,015, respectively. The number of subjects assigned to the different conditions was not significantly different whether comparing the two box conditions (7 or 15), the three distributions of box values (low, medium, or high), or all six cells of the experiment by chi-squared tests (all p-values were > 0.05).

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. Recall that the player sequentially opens boxes with values that are drawn independently from a common distribution F. The players win only if they select the highest value out of all of the boxes: opened or unopened. The problem is nontrivial, because they are forced to make a stopping decision for each box without knowing the contents of the as-yet unopened boxes.

Table 1. Critical Values and Probability of Winning Given a Known Distribution of Values for Up to 15 Boxes

Boxes left	Critical percentiles	Pr(win)
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

Optimal players will adopt a threshold rule: accept the current value if it is greater than a critical value (Gilbert and Mosteller 1966). It is a dominant strategy to reject any value worse than the best value previously observed.[1] With a known distribution independently distributed across boxes, the critical dollar value will be the maximum of the historically best value and a critical value that does not depend on the history. Gilbert and Mosteller (1966) derived the critical values for the uniform distribution over $[0, 1]$. Because the percentiles of an arbitrary distribution are uniformly distributed on $[0, 1]$ by definition, this same analysis gives the critical *percentile* values for all distributions listed in Table 1.[2]

The relevant entries for our study are the games of 7 and 15 boxes. These calculations show that experienced players who know the distribution can hope to win at best 62.2% of the games for 7-box games and just under 60% of the time for 15-box games. Note that these numbers compare favorably with the usual secretary results, which are less for all game lengths, converging to the famous $1/e$, about 37%, as the number of boxes increases. Thus, there is substantial value in knowing the distribution.

This favorable comparison holds for the specific game lengths that we consider as well as in the limit. 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 boxes as a critical value. Optimizing over k yields the probabilities of winning for given game lengths given in 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.

Table 2. The Probability of Winning a Game in the Classical Secretary Problem (Unknown Distribution of Values) for Up to 15 Boxes

Game length (boxes)	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

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 percentile values from Table 1 and learns the percentiles of the distribution from experience; it will be referred to as the learn percentiles (LP) 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 that 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 (i.e., the critical values given by the first 7 or 15 rows of Table 1). The agent does not yet know the critical value as raw box(dollar) values, because the distribution is unknown before the first play. Armed with these critical values, the LP agent converts the box values that it observes into percentiles to compare them with the critical values. The first box value that the LP agent sees gets assigned an estimated percentile of 0.50. If the second observed box value is greater than the first, it estimates the second value's percentile to be 0.75 and reestimates the first value's percentile to be 0.25. If the second value is smaller than the first, it assigns the estimate of 0.25 to the second value and 0.75 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 that 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 with 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 that 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 with 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 achieves the theoretical maxima (horizontal lines in Figure 2) in seven or fewer games. In any given game, a player can either stop when it sees the maximum value, in which case it wins, or 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, and therefore, the steep gain in one implies a steep drop in 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.

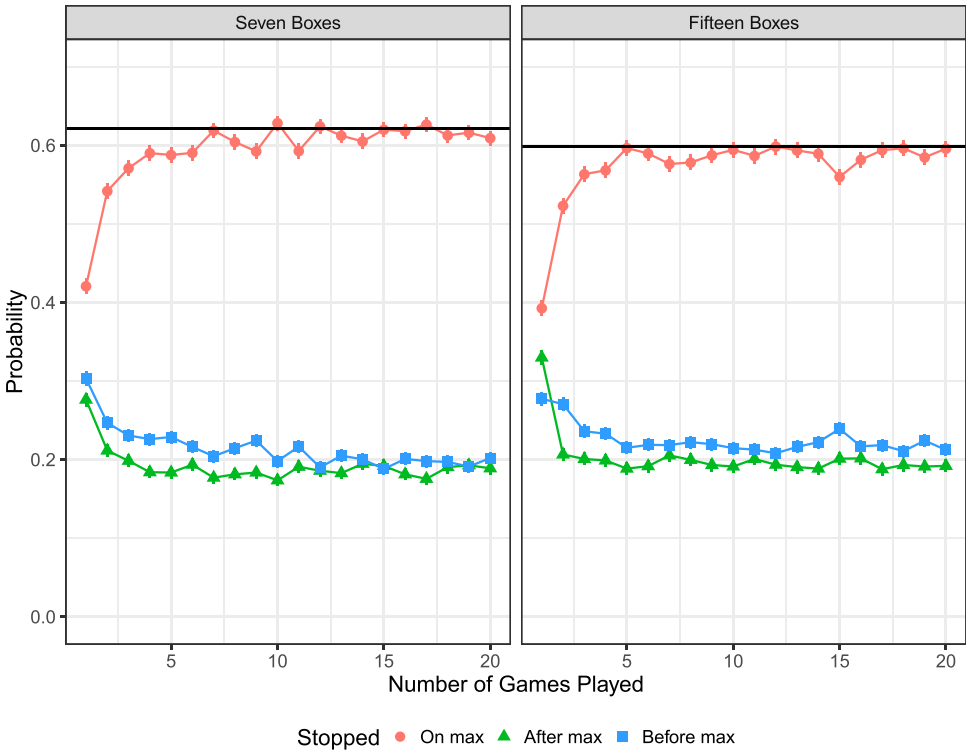
The LP agent serves as strong benchmark against which human performance can be compared. It is useful to study its performance in simulation, because the existing literature provides optimal win rates for many variations of the secretary problem but is silent on how well an idealized[3] agent would do 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 with those of the human players in the experiment.

4. Behavioral Results: Learning Effects

Because 48,336 games were played by 6,537 users, the average user played 7.39 games. Roughly one-half (49.6%) of users played five games or more, and one-quarter (23.2%) played nine games or more. One-tenth (9.3%) played 16 games or more.

Prior research (e.g., Lee 2006) has found no evidence of learning in repeated secretary problems with

Figure 2. (Color online) Rates of Winning, Stopping Too Soon, and Stopping Too Late for the LP Agent



Note. The theoretically maximal win rates for 7 and 15 boxes are given by the solid black lines.

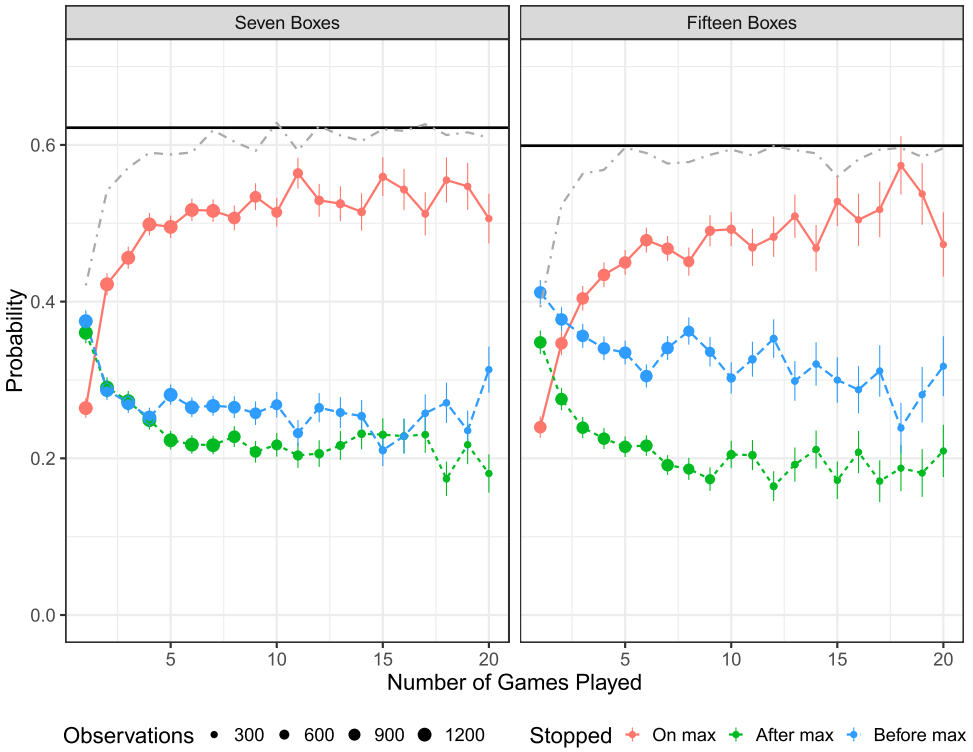
known distributions. What happens with unknown but learnable distribution? As shown in Figure 3, players rapidly improve in their first games and come within 5–10 percentage points of theoretically maximal levels of performance. The leftmost point on each solid curve in Figure 3 indicates how often first games are won. The next point to the right represents second games and so on. The solid horizontal lines in Figure 3 at 0.622 and 0.599 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 in Figure 3.

