Simple threshold rules solve explore/exploit tradeoffs in a resource accumulation search task

Running title: Threshold rules for exploration/exploitation

Ke Sang[1,2], Peter M. Todd[1], Robert L. Goldstone[1], and Thomas T. Hills[3]

[1] Cognitive Science Program and Department of Psychological and Brain Sciences, Indiana University Bloomington

[2] Indeed, Inc.

[3] Department of Psychology, University of Warwick

Keywords: exploration/exploitation tradeoff; optimal search; resource patches; model comparison; threshold strategy; Secretary Problem

Contact information:
Peter M. Todd
Indiana University, Cognitive Science Program
1101 E. 10th Street, Bloomington, IN 47405 USA
phone: +1 812 855-3914
email: pmtodd@indiana.edu

Abstract

How, and how well, do people switch between exploration and exploitation to search for and accumulate resources? We study the decision processes underlying such exploration/exploitation tradeoffs by using a novel card selection task that captures the common situation of searching among multiple resources (e.g., jobs) that can be exploited without depleting. With experience, participants learn to switch appropriately between exploration and exploitation and approach optimal performance. We model participants' behavior on this task with random, threshold, and sampling strategies, and find that a linear decreasing threshold rule best fits participants' results. Further evidence that participants use decreasing threshold-based strategies comes from reaction time differences between exploration and exploitation; however, participants themselves report non-decreasing thresholds. Decreasing threshold strategies that "front-load" exploration and switch quickly to exploitation are particularly effective in resource accumulation tasks, in contrast to optimal stopping problems like the Secretary Problem requiring longer exploration.

1. Introduction

Search is a ubiquitous requirement of everyday life. Applicants look for the best job to match their skills; scientists search for information to help their research; and web surfers use search engines like Google to obtain information and products from the internet. In many situations, whether to search (i.e., explore for better options) or to stop searching (and exploit the fruits of search already done) is a key issue for making good decisions. Because of its importance, strategies for balancing between exploration and exploitation have been widely studied across many fields, including animal behavior, psychology, management, and computer science (Todd, Hills, & Robbins, 2012; Hills, Todd, Lazer, Redish, Couzin, & the Cognitive Search Research Group, 2015; Christian & Griffiths, 2016). The kinds of strategies that humans and other organisms can use to deploy exploration and exploitation effectively depend on the details of the search tasks they face. In this paper, we investigate a common type of search task to study how, and how well, people regulate and adjust their use of these two components of search.

Many common search tasks involve exploring a sequence of options and deciding whether and how long to exploit (i.e., stick with) each one, with the searcher getting payoff from each selected option at each point of the search. For instance, a person's career can be thought of as a search process over a sequence of jobs—the searcher explores and finds a new job and accepts it, receiving a payoff for as long as they exploit that opportunity; but at any point they can decide to explore again to find a new job. In some of these cases, the rewards from exploiting an option can decrease over time—this happens when the option is actually consumed or used up, as in patches of food eaten by foraging animals (Charnov, 1976; Hills, Kalff, &

Wiener, 2013; Hutchinson, Wilke, & Todd, 2008) or clusters of webpages found and read by an individual online (Pirolli, 2005, 2007) or when the challenges from a particular job become routine and uninteresting. These situations require an ongoing switch back and forth between exploring for a new bountiful option, exploiting that option and extracting its benefit for some time, and then exploring again once the option is sufficiently used up. The searcher here decides how long to exploit the diminishing option before leaving to explore and find a better one.

In other cases, the resources do not deplete, but rather stay constant over time (e.g., a job with a salary that tracks inflation), or go up (e.g., a job with substantial raises, or a romantic relationship that deepens with time), or even alternate (as in a cyclically rising and falling stock, or the Leapfrog task of Knox, Otto, Stone, and Love, 2012, where the value of two resource options switch rank over time). Even for these situations of ongoing benefit, there can still be a reason to leave a current option and explore for possibly better ones (e.g., if one's current salary is insufficient); sit-and-wait foragers such as web-building spiders and some ocean-dwelling filter feeders also face this kind of search as they leave a currently stable feeding site to seek a better one (Beachly, Stephens, & Toyer, 1995). Here the searcher must tradeoff between exploiting their current option and thereby getting some sure return, versus exploring more to find a better option but possibly obtaining a lower payoff until then.

This is the frequently-encountered search task we explore in this paper—deciding when to explore more and when to stop (or stay) and exploit what is available, in a situation where both exploring and exploiting provide rewards, and where there are multiple non-depleting options with known rewards to choose among, but there could be better options still to be found. Consequently, along two important dimensions of search (among several), exploitation with depleting versus non-depleting resources and exploration with ongoing reward versus without,

we focus on one commonly occurring combination: search over non-depleting resources with ongoing rewards during exploration. Other dimensions, such as fixed versus open-ended time horizon and known versus unknown reward distributions, are not explored here; we use a known and fixed horizon and inform participants of the reward distribution in these studies. While there has been a fair amount of research (described in the next section) on how people search for non-depleting resources and what the optimal strategies are in cases like the Secretary Problem where exploration has no direct reward, less is known about the common cases we investigate here where both exploration and exploitation provide rewards (see Mehlhorn et al., 2015, for a review). Do people use the same strategies for these two types of search and are the optimal strategies similar? To find out, we devise a task to study the strategies that people use to make the explore/exploit tradeoff in situations where both are rewarding, and we assess how well people perform in comparison to random baseline and optimal strategies. We also compare the strategies people use with those previously reported for the Secretary Problem, to assess how sensitive the strategies are to the presence of rewards during exploration. Finally, given that people can learn to improve their performance without rewards while exploring (Seale & Rapoport, 1997) we look at the extent of learning when rewards are present (for both exploration and exploitation). We begin in the next section with a consideration of possible strategies for search with and without rewarding exploration, before turning to our experimental design and empirical and simulation results.

2. Search strategies for non-depleting resources

2.1. Exploration without reward: The Secretary Problem

To provide a point of comparison for our novel task, where the searcher can switch back and forth between exploration and exploitation, it is useful to first introduce a form of search that has been well-studied by mathematicians, economists, and psychologists: optimal stopping problems, in which the task is to make a one-time decision about when to stop exploration and switch to exploiting the current discovered option (Ferguson, n.d.). In many optimal stopping problems, individuals must first explore through some of the available options without accruing any reward, until they find the one option they want to choose and exploit for its payoff thereafter. For example, employers may consider many job applicants over an extended period without employing anyone until they finally decide to hire a particular person and start benefitting from having their position filled. This type of search task is embodied in the classic Secretary Problem (Ferguson, 1989), in which a searcher (employer) sees one secretarial candidate at a time and aims to hire the single best applicant in the population, without knowing the overall distribution of abilities in the population and without being able to return to any previously-interviewed candidate (e.g., because those people were hired elsewhere in the meantime). The employer must decide when they think they have found the best candidate and then stop their exploration and make the hire to get the benefit from "exploiting" the worker. In this case, the searcher must trade off exploring more and possibly finding a better candidate but also possibly passing over the best candidate, against exploiting sooner with the best candidate seen so far who may not be the best overall. Similar search tasks also appear in domains including selecting a mate (Todd & Miller, 1999), finding a parking space (Hutchinson, Fanselow, & Todd, 2012), and buying a house or other unique items (though in these cases, there can be both search costs and also benefits from stopping search sooner and exploiting a chosen

option for longer, aspects that are absent in the generic Secretary Problem).

In the Secretary Problem, the solution for optimizing the probability of selecting the very best applicant involves two stages: First, the searcher must explore (and pass by) an initial set of applicants, the size of this set approaching N/e for large numbers of applicants N (where e is the base of the natural logarithm, ≈2.71828); second, the searcher must stop and accept the first applicant after the initial set who is better than all other applicants seen so far (Ferguson, 1989). The first stage can be thought of as gathering information about the range of possible values and setting a threshold (the highest value seen so far) for the second stage. This optimal strategy gives the searcher a 37% probability of selecting the best applicant.

To see how and how well people actually solve this problem, Seale and Rapoport (1997) presented participants with fixed-length sequences of values (using ranks rather than actual values so that distributions could not be learned) and had participants stop the search whenever they thought they were on the highest value. Participants most often appeared to use a cutoff rule having the same form as the optimal strategy—passing over an initial number of options and then taking the first subsequent option seen that exceeds all the preceding options. Across 100 searches with N=40 options each, participants achieved a 30% mean proportion of success in selecting the single best option. This was lower than the optimal performance of 37% in part because participants did not search long enough. Participants did show quite effective learning, improving their mean rate of success from 26% in the first 50 trials to 35% in the second 50 trials, largely by searching longer. Participants thus could learn to perform rather well in this sequential search task by increasing the amount of exploration they did, even though there was no direct payoff during exploration.