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 0.41) or 15 boxes (optimal win rate 0.39). Thus, some of the learning has to do with starting from a low base. However, the classic version's optima are reached by about the second game, and improvement continues another 10–15 percentage points beyond the classic optima.

One could argue that the apparent learning that we observe is not learning at all but a selection effect. By this logic, a common cause (e.g., higher intelligence) is responsible for players both persisting longer at the game and winning more often. To check this, we created Figure A.2 in the appendix, which is a similar plot except that it restricts to players who played at least seven games. Because we see very similar results with and without this restriction, we conclude that Figure 3 reflects mostly learning effects.

Recall that our experiment had a 2×3 design with either 7 or 15 boxes and one of three possible underlying distributions of the box values. Figure 3 shows the average probability of our subjects winning in the 7- and 15-box treatments aggregated over the three different distributions of box values. Figure 4 shows the probability of people winning in each of the six treatments of our experiment. Observe first that the probability of winning, indicated by the circles in Figure 4, increases toward the maximal win rate in each of the six treatments. In all six treatments, most of the learning happens in the early games, with diminishing returns to playing more games. This qualitative similarity shows the robustness of the finding that there is rapid and substantial learning in just a few repeated plays of the secretary problem. By comparing the probability of winning across all six treatments, one can see that the low and medium distributions were about equally as difficult and that the

Figure 3. (Color online) Rates of Winning the Game and Committing Errors for Players with Varying Levels of Experience



Notes. 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, because less than 1% of games played were beyond a user's 20th game. Solid lines indicate stopping on the maximum, long dashes indicate stopping before the maximum, and short dashes indicate stopping after the maximum. The dashed and dotted gray lines are the rate of winning the game for the LP agent. The solid black horizontal lines indicate the maximal win rate for an agent with perfect knowledge of the distribution.

high distribution was the hardest, because it had the lowest probability of winning. To understand this, we next examine the types of errors that the participants made. The probability of stopping after the maximum box value, indicated by the triangles in Figure 4, is fairly similar across all six treatments. There is, however, variation in the probability of stopping before the maximum, indicated by the squares in Figure 4, across the six treatments. Participants were more likely to stop before the maximum box in the high condition than in the others, which explains why subjects performed worse in this treatment than in the others. Because the overall qualitative trends are fairly similar across the three different distributions of box values, we will aggregate over them in the analyses that follow.

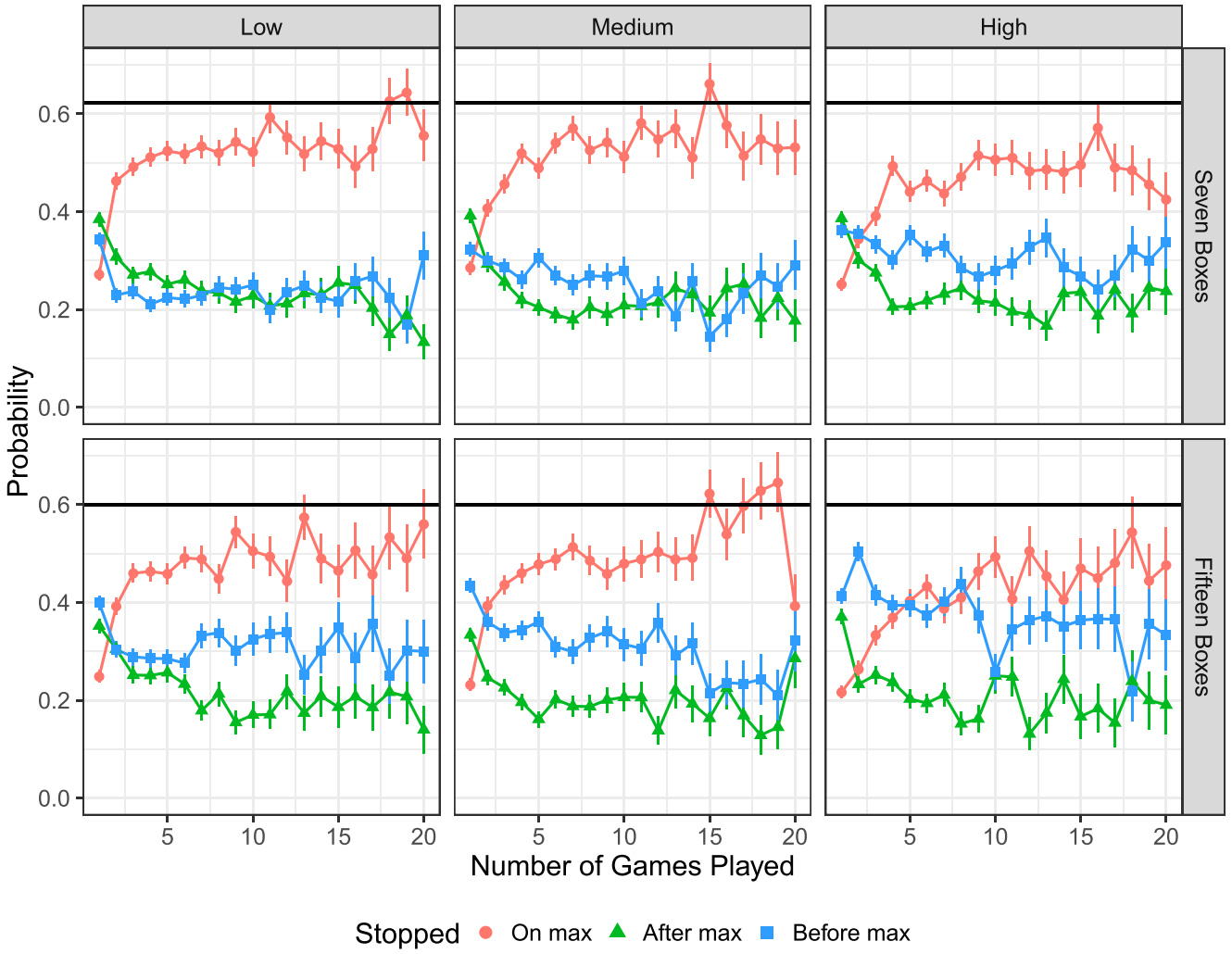
Having established that players' behavior changes with experience, we turn our attention to characterizing those changes. One overarching trend is that, soon after their first game, people search less. As seen in Figure 5, in the first eight games, the average depth of search decreases by about one-quarter to one-third of one box. Players can lose by stopping too early or too late. These search depth results suggest that 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 Figures 3 and 4. In both Figures 3 and 4, 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 searching by comparing the percentile of an observed box value with 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 searching. In Figure 6, the 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 the search when values above or below the critical values are encountered. The data in Figure 6(a) 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 zero) when encountering a value with a percentile that is

Figure 4. (Color online) Rates of Winning the Game and Committing Errors for Players with Varying Levels of Experience



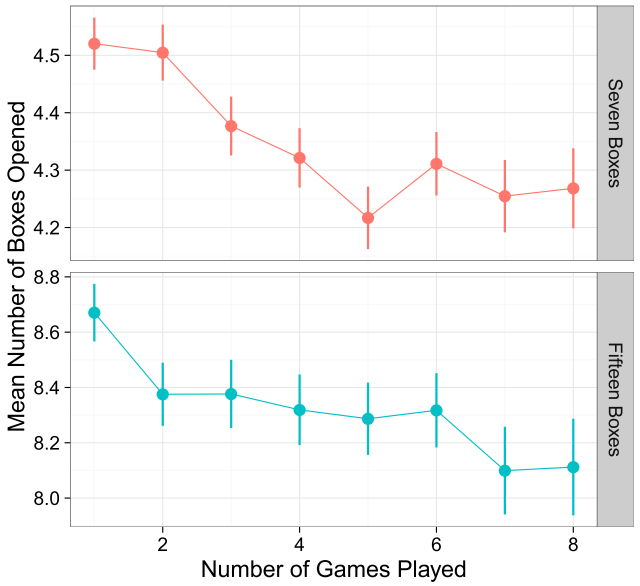
Notes. Error bars indicate ±1 standard error; when they are not visible, they are smaller than the points. The solid black horizontal lines indicate the maximal win rate for an agent with perfect knowledge of the distribution.

below the critical value. Similarly, such an optimal player would always stop searching (stop with probability one) when encountering a value with a percentile that exceeds the critical value. Together, these two behaviors would lead to a step function: stopping with probability zero to the left of the critical value and stopping with probability one above it.

Figure 6(a) shows that, on first games (denoted by circles), players tend to both undersearch (stopping about 25% of the time when below the critical value) and oversearch (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 (triangles in Figure 6(a)), performance is much improved, and the probability of stopping the search is close to the ideal 0.5 at the critical value. The squares in Figure 6(a), showing performance in later games, approaches ideal step function. To address possible selection effects in this analysis, Figure A.3 in the appendix is similar to Figure 6, except that 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 primarily because of learning.

Attaining ideal step function performance is not realistic when learning about the distribution from experience. Comparison with the LP agent provides a baseline of how well one could ever hope to do. Figure 6(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 may be a necessary consequence of learning about a distribution from experience. Compared with the human players, however, the LP agent approaches optimality more rapidly. Furthermore, in the first game, it is less likely to make large-magnitude

Figure 5. (Color online) Search Depth for Players in Their First Games Measured by the Number of Boxes Opened



Note. Error bars indicate ±1 standard error.

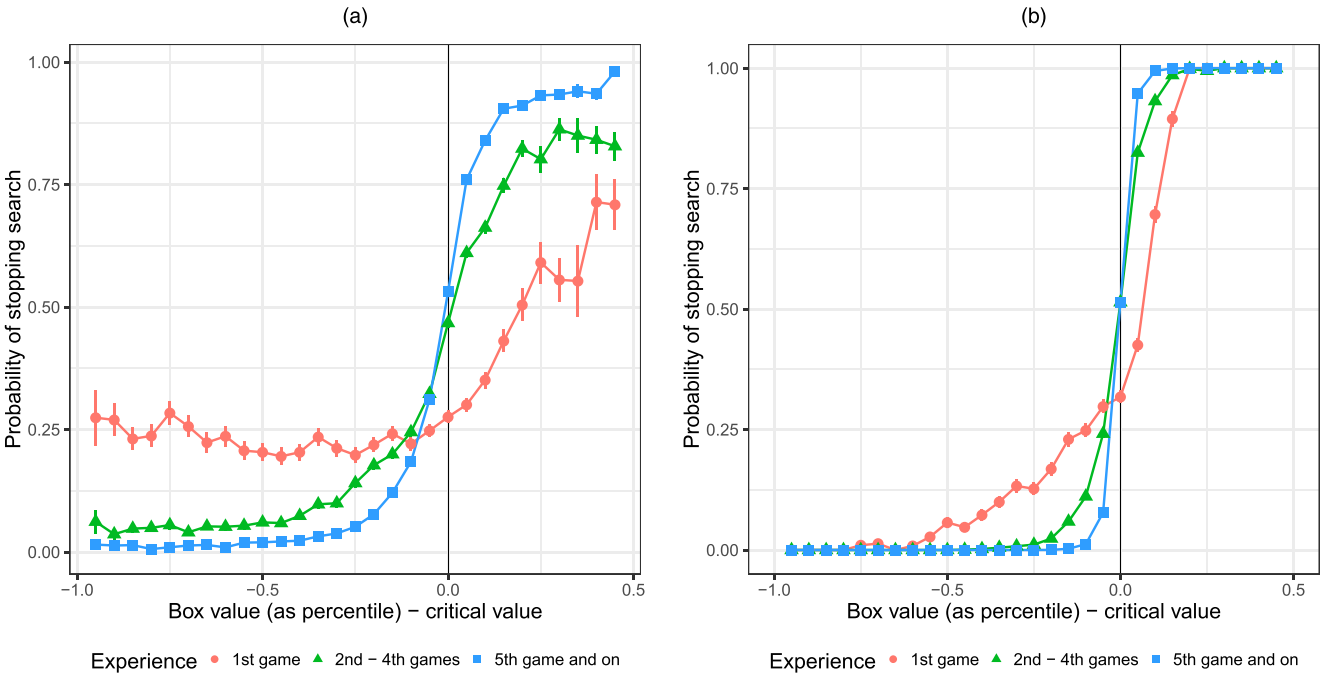
errors. Although the human players never reach the ideal stopping rates of zero and one in the first game, the LP agent does so when the observed values are sufficiently far from the critical vales.

Figure 6(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 that this player oversearched 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, and therefore, their feedback was unhelpful. The middle panel of Figure 7(a) shows the errors made during a second game after oversearching and either winning or losing during their first game. The circles in the middle panel of Figure 7(a) tend to be above the triangles, meaning that players who stopped too late but did not get punished (triangles) are less likely to stop on most box values in the next game compared with players who stopped too late and got punished (circles).