2.2. A search task with rewarding exploration and exploitation

Optimal stopping problems like the Secretary Problem involve exploring some number of sequential options with no payoff other than gathering information to determine when to stop exploring and make the single switch to the final exploitation phase. Other paradigms allow for transitions back and forth between information-gathering exploration (which is again non-rewarding) and reward-accumulating exploitation (Navarro, Newell, & Schulze, 2016). However, many common search domains provide payoff during exploration *and* exploitation (which may serve to increase the amount of exploration done) and also allow switching back and forth between periods of exploring, exploiting the found resources, and exploring again for something better. Such search can occur, for example, when checking out and selecting restaurants, genres of books to read, movies to watch, music to listen to, specific products to buy, and relationships to engage in (Cohen & Todd, 2018). It is commonly studied in the form of bandit problems with multiple rewarding options (arms) to explore and select among; in some such experimental settings people have been found to switch once from exploration to exploitation, as in optimal stopping problems (Lee, Zhang, Munro, & Steyvers, 2011). We are interested here in the effective strategies for this widespread type of search, the actual strategies that people apply, and how they differ from optimal strategies and from strategies people use in stopping problems without ongoing rewards such as the Secretary Problem.

To study the exploration/exploitation tradeoff strategies that people employ when exploration (as well as exploitation) can provide a payoff, we developed a search task game in which participants must accrue resources over a sequence of 20 turns (Sang, Todd, & Goldstone, 2011). The resources are represented as points on cards, which individual participants search

through on a computer screen. The participant begins with a deck of 20 cards all face-down in the lower left corner of their screen; they are told (accurately) that each card has a number from 1 to 99 on it, that the card values are uniformly distributed in the decks (with repetitions of particular values possible in each deck) so the expected value of each new card is 50, that they have 20 turns in a game, and that their task is to accrue as many points as they can during the 20 turns in each game. (We also occasionally refer to games as trials.) There are two distinct actions possible on each turn in the game, corresponding to exploration and exploitation, and participants get points from each action as follows: A participant can either explore by flipping over a card from the deck and getting the points revealed on that card, at which point the card is displayed face-up across the top of the screen (see Fig. 1); or they can exploit a card they have previously found by clicking on it in the display, and getting the points shown on that card added to their total score. (In this way our task differs from a standard bandit problem, as here exploring creates new options that can be exploited again on subsequent turns, and which never change their value.) Thus a participant's total score for a game is the sum of all of the points accrued by their explorations (choosing cards from the deck) and their exploitations (choosing cards already on the screen) over all 20 turns.

---------- Insert Figure 1 about here ----------

This search problem differs from optimal stopping problems like the Secretary Problem in a number of ways. As described earlier, the key difference is that individuals in our card task do not decide when to stop their search, but rather decide how to allocate their actions as they see fit between exploration and exploitation across the duration of the task. They also receive payoff

during both types of search actions, in contrast to the Secretary Problem where payoff is solely determined by what value the searcher chooses to stop on and exploit. Reflecting many real-world situations, searchers in the card task have knowledge of the possible outcomes they face, they can return to previously seen options, and they can switch back and forth repeatedly between bouts of exploration and exploitation—aspects missing from the Secretary Problem. The card task and Secretary Problem do, though, both share a known fixed time horizon and a lack of explicit search costs[1].

To provide an upper bound on participants' performance on this task and determine a possible approach that they might use to solve it, we determined the optimal strategy using backward mathematical induction (see Supplemental Materials file). The optimal strategy for the card task with rewarding exploration and exploitation is to begin by exploring and then switch once from exploration to exploiting the best card found so far (i.e., one displayed on the screen) for all remaining turns whenever that best card's value exceeds a predetermined threshold level that falls with increasing turns. This decreasing threshold curve is influenced by the range of possible card values (highest and lowest) and the number of turns remaining at each point in the search game. For the settings here, with card values ranging from 1 to 99 and 20 total turns in the game, the optimal threshold curve is shown in Fig. 2. This function is based not only on comparing the results of exploiting the value of the current highest card versus the expected value of exploring again once, but also on the expected value of exploring further and then exploiting a better card value found later in the search. While this optimal strategy does not call for switching back and forth between multiple phases of exploration and exploitation, its time-varying threshold

[1] For an interesting hybrid search task that involves payoff from all items seen, but without knowing their values, only their ranks, see Richard Feynman's restaurant problem: http://www.feynmanlectures.info/solutions/restaurant_problem_sol_1.html

is also quite different from the optimal 37% rule for the Secretary Problem. One striking consequence is that searchers could pass by a value early in their search that they should accept (i.e., return to and exploit) later in their search—this happens in about 9% of card sequences in the current setting, but cannot occur in the optimal Secretary Problem solution.