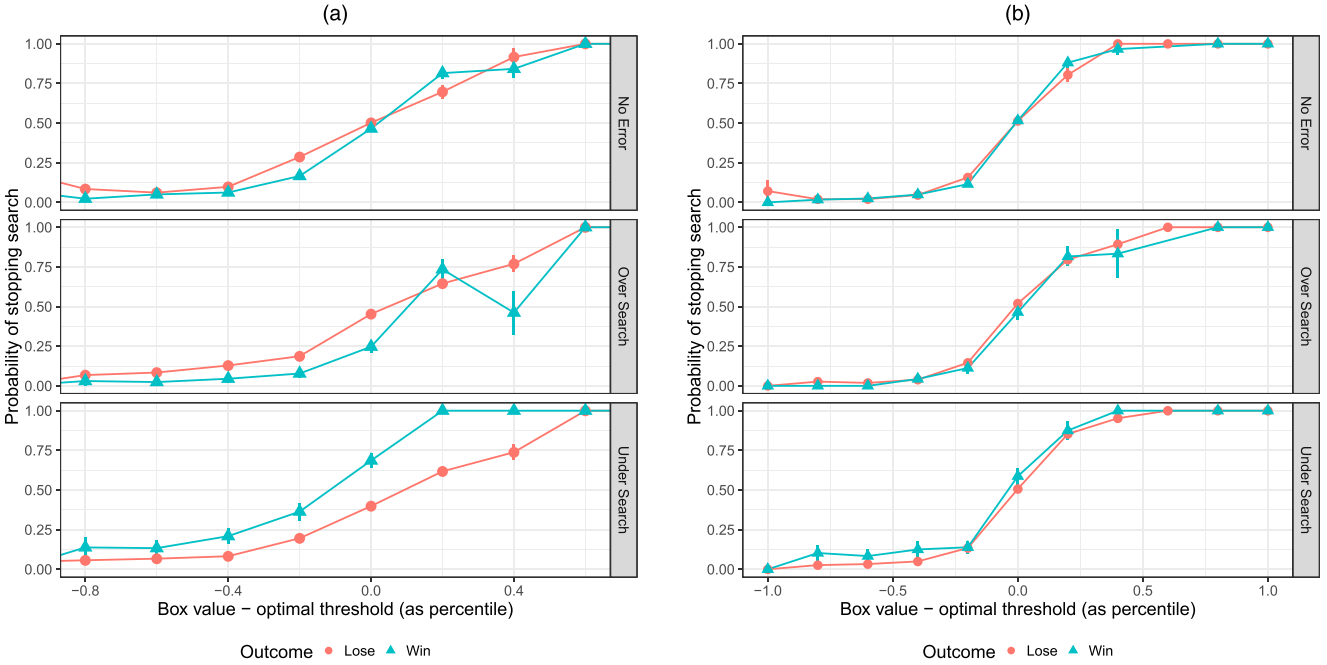
Figure 6. (Color online) (a) Empirical Rates of Stopping the Search for Box Values Above and Below the Critical Values; (b) a Version of Panel (a) with Data from Simulated Agents Instead of Human Players



Human Players LP Agent

Note. Only nondominated boxes are included in this analysis.

Figure 7. (Color online) Errors in the Second (a) and Fifth (b) Games Conditional on Whether the First Game Was Won or Lost



Notes. Vertically arranged panels indicate what type of error, if any, was made on the first game. The fifth game was the first game where the differences between the distributions of behavior between players who initially won and those who initially lost were not significant at any standard level for any of the three types of error.

Similarly, the bottom panel in Figure 7(a) shows the triangles to be above the circles meaning that players who stopped too early but did not get punished (triangles) are more likely to stop on most box values in the next game compared with players who stopped too early and got punished (circles).

This finding makes the results in Figures 3 and 6 even more striking, because it is a reminder that the participants are learning in an environment where the feedback is often misleading. Figure 7(b) shows the errors in the fifth game given the feedback from the first game.[4] 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, then describe the models, and finally, compare their performance.

5.1. Evaluation and Comparison

Our goal in this section is to compare several models in terms of how well they capture human behavior in the repeated secretary problem to give us some insight into to how people are learning to play the game. Because 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 that 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 to stop. Our data set consists of a set of *stopping decisions* $y_t^g \in \{0,1\}$ that the player made in game g after seeing nondominated box t. If the player stopped at box t in game g, then $y_t^g = 1$; otherwise, $y_t^g = 0$. Our data set also contains the *history* $x_{T:t}^g = (x_T^g, x_{T-1}^g, \ldots, x_t^g)$ of box values that the player had seen until each stopping decision. We represent the full data set 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 θ of parameters as input. We assume that every decision is independent of the others given the context. Hence,

given a model and a vector of parameters, the likelihood of our data set is the product of the probabilities of its decisions: that is,

$$p(\mathcal{D}|h,\theta) = \prod_{(x^g_{T:t},y^g_t)\in\mathcal{D}} \left[f(x^g_{T:t}|\theta)y^g_t + (1-f(x^g_{T:t}|\theta))(1-x^g_{T:t})\right].$$

In Bayesian model comparison, models are compared by how probable they are given the data. That is, a model f^1 is said to have better predictive performance than model f^2 if $p(f^1|\mathcal{D}) > p(f^2|\mathcal{D})$, where

$$p(f|\mathcal{D}) = \frac{p(f)p(\mathcal{D}|f)}{p(\mathcal{D})}. \tag{2}$$

With no a priori reason to prefer any specific model, we can 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*, which is defined as

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

The ratio of model evidences $p(\mathcal{D}|f^1)/p(\mathcal{D}|f^2)$ is called the *Bayes factor* (e.g., see Jeffreys 1935, 1961; Kass and Raftery 1995; 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 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.

One additional advantage of the Bayes factor is that it compensates for overfitting. Models with a higher-dimensional parameter space are penalized because of 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 to attain the same evidence as a lower-parameter model.

The amount by which high-dimensional models are penalized by the Bayes factor depends strongly on the choice of prior. The standard Bayesian view is that a model consists of both a prior and a likelihood; in this view, the dependence of the overfitting penalty on the prior is unproblematic, because the prior is part of the model. An alternative view is that likelihood is the model, whereas the choice of prior is relatively arbitrary. In this view, the choice of prior constitutes a "researcher degree of freedom," the influence of which should be minimized. One way to minimize the impact of the choice of prior is to evaluate models using *cross-validation*, in which the data are split into a *training set* that is used to set the parameters of the model and a *test set* that is used to evaluate the model's performance.

We use a hybrid of the cross-validation and Bayesian approaches. We first randomly select a split $s = (\mathcal{D}^{\text{train}}, \mathcal{D}^{\text{test}})$, with $\mathcal{D}^{\text{train}} \cup \mathcal{D}^{\text{test}} = \mathcal{D}$ and $\mathcal{D}^{\text{train}} \cap \mathcal{D}^{\text{train}} = \emptyset$. We then compute the *cross-validated model evidence* of the test set $\mathcal{D}^{\text{test}}$ with respect to the prior updated by the training set $p(\theta|\mathcal{D}^{\text{train}})$ rather than computing the model evidence of the full data set $\mathcal{D}$ with respect to the prior $p(\theta)$. To reduce variance owing to the randomness introduced by the random split, we take expectation over the split, yielding

$$\mathbb{E}_s p(\mathcal{D}^{\text{test}}|f,\mathcal{D}^{\text{train}}) = \int \left[\int_\Theta p(\mathcal{D}^{\text{test}}|f,\theta)p(\theta|\mathcal{D}^{\text{train}})d\theta\right] \cdot p(s)ds, \tag{4}$$

where $p(s)$ is a uniform distribution over all splits. The ratio of cross-validated model evidences

$$\frac{\mathbb{E}_s p(\mathcal{D}^{\text{test}}|f^1,\mathcal{D}^{\text{train}})}{\mathbb{E}_s p(\mathcal{D}^{\text{test}}|f^2,\mathcal{D}^{\text{train}})} \tag{5}$$

is called the *cross-validated Bayes factor*. As with the Bayes factor, larger values of (5) indicate stronger evidence in favor of f^1 versus f^2.

The integral in Equation (4) is analytically intractable, and therefore, 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 25,000 samples from each posterior distribution of interest, discarding the first 5,000 as a "burn in" period. We then used the "bronze estimator" of Alqallaf and Gustafson (2001) to estimate Equation (4) based on this posterior sample.

5.2. Models

We start by defining our candidate models, each of which assumes that an agent decides at each nondominated 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 nondominated boxes seen i^*, and the percentile of 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 1 (Value Oblivious). In the value oblivious model, agents do not attend to the specific box values. Instead, conditional on reaching a nondominated box i, an agent stops with a fixed probability p_i:

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

Definition 2 (Viable k). The viable k model stops on the kth nondominated box:

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

In this model and the next, agents are assumed to err with probability ϵ on any given decision.

Definition 3 (Sample k). The sample k model stops on the first nondominated box that it encounters after having seen at least k boxes, regardless of 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 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 τ_i, an agent stops with probability 0.5; an agent stops with greater (less) than 0.5 probability on values higher (lower) than τ_i, with the probabilities becoming more certain as the value's distance from τ_i grows. An additional parameter, λ, controls the sensitivity of the estimated probabilities to the distance from τ_i. Unlike the thresholds τ_i, we estimate a single value of λ that is shared across all boxes. Intuitively, λ controls the slope of the sigmoid[5]:

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

When the thresholds are set to the critical values of Table 1 such that $\tau_i = z_{(T-i+1)}$ and λ is set sufficiently high, this model corresponds to the optimal solution of the secretary problem with a known distribution.[6]

Definition 5 (Single Threshold). The single-threshold model is a simplified threshold model in which agents compare all box values with a single threshold τ rather than box-specific thresholds:

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

Definition 6 (Two Threshold). The two-threshold model is another simplified threshold model in which agents compare all "early" box values (the first $\lfloor T/2 \rfloor$ boxes) with one threshold (τ_0) and all "late" box values with another (τ_1):

$$f^{\text{two-threshold}}(i, i^*, q_i \,|\, \lambda, \tau_0, \tau_1) = \frac{1}{1 + \exp[\lambda(q_i - \tau_v)]},$$

where

$$\tau_v = \begin{cases} \tau_0 & \text{if } i < \frac{T}{2}, \\ \tau_1 & \text{otherwise.} \end{cases}$$

This is a low-parameter specification that nevertheless allows for differing behavior as a game progresses (unlike single threshold).

5.2.1. 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:

$$p_i \sim \text{Uniform}[0,1] \qquad \tau, \tau_i, \tau_0, \tau_1 \sim \text{Uniform}[0,1],$$
$$k \sim \text{Uniform}\{1, 2, \ldots, T-1\} \quad \lambda \sim \text{Exponential}(1000).$$
$$\epsilon \sim \text{Uniform}[0, 0.5].$$

The hyperparameter for precision parameters λ was chosen manually to ensure good mixing of the sampler. The prior for parameter k is a discrete uniform distribution; the other uniform distributions are continuous uniform distributions over intervals. Each parameter's prior is independent (e.g., in the single-threshold model, a given pair (λ, τ) has prior probability $p(\lambda, \tau) = p(\lambda)p(\tau)$).

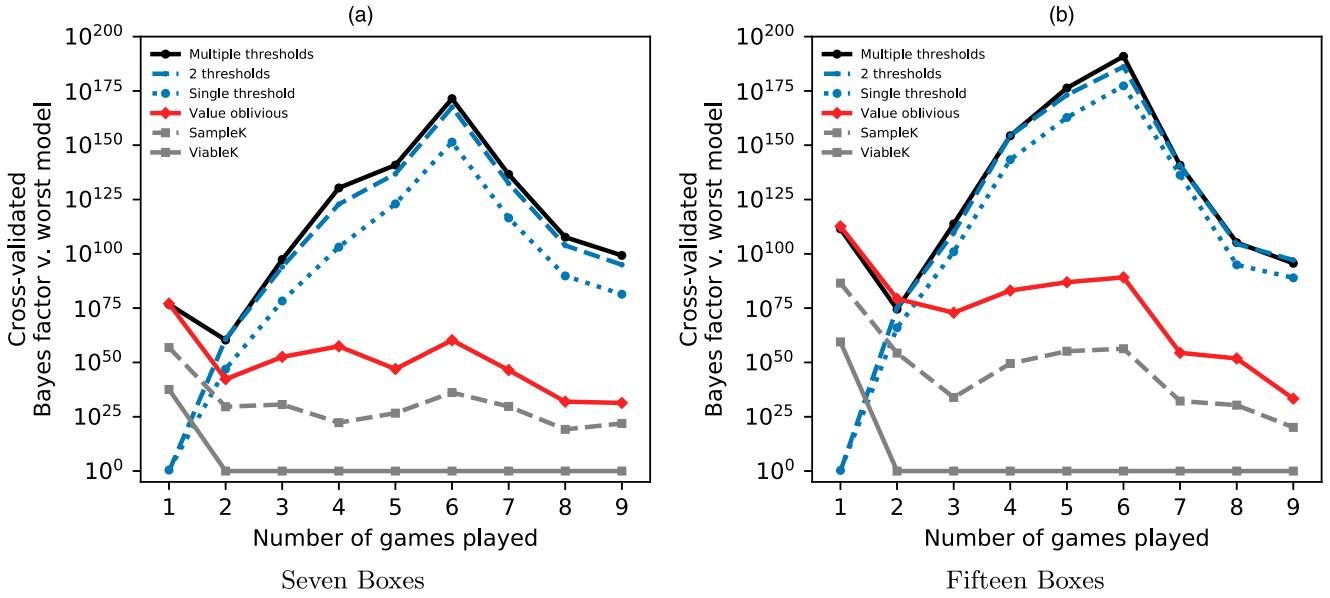
5.2.2. Related Work.

Previous work has also evaluated threshold-based models of human behavior in optimal stopping problems. Guan et al. (2014) study a threshold model in which each participant has a personal latent threshold for each box constrained to decrease as more boxes are opened. Guan et al. (2015) compare this model with a hierarchical model in which the deviation of participants' thresholds from the optimum is directly modeled. Lee (2006) estimates a family of threshold-based models, which include models with single thresholds, two thresholds, and separate thresholds for each box. Importantly, all of these models estimate thresholds separately for each participant, but the thresholds are shared across all repetitions of the task; in contrast, we estimate the multiple-thresholds model separately for each repetition but share thresholds across participants.

5.3. Model Comparison Results

Figure 8 gives the cross-validated Bayes factors for each of the models of Section 5.2. The models were

Figure 8. (Color online) Cross-Validated Bayes Factors for Each Model Compared with the Lowest-Evidence Model in Each Game



estimated separately for each number of games, that is, each model was estimated once on all of the first games played by participants, again on all of the second games, etc. This allows us to detect learning by comparing the estimated values of the parameters across games. The cross-validated Bayes factor is defined as a ratio between two cross-validated model evidences. Because we are instead comparing multiple models, we take the standard approach of expressing each factor with respect to the lowest-evidence model for a given number of games. These normalized cross-validated Bayes factors are consistent in the sense that, if the normalized cross-validated Bayes factor for model h^1 is k times larger than the normalized cross-validated Bayes factor for h^2, then the cross-validated Bayes factor between h^1 and h^2 is k. As a concrete example, the two-threshold model had the lowest cross-validated model evidence for participants' first games in the seven-box condition; the cross-validated model evidence for the value oblivious model was 10^{21} times greater than that of the sample k model and 10^{40} times greater than that of the viable k model.