---------- Insert Figure 2 about here ----------

3. Methods

To find out how people search in settings when both exploration and exploitation are rewarding, we conducted a search experiment to study participants' decisions, and compared their behavior to a range of strategies including optimal and random baselines, decreasing threshold rules based on the optimal strategy, and sample-based rules like the cutoff rule found for the Secretary Problem. We investigated participants' strategies both through model fitting and by explicitly asking them what thresholds they may have been using across their search. We recruited 191 participants from the Indiana University Bloomington psychology student participant pool in exchange for credit for their courses. They were told that their goal was to accumulate as many points as possible in each search game, by flipping over cards from the deck to get their points or clicking on cards already found and displayed on the screen and getting the points shown on those cards. Participants were also informed about the distribution of card values as indicated earlier. The general framework of the experiment is shown in Fig. 3.

---------- Insert Figure 3 about here ----------

In the experiment, a turn refers to one exploration or exploitation decision, and every game consists of 20 turns. On the first turn, the participant had to explore, flipping over the top card on the deck, and thereafter the participant decided at each turn whether to explore from the deck or exploit a displayed card value. Participants' choices and response times between choices were recorded. After completing each game by playing 20 turns, the screen was cleared and the next game began (with a new deck with the same parameters); participants played 30 games so we could evaluate their change in strategy and performance over time.

For example, in Fig. 1, four cards have been taken from the card deck so far, with the first three values shown in a small font and the highest (and most recently found) card value, 91, in a larger, red font. The screen shows that the number of turns taken thus far is 15, there are 5 turns left, and the point total so far for this game is 1245. The number of points received by the participant on each turn in this game is shown in the list beside the deck. On this 16th turn, the participant must decide whether to exploit the highest value 91 again, as has been done for the previous 12 turns (and which the optimal strategy dictates), or explore the deck further, hoping for an even higher card value.

After each of the 30 independent games, participants were told the points they received on that game, the points that the optimal strategy would have earned, and (redundantly) whether the participant did better, worse, or the same as the optimal strategy. After finishing all 30 games, participants were asked to state explicitly what card-value thresholds they may have had in mind while searching. For turns 2, 5, 9, 13, 17, and 20, participants indicated the minimum

card value that they would have been satisfied with at that point, and hence would have made them stop exploring and start exploiting this card for the rest of the turns in the game. (Whether or not they were actually using a threshold rule to make their decisions, they could still have a sense of what card values would be good enough to make them switch to exploitation at each turn, and this potential knowledge should be stronger after 30 games of practice and improvement toward an effective strategy.) Participants were asked to report explicitly this threshold value with the following instructions:

> **Post-experiment Questionnaire:** We would like to ask you about your general strategy for doing this task. At different points in the game, you may have felt that if you had a large enough card value showing on the table, then you would be satisfied with it, and would pick it instead of drawing a new card from the deck. For example, on the first round, if you happened to draw a 99, then you may have been satisfied enough with it to pick it for the rest of the game. For each of the rounds listed below, what was the LOWEST value for a card that would be enough for you to pick it for the rest of the game? When you are finished, click the <space bar> to submit your numbers.
>
> For the 2nd [5th, etc.] round out of 20, the lowest card that would be enough was _____.

4. Decision behavior

Across all of the turns taken by all participants (191·30·20 = 114600 turns), there was 73.3% exploitation and 26.7% exploration. The optimal strategy calls for more exploitation, 81%, across the same games that participants saw[2] ($t(190) = 6.31$, $p < 0.001$; pairwise t-test). In other words, participants explored too much. Participants' mean total points per 20-turn game was 1528 (*SD* 266), lower than that achieved by the optimal strategy applied to the same values that participants saw, 1595 (*SD* 224) ($t(190) = 16.07$, $p < 0.001$; pairwise t-test).

4.1. Switches and final exploitation patterns

The optimal strategy for this search task dictates that there will be at most one switch from exploration to exploitation per 20-turn game—only when the highest card seen so far exceeds the current optimal threshold. There should never be a switch from exploitation back to exploration—all exploration should be "front loaded" to the first portion of a game because this maximizes the opportunities to subsequently take advantage of (i.e., exploit) high values that are found during explorations. Participants behaving non-optimally, by contrast, might switch back from exploitation to exploration for many reasons, including intrinsic stochasticity, boredom, employing particular strategies, and/or changing strategies over time. Subsequently, as the end of the game approaches, any participants who have switched from exploiting to exploring may well switch to exploitation again to take advantage of previously-found high values.