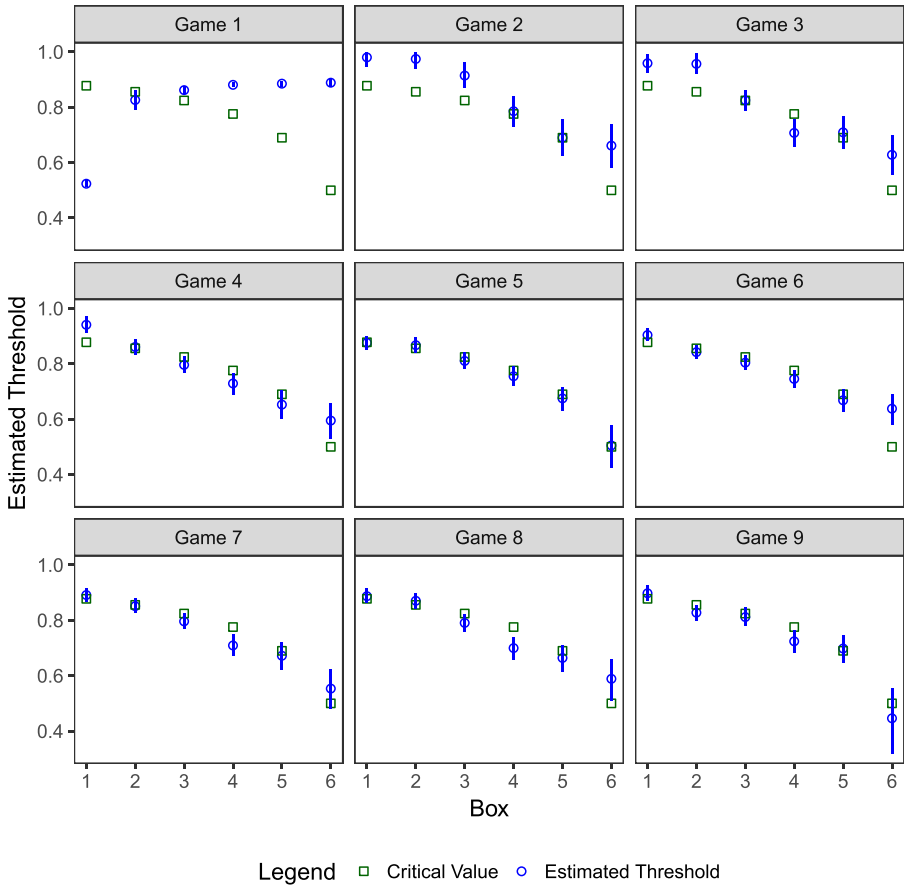
In first game played in both the 7- and 15-box conditions and in second game in the 15-box condition, the best-performing model was value oblivious. In all subsequent games and in both conditions, viable k was the worst-performing model,[7] and the multiple-threshold model was the best-performing model. The two-threshold model was the next best-performing model.[8]

In Figure A.4 in the appendix, we present the results of Figure 8 using the standard Bayes factor, in which the performance of the model on the entire data set is marginalized over the prior distribution rather than marginalizing the performance of the model on a test data set over the posterior distribution with respect to a disjoint training data set. The primary difference between the evaluation by standard Bayes factors and the results of this section is the performance of the two-threshold model. In the seven-box condition, the two-threshold model has a very similar Bayes factor to the multiple-threshold model (after game 2) rather than having decisively lower performance. In the 15-box condition, the two-threshold model has a higher Bayes factor than the multiple-threshold model in contrast to the cross-validated Bayes factors, which are very similar. These differences are consequences of the uninformative priors that we impose on threshold values in these models, which cause the Bayes factor to favor lower-dimensional models (such as two threshold) more strongly than the cross-validated Bayes factor.

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 6(a) but improves on that analysis, because here, the most common stopping points—the early boxes—do not dominate the average.

Furthermore, players' estimated thresholds approached the theoretically optimal values remarkably quickly. Figure 9 shows the estimated thresholds for the seven-box condition along with their 95% posterior credible intervals. The estimated thresholds for the second

Figure 9. (Color online) Estimated Thresholds in the Seven Box Games



Note. Error bars represent the 95% posterior credible interval.

and subsequent games are strictly decreasing in the number of boxes seen, like the optimal thresholds. Overall, the thresholds seem to more closely approximate their optimal values over time. After only four games, each threshold's credible interval contains the optimal threshold value.[9] 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 (Figure 8) 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 because of the selection effect from only stopping on nondominated boxes.

In sum, the switch from increasing to decreasing thresholds in Figure 9 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.

One possible explanation for the apparent change in strategy is that players spent the first few games primarily collecting information about the distribution and then switched to trying to actually win the game only in later games: that is, they spent the first few games *exploring* and then switched to *exploiting* only later (Fang and Levinthal 2009). In that situation, one would expect to see players opening more boxes in early games than in later games, because that is the way to learn the most about the distribution. We do indeed see that players' search depth decreases with experience (Figure 5). However, the most common behavior in the first game (for both the 7- and 15-box conditions) was for players to stop before encountering the maximum (see the dashed lines in Figure 3). This suggests that players are attempting to win—even in the first game—rather than purely exploring.

6. Conclusion: Behavioral Insights

The main research questions that we addressed in this work were whether people improve at the secretary problem—a paradigmatic optimal search problem—through repeated play and whether they approach optimality with experience. The investigation is both qualitative, inferring which decision rules players use, and quantitative, estimating decision thresholds and computing rates of success. As Lee (2006) observed, in addition to its intrinsic interest, the secretary problem is a valuable setting for studying human problem solving. Its simple rules enable it to be studied empirically, whereas its complex solution allows for the comparison of human success rates with optimal success rates and, importantly, human decision procedures with optimal decision procedures.

In contrast to prior research (Seale and Rapoport 1997, Campbell and Lee 2006, Lee 2006), across thousands of players and 10s 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 10 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 to ten percentage points of the maximum win rates possible, and this average is taken without cleaning the data of players who were obviously not trying; removing such players would bring the win rates even closer to the maximum. In looking at candidate-by-candidate decision making, players' probabilities of stopping came to approximate an optimal stop function after a handful of games (Figure 6). Additionally, 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, people were able to do this with no prior knowledge of the distribution and, consequently, sometimes unhelpful feedback (Figure 7).

In the first game, player behavior was relatively well fit by the value oblivious model, which had a fixed probability of stopping at each candidate(box) independent of the values of the candidates. In later plays, threshold-based decision making—the optimal strategy for known distributions—fit the data best (Figure 8). Additional analyses revealed that players' implicit thresholds were close to the optimal critical values (Figure 9), which is surprising given the small likelihood that players actually would, or could, calculate these values.

In domains that are well described as optimal stopping problems, these results mean that the optimal procedure is likely to give a close approximation of human behavior. This is in contrast to many areas of economics, where human behavior is known to qualitatively and consistently deviate from optimal (e.g., Tversky and Kahneman 1992, Goeree and Holt 2001, Camerer 2003).

A few points of difference could explain the apparent departure from prior empirical results. First, to our knowledge, our study is the first study to begin with an unknown distribution that players can learn over the course of the experiment. Previous studies have provided rank information only (e.g., Seale and Rapoport 1997), given a full description of the distribution (leaving no scope for learning) (e.g., Lee 2006), or presented samples from the distribution to the participants before the main experiment (e.g., Rapoport and Tversky 1970, Guan and Lee 2017). 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 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 12–14 participants per condition and assess learning by binning the first 40, second 40, and third 40 games played. In contrast, with over 6,000 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 because most of it happens in the first 10 games. Although 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 of unhelpful feedback (Figure 7), the transition from value oblivious to threshold-based decision making (Figure 8), and the learning of near-optimal thresholds (Figure 9). 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.

Appendix. Additional Figures and Tables

Table A.1. Table of p-Values of Comparisons of Distributions of Winning and Losing Players Broken Out by the Three Possible Error Types in the First Game (No Error, Undersearch, and Oversearch)

Game	2	3	4	5	6
No error	0.02	0.56	0.78	0.27	0.96
Undersearch	$< 10^{-10}$	0.05	0.08	0.19	0.48
Oversearch	$< 10^{-10}$	0.01	0.01	0.34	0.17

Notes. The p-values were computed on a 2×2 contingency table for each game and initial error type. The columns are number of nondominated boxes stopped on and number of nondominated boxes passed on. There is one row for subjects who won the first game and one row for subjects who lost the first game.