Participants switched between exploration and exploitation a mean of 1.83 times per game. In most cases, after some number of turns (during which they have explored and possibly

[2] Conceptually, 30 sequences of 20 values were generated for each participant, and the optimal strategy was applied to those same values, which could mean that the optimal strategy "saw" fewer or more values than a participant did in any particular game sequence if it specified switching to exploitation at an earlier or later turn than the participant selected.

exploited over one or more stretches), participants made a final switch to exploitation for the rest of the turns until the end of the game (and hence ended up exploiting on the last turn in 94.9% of all games). The turn on which this final run of persistent exploitation begins will vary depending on the search strategy used. For example, a strategy with a constant exploitation threshold of 90 would lead to a later mean switch point than the optimal strategy does, because cards exceeding this high threshold are less common than cards exceeding the decreasing optimal threshold. The mean position of the initial turn for this persistent exploitation (that is, the first exploitation turn of a game that has no later exploration turns after it) is 7.64 across all participants. The optimal strategy, when applied to all of the same data that participants saw, switched to persistent exploitation at (mean) turn 5.14. This indicates that people continue exploring for longer than optimal by about two turns. Fig. 4 shows the frequency distributions of these initial turns of persistent exploitation for both participants and the optimal strategy. Compared to the optimal strategy, the distribution of participants' behavior has a fatter tail, with an appreciable proportion of participants exploring even until the very end of some games. (See Section 6 for how these results change with learning across games.)

---------- Insert Figure 4 about here ----------

4.2. Response times

How long participants took to decide whether to switch between exploring and exploiting or to continue doing what they were doing can give insight into the decision strategies they use.

Table 1 shows descriptive statistics for response times (RTs)—that is, times leading up to the click on a certain action (since the last click)—for deciding to *continue* to explore or exploit (i.e., the previous action and current action are the same) and deciding to *switch* to explore or exploit (i.e., the previous and current actions are different; all switches to exploit are included, not just the final switch). These response times are calculated as follows.

When a participant switches from exploration to exploitation (or vice versa), part of the RT comes from the motoric behavior of moving the mouse from the deck to a card displayed on the screen (or verse versa). This movement time should not be included in the RT for the psychological decision process. To account for this motoric effect, we built a linear regression model for each participant to predict the *log*(RT) values associated with each action based solely on whether that action is a switch or not; the residuals from this model then should reflect only the decision process involved (Knox, Otto, Stone, & Love, 2012). To show the mean RTs in seconds in Table 1, we convert the residuals from the *log*(RT) predictions back into RTs by exponentiation; but because the RTs are not normally distributed, we analyze them further in their *log*(RT) form.

---------- Insert Table 1 about here ----------

We next compared categories of residual RTs in the log scale to see whether some types of search decisions took longer than others. Deciding to continue to explore takes longer than deciding to switch to exploit ($t(380) = 4.84$, $p < 0.001$; two-sample t-test at the individual level), which takes longer than deciding to continue to exploit ($t(380) = 29.21$, $p < 0.001$; two-sample t-test). A second analysis of this data that explicitly seeks to remove motor response time found

the same ordering for these three key decision times: continue to explore > switch to exploit > continue to exploit—see Supplemental Materials file. This fits with the cognitive steps involved in following a decreasing threshold rule (like the optimal strategy introduced above and others discussed in section 5): Once the searcher decides to exploit, she should continue to exploit without having to consider any further information, making this a quick decision. But to decide to continue to explore (after any turn of exploration) takes longer, because the searcher needs to verify (1) that the card just found is not above the current exploitation threshold, and also verify (2) that the best card found previously is not now better than the current threshold (which may be lower than the threshold when that card was first seen). If either of those verification steps fail, then the searcher would decide to switch to exploitation. Since this fail and switch could happen after one *or* two verification steps, on average we would expect the switch-to-exploit decision to take less time than the continue-to-explore decision which requires passing *both* verification steps, but take longer than the continue-to-exploit decision which involves *no* verification steps—and this is the pattern we see in the RT data. (A fixed-threshold rule would only require the first verification step, making the switch-to-exploit and continue-to-explore decisions take similar amounts of time, which goes against the pattern seen in these data.)

The intermediate duration of the rarer decisions to switch back from exploitation to exploration may suggest a different strategy (or strategy component) for this decision. It could involve one or more computational comparisons, as for the switch-to-exploit decision, or could just involve an internal threshold that does not require checking any card values, such as an "impatience" to return to exploring.

4.3. Explicitly reported versus modeled thresholds

Given that the optimal strategy takes the form of a threshold rule, we next analyzed what thresholds participants may have been using, both by asking them explicitly to report their thresholds and by modeling their best-fitting thresholds. Participants' reported thresholds for exploiting are plotted in Fig. 5, first averaged across all participants for each of the 6 specific turns for which participants were asked to give thresholds and then linearly interpolated between those mean values for the 6 turns. The general trend of the mean reported threshold is flat over turns (the 95% confidence interval for a regression coefficient of reported threshold values on turns is [−.0004, .3684], including 0), in contrast to the optimal threshold also shown which decreases over turns. This flat aggregated pattern reflects wide variance in individuals' reported thresholds, some of which even increase across turns (see Fig. S1 in the Supplemental Materials file).

---------- Insert Figure 5 about here ----------

---------- Insert Figure 6 about here ----------

In addition to participants' explicitly reported thresholds, we also attempted to infer the thresholds that may underlie their decisions in the experiment. We first looked at whether participants were adjusting the range of card values that they would exploit at different turns in each game. Fig. 6 plots the participants' probabilities at each turn of exploiting the highest-available card if its value fell within a certain range (starting with turn 2, the first turn where exploitation is possible), calculated across all exploitations in all games for all participants.

(Participants only rarely exploit a card value that is not the highest one available, in just 1.6% of all exploitation decisions.) For example, the top curve shows the probability at each turn that a participant would choose exploitation, given that the highest card value available to exploit on the screen at that turn was between 90 and 99[3]. All of the curves generally increase over turns, with the lower highest-value curves increasing more rapidly. For turns toward the end of a game, the exploitation probability nears 1.0 for all card values, indicating that participants will exploit whatever highest card value they have once they get to the end of their search. These results indicate that people are taking the turns into account by gradually shifting their exploitation tendency for all card values upward, which is consistent with a gradually decreasing threshold, applied with noise. It is not compatible with a fixed or increasing threshold. (If everyone used exactly the deterministic decreasing threshold of the optimal strategy of Fig. 2, this plot would show abrupt changes in acceptance probabilities for each bin across turns: For example, for highest-value cards in the 70-74 bin, a searcher following the optimal strategy would have a 0% probability of exploiting for turns 2-8 and then would quickly rise to a 100% probability of exploiting for turns 12-20. Mixtures of participants using something close to the optimal strategy could however produce this pattern in aggregate.)

To estimate the underlying thresholds that participants may have been using (which could differ from what they reported using), we treated the thresholds at different turns as model parameters and used maximum likelihood estimation (MLE) to find values that best fit participants' decisions. We built a model that uses a stepwise threshold to decide when to switch

[3] Bins over ranges of highest card value available are used because plotting probabilities for single card values would result in considerable noise in the graph. There is still some noise at the end of the trial for some lower highest-value bins, because, for example, there are very few trials in which a participant gets to turn 18 and still has not uncovered a card with a value greater than 60.

from exploration to exploitation. To allow easy comparison with the participants' reported thresholds, the model's stepwise thresholds are estimated for the same 6 turns that we asked participants about (turns 2, 5, 9, 13, 17, and 20). Given that this model is focused on estimating thresholds rather than reflecting the psychological choice process, it may not capture the details of participants' turn-by-turn behavior well; we compare it to other plausible models in the next section.

This *six-threshold model* has 7 parameters: T_1-T_6 represent the thresholds, from 1-99, that apply across turns 2, 3-5, 6-9, 10-13, 14-17 and 18-20 respectively, and the strength parameter *s* is a positive value that reflects how strongly and consistently the participant follows the threshold rule—if *s* is large, then participants usually make a choice that is consistent with the current threshold $T(t)$ that holds at turn *t*, and if *s* is small, then the model shows considerable randomness in determining whether to explore or exploit on a given game. The model specifies the probability of exploring at each turn *t* with respect to the current threshold as follows:

$$Pr_{explore}(t) = \frac{1}{1 + e^{-s[T(t)-Max]}} \quad (1)$$

where *s* is the strength parameter, *Max* is the highest card value seen (i.e., on the table) before turn *t*, and $T(t)$ is the threshold (T_1-T_6) that holds at turn *t*. When the model indicates exploitation rather than exploring, it exploits the highest available card.

We used MLE to estimate parameter values for each individual. The average best-fitting model threshold function is plotted in Fig. 5, interpolating between the medians of the 6 threshold parameters across participants (using medians because the parameter distributions are skewed). This modeled threshold falls across turns, though relatively evenly and not as steeply

as the optimal threshold at the end of each game[4]. There is also an evident mismatch between the flat reported threshold and the falling modeled threshold, which could be a consequence of participants using thresholds but reporting them incorrectly (e.g., not knowing or remembering them), or of participants making their decisions in some way other than just using thresholds (see next section). However, as discussed in Section 5.2, these differences in threshold patterns do not have a very large effect on the performance achieved by using them with threshold rules in the card search task—even if participants are not describing closely what they may be doing, what they do describe would perform well if it were used. Finally, the median of the strength parameter *s* is 0.13, indicating substantial randomness in choices—that is, there are many situations in which the model chooses to exploit even though the highest card on the table is below the current estimated threshold, and chooses to explore even though the highest card is above the threshold. This randomness reflects the stochasticity of participants' choices.