Figure A.1. (Color online) Screenshot of the 15-Box Treatment with 3 Boxes Opened

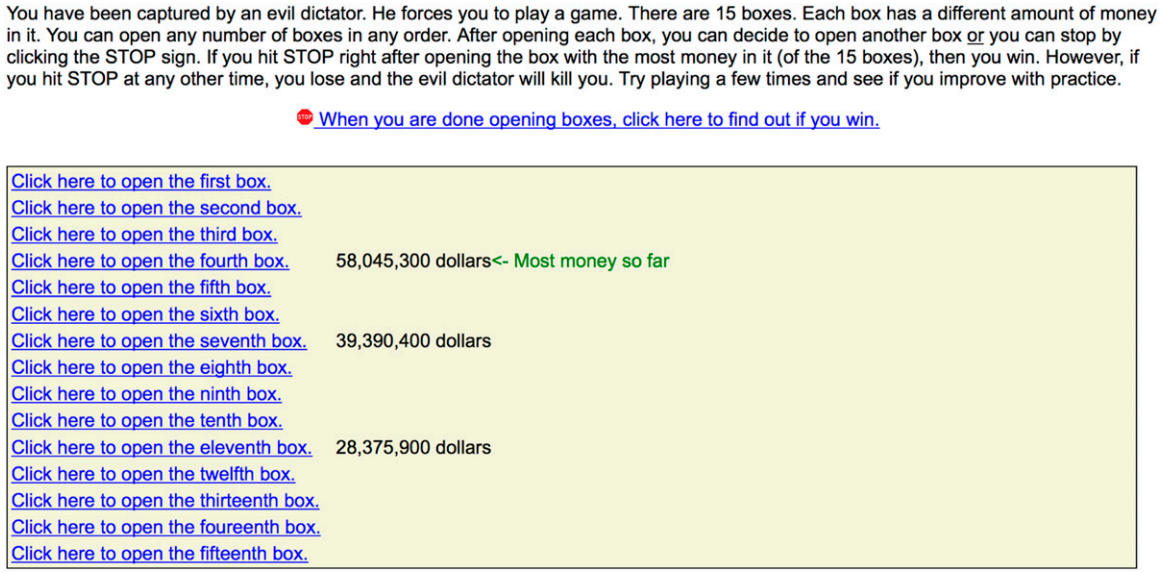
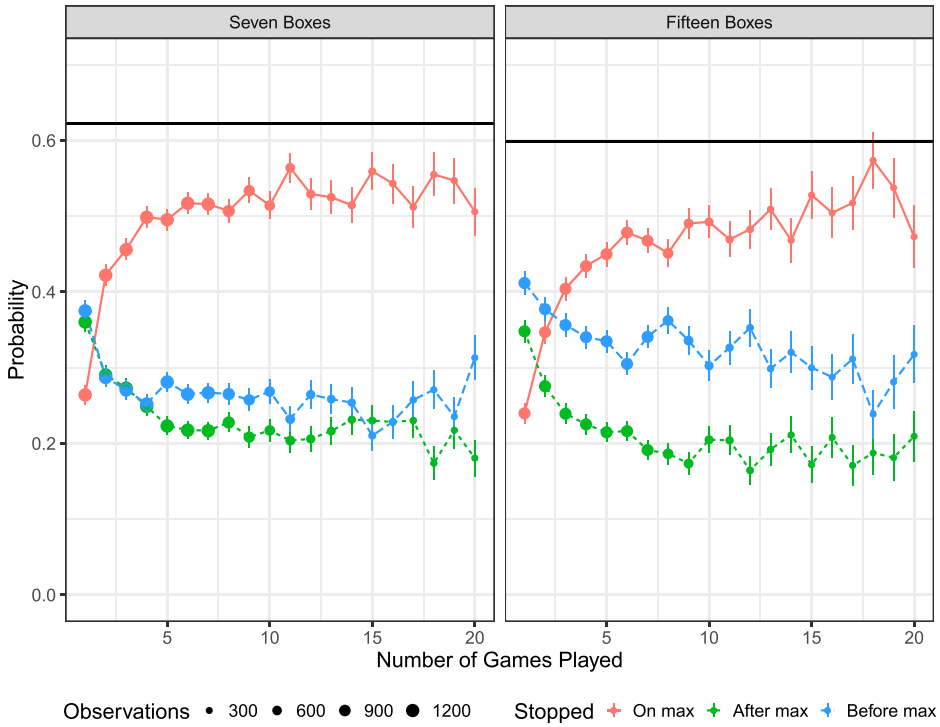
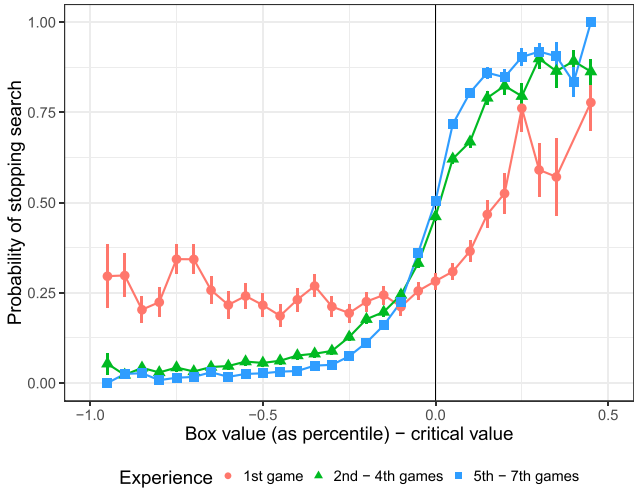


Figure A.2. (Color online) Rates of Winning the Game and Committing Errors for Human Players Where Each Player Played at Least Seven Games



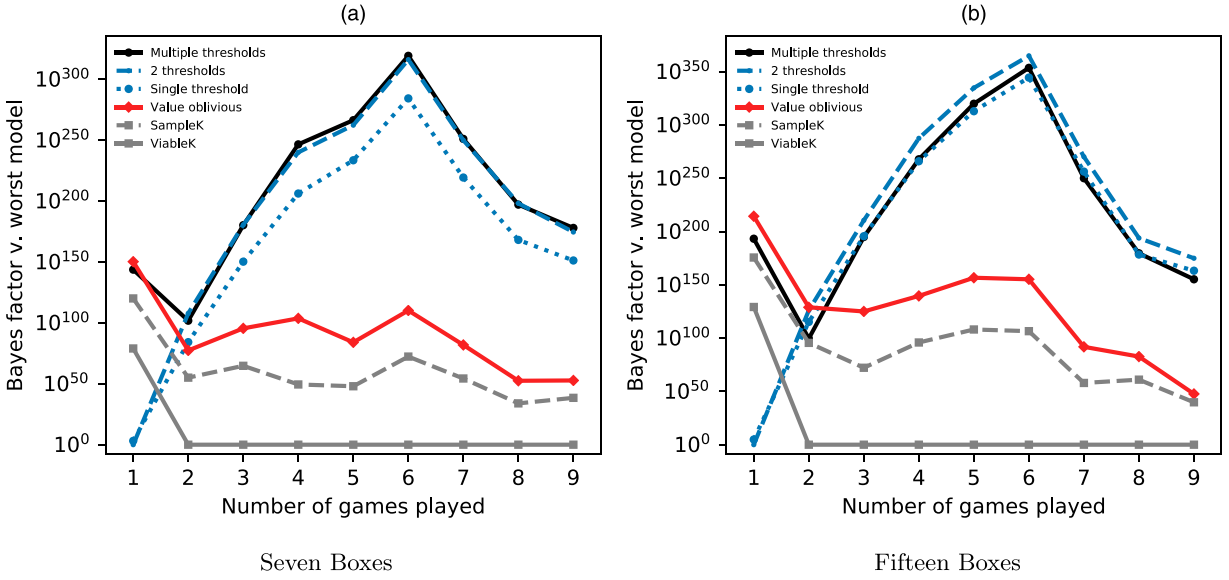
Notes. Error bars indicate ±1 standard error; when they are not visible, they are smaller than the points. The area of each point is proportional to the number of players in the average. Solid lines indicate stopping on the maximum, long dashes indicate stopping before the maximum, and short dashes indicate stopping after the maximum.