5. Comparison of decision strategies

Of course, a multi-stage changing-threshold mechanism may not be what people are actually using to make their decisions. We therefore tested a variety of further strategies in terms of both their fit to participants' data, and their performance compared to the optimal threshold rule. We assessed three types of strategies: random baseline models, threshold models, and sampling models that, like the cutoff rule for the Secretary Problem, decide when to stop based

[4] This falling threshold was also found when modeling subsets of participants with different patterns of reported thresholds, including increasing—see Fig. S2 in the Supplemental Materials file.

on an initial sample of values. We first describe the models and then report their performance on the card search task and their fit to participants' data.

5.1. Strategies compared

For all of the strategies below we describe a stochastic model used to find the parameters that best fit the model to participants' data. We also report the performance on the card search task of the corresponding deterministic form of each model (except the epsilon-greedy baseline).

5.1.1. Random baseline models

A standard type of random model for exploration/exploitation tasks is the *epsilon-greedy* model (Sutton & Barto, 1998), which uses a single parameter ε to control the probability of exploration (selecting a new card) on each turn, versus exploitation (taking the value of the highest card seen so far). A model more specific for this card search task is the *random switch* model (also with one parameter), which randomly picks a turn k at which to switch from exploration to exploitation, with k chosen from the range [2, 20] using the distribution of switch turns from participants shown in Fig. 4 (left). The model then has a tendency to explore for k–1 turns and to exploit on subsequent turns in that game, with the probability influenced by the strength parameter s:

$$Pr_{explore}(t) = \frac{1}{1 + e^{-s(k-t)}} \tag{2}$$

where $Pr_{explore}(t)$ is the probability of exploration on turn t, s is the strength parameter, and k is the randomly-chosen switch turn (not a fitted parameter). This equation results in higher likelihood of exploring at the beginning of the game and higher likelihood of exploiting at the

end, with equal probability of both at the switch turn (when $t = k$).

5.1.2. Threshold models

Given the prevalence of simple threshold rules in human bounded rationality (e.g., satisficing rules—Simon, 1990) and the fact that optimal behavior in the card search task follows such a strategy, we also tested three forms of threshold models. The simplest *one-threshold model* (two parameters) has a single fixed threshold T that applies across all turns and a strength parameter s controlling the probability of exploring on a given turn based on how far the highest card seen so far is from the threshold T (above or below), according to Equation 1. The *linear decreasing threshold* model (three parameters) uses Equation 1 with a falling threshold $T(t) = b + m(t–2)$ where b is the initial threshold at turn 2 and m is the (negative) slope, along with strength parameter s. The *two-threshold model* (four parameters) has two threshold values T_1 and T_2 and a "jump turn" parameter k in the range [2, 20] that determines how long each threshold is used: T_1 for turns 2 to k, and T_2 for turns k+1 to 20. It also uses a strength parameter s and determines the probability of exploring by Equation 1. Finally, the *six-threshold model* (with seven parameters) described earlier has six threshold values T_1-T_6 that are used for turns 2, 3-5, 6-9, 10-13, 14-17, and 18-20, respectively, and a strength parameter s combined via Equation 1.

5.1.3. Sampling models

Another class of simple search rules base their stopping decisions on information gained from an initial sample of options. Each of these models has two parameters: one parameter that controls the size of the sample used, and another, h (for "trembling hand"—see Selten, 1975), that introduces stochasticity by setting the probability of the model's deterministically-selected

action (exploring or exploiting) to $1-h$ and the probability of the other action to h (so larger h indicates more stochasticity, while larger s in the previous models indicates less stochasticity). The *fixed-sample model* first assesses a sample of fixed size by exploring for k turns and then exploits the highest value card seen in that sample starting on turn $k+1$ (and for all remaining turns)[5]. (This is very similar to the epsilon-first strategy in multi-armed bandit problems—see Lee, Zhang, Munro, & Steyvers, 2011, where it is also called π-first.) The *cutoff model* similarly explores for k turns, determines the highest value seen in that initial sample and sets it as the cutoff threshold (rather than exploiting it), and then continues exploring until it finds a card that is above that cutoff threshold, which it exploits for the rest of the turns. Use of this rule has been studied particularly for situations where the distribution of available values is not known, as for the Secretary Problem discussed earlier. The *successive non-candidate count model*, also studied as a potential strategy to solve the Secretary Problem (Seale & Rapoport, 1997), is defined in terms of "candidates," which are those cards that have the highest values seen so far in the current game (and hence are candidates for exploitation), and "non-candidates," which are all other cards (i.e., those not the highest seen so far, hence not appropriate to exploit). The model starts with the necessary exploration on turn 1; that first card is by definition a candidate (as it is the highest-value card seen so far). It continues exploring, assessing whether each new card is a candidate or a non-candidate and counting up how many non-candidates in a row it encounters. If the number of successive non-candidates seen in a row reaches a threshold value j, then the model will exploit (for all remaining turns) the next candidate it encounters. This model can thus be interpreted as being based on impatience—if it has been too long since finding the previous

[5] Note that the size of the initial sample will be reduced if any of the first k turns individually "tremble" into unintended exploitation, decreasing the number of cards explored; but the intended ongoing exploitation will still start at turn $k+1$.

exploitable candidate option, then the searcher "gets impatient" and takes the next option that is higher than all those encountered previously.