Figure A.3. (Color online) In Figure 6(a), Different Players Contribute to Different Curves



Notes. For example, a player who only played 1 time would only contribute to the circle-denoted curve, whereas someone who played 10 times would contribute to all three curves. To address these selection effects, in this plot, we restrict to the first seven games of those who played at least seven games.

Figure A.4. (Color online) Standard Bayes Factors for Each Model Compared with the Lowest-Evidence Model in Each Game



Endnotes

[1] The final box, which must be accepted if opened, is an exception. In our game, a pop-up warning prevented players from choosing dominated boxes.

[2] Despite the term "percentile," we use decimal notation instead of percentages for convenience.

[3] Although it is idealized, the LP agent is nevertheless not an optimal Bayesian learner. It seems very reasonable to want an equal posterior probability of the next value occurring in each interval between the data observed so far; the estimate of Equation (1) has this property. However, there is no countably additive prior that will produce a sequence of estimates having this property when conditioned on successive observations (Hill 1968). We are not aware of any tractable, analytic, optimal Bayesian procedures for sequential quantile estimation.

[4] The fifth game was the first game where the differences between the distributions of behavior of players who initially won and those who initially lost were not significant at any standard level for any of the three possible error types in the first game (no error, undersearch, and oversearch). See Table A.1 in the appendix for p-values.

[5] We considered models with one λ per box, but they did not perform appreciably better than the single-λ models.

[6] More precisely, it approximates the optimal model arbitrarily closely as $\lambda \to \infty$.

[7] Recall that, although sample k was inspired by the optimal strategy for the classical secretary problem, neither sample k nor viable k are optimal strategies for the repeated secretary problem.

[8] We tested whether a different boundary between the two thresholds would perform better by estimating a model in which the choice of boundary was a separate parameter. This model actually performed worse that the two-threshold model. This indicates that the data do not argue strongly for a different boundary, and hence, the only effect of adding a boundary parameter was overfitting.

[9] 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. We caution that this is not necessarily equivalent to the true values differing significantly from optimal in games 5–8; see Wagenmakers et al.(2019) for a detailed discussion of this issue.

References

Alpern S, Baston V (2017) The secretary problem with a selection committee: Do conformist committees hire better secretaries? *Management Sci.* 63(4):1184–1197.

Alqallaf F, Gustafson P (2001) On cross-validation of Bayesian models. *Canadian J. Statist.* 29(2):333–340.

Bearden JN (2006) A new secretary problem with rank-based selection and cardinal payoffs. *J. Math. Psych.* 50(1):58–59.

Bearden JN, Rapoport A, Murphy RO (2006) Sequential observation and selection with rank-dependent payoffs: An experimental study. *Management Sci.* 52(9):1437–1449.

Camerer CF (2003) *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton University Press, Princeton, NJ).

Campbell J, Lee MD (2006) The effect of feedback and financial reward on human performance solving secretaryproblems. Ron Sun R, Naomi Miyake N, eds. *Proc. Annual Conf. Cognitive Sci. Soc.* (Cognitive Science Society, Austin, TX) 1068–1073.

Corbin RM, Olson CL, Abbondanza M (1975) Context effects in optional stopping decisions. *Organ. Behav. Human Performance* 14(2):207–216.

Fang C, Levinthal D (2009) Near-term liability of exploitation: Exploration and exploitation in multistage problems. *Organ. Sci.* 20(3):538–551.

Ferguson TS (1989) Who solved the secretary problem? *Statist. Sci.* 4(3):282–289.

Freeman PR (1983) The secretary problem and its extensions: A review. *Internat. Statist. Rev.* 51(2):189–206.

Gilbert JP, Mosteller F (1966) Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.* 61(313):35–73.

Goeree JK, Holt CA (2001) Ten little treasures of game theory and ten intuitive contradictions. *Amer. Econom. Rev.* 91(5):1402–1422.

Guan M, Lee MD (2017) The effect of goals and environments on human performance in optimal stopping problems. *Decision* 5(4): 339–361.

Guan M, Lee M, Silva A (2014) Threshold models of human decision making on optimal stoppingproblems in different environments. Bello P, McShane M, Guarini M, Scassellati B, eds. *Proc. 36th Annual Conf. Cognitive Sci. Soc.* (Cognitive Science Society, Austin, TX).

Guan M, Lee MD, Vandekerckhove J (2015) A hierarchical cognitive threshold model of human decision making on different length

optimal stopping problems. Noelle DC, Dale R, Warlaumont AS, Yoshimi J, Matlock T, Jennings CD, Maglio PP, eds. *Proc. 37th Annual Conf. Cognitive Sci. Soc.* (Cognitive Science Society, Austin, TX).

Hill BM (1968) Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *J. Amer. Statist. Assoc.* 63(322): 677–691.

Jeffreys H (1935) Some tests of significance, treated by the theory of probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 31 (Cambridge University Press, Cambridge, UK), 203–222.

Jeffreys H (1961) *Theory of Probability* (Oxford University Press, Oxford, United Kingdom).

Kahan JP, Rapoport A, Jones LV (1967) Decision making in a sequential search task. *Perceptions Psychophysics* 2(8):374–376.

Kass RE, Raftery AE (1995) Bayes factors. *J. Amer. Statist. Assoc.* 90(430):773–795.

Kruschke J (2015) *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd ed. (Academic Press Waltham, MA).

Lee MD (2006) A hierarchical Bayesian model of human decision-making on an optimal stopping problem. *Cognitive Sci.* 30(3):1–26.

Lee MD, OConnor TA, Welsh MB (2004) Decision-making on the full informationsecretary problem. Forbus K, Gentner D, Regier T, eds. *Proc. 26th Annual Conf. Cognitive Sci. Soc.* (Cognitive Science Society, Austin, TX), 819–824.

Mahdian M, Mcafee RP, Pennock D (2008) The secretary problem with a hazard rate condition. *Proc. 4th Internat. Workshop Internet Network Econom.* (Springer, New York), 708–715.

Neal RM (2003) Slice sampling. *Ann. Statist.* 31(3):705–741.

Palley AB, Kremer M (2014) Sequential search and learning from rank feedback: Theory and experimental evidence. *Management Sci.* 60(10):2525–2542.

Rapoport A, Tversky A (1970) Choice behavior in an optional stopping task. *Organ. Behav. Human Performance* 5(2):105–120.

Rieskamp J, Otto PE (2006) SSL: A theory of how people learn to select strategies. *J. Experiment. Psych. General* 135(2):207–236.

Rubinstein A (2013) Response time and decision making: An experimental study. *Judgment Decision Making* 8(5):540–551.

Salvatier J, Wiecki TV, Fonnesbeck C (2016) Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* 2:e55.

Seale DA, Rapoport A (1997) Sequential decision making with relative ranks: An experimental investigation of the "secretary problem." *Organ. Behav. Human Decision Proces.* 69(3):221–236.

Todd PM (1997) Searching for the next best mate. Conte R, Hegselmann R, Terna P, eds. *Simulating Social Phenomena*, Lecture Notes in Economics and Mathematical Systems (Springer, Berlin), 419–436.

Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertainty* 5(4): 297–323.

Wagenmakers EJ, Lee M, Rouder J, Morey R (2019) Another statistical paradox. Working paper, University of Amsterdam, Amsterdam.

Worthy DA, Maddox WT (2014) A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W. K. Estes. *J. Math. Psych.* 59:41–49.

Zwick R, Rapoport A, Lo AKC, Muthukrishnan A (2003) Consumer sequential search: Not enough or too much? *Marketing Sci.* 22(4): 503–519.