5.2. Model performance

How well do these different models perform on the search task, balancing exploration and exploitation? To find out, we used grid search to find the best performing parameter values for the deterministic version of each model when applied to a set of 50,000 randomly generated sequences of 20 card values (different from the sequences seen by participants). The mean scores of the best performing models over the 50,000 simulated runs are shown in Table 2. The optimal strategy scored 1601.8; participants scored 1528 on average. The threshold strategies all scored very close to the optimal (around 1600), while the two random strategies both performed more poorly (around 1300). The sample-based strategies fell in between, with the cutoff strategy from the Secretary Problem doing worst of these (at 1391). Thus for this problem, a simple threshold rule—even one that uses a single fixed threshold, set at 79—can perform about as well as the optimal decreasing threshold strategy.

---------- Insert Table 2 about here ----------

One potential drawback of the threshold rules is that they need to have their thresholds and any switch points for changing the thresholds predetermined based on knowledge of the distribution of values that will be encountered. In contrast, the sampling rules learn their stopping thresholds through the initial sample of values they collect as they are used, making them robust across differences in underlying distributions. For example, the sample-based cutoff

rule scored 87% of the optimal score in the original 1-100 card value range with its sample size parameter set at 2, and it actually improves its performance to 94% of optimal when applied with the same sample size to an expanded card value range of 500-1000. In contrast, the one-threshold rule does not show this robustness to changing inputs: With its threshold set at the best-performing value of 79 for card value range 1-100, it scored 99.8% of optimal, but its performance with that threshold falls to 83% when applied to cards in the range 500-1000.

However, the threshold rules make up for their lower input-based robustness by being very robust with respect to changes in their parameters, specifically the threshold level, given a particular card value distribution. The one-threshold rule applied to card values 1-100 for instance scores above 1500 (94% of optimal) for thresholds from 58 to 90, and even with the mid-distribution threshold of 50 it still performs well above the best score of the cutoff rule for (reaching 1442, vs. 1391 for the cutoff rule). We can also see this robustness of the threshold rules in the performance comparison of participants' average self-reported threshold versus the optimal and modeled thresholds shown in Fig. 5: When applied to the actual card value data each participant saw, the optimal threshold strategy scored 1595, the average reported threshold (which looks quite different) came very close at 1589, and the modeled threshold scored in between with 1591. In contrast, the sample-based rules are not robust with respect to their sample size parameter—the performance of the cutoff rule for instance quickly falls as the sample size increases, to a score of 1302 (81% of optimal) at sample size 5 (cf. Todd & Miller, 1999).[6]

5.3. Model fits to participant data

[6] Note that the best sample size for the cutoff rule in the original version of the Secretary Problem with 20 options is 7, much larger than the best sample size in our current card search problem.

Simple threshold rules do very well on this task, but is that how people actually navigate it themselves? To compare how well the different models describe participants' choices, we used the Bayesian Information Criterion (BIC = $-2\,LL + k\log(N)$, where LL is the maximized value of the log likelihood function of the model, k is the number of free parameters in the model, and N is the number of observations of each participant). The BIC value was computed for each participant, and the mean over all participants is shown in Table 2. Models with smaller BIC values are preferred. For this fitting comparison, the stochastic version of each model was assessed, as described in Section 5.1 (along with a stochastic version of the optimal threshold model, using a strength parameter s and Equation 1). The best-fitting parameter values for each stochastic model were determined for each participant's data (using the exact sequences of values that each participant saw), and then the medians of these values were calculated and reported in Table 2.

As seen in the right-most column of Table 2, the threshold strategies (other than the optimal threshold) also achieved the best fit to participant data, with the linear decreasing threshold strategy having the lowest BIC score, followed closely by the two-threshold strategy. The best-fitting linear decreasing threshold strategy begins with a threshold of 80, and then lowers the threshold by nearly 1.8 on each successive turn. This strategy (along with the rapidly decreasing best-fitting six-threshold strategy, shown in Fig. 5) supports the idea that participants generally may have been lowering a threshold quickly as turns progressed, enabling them to begin long-term exploitation early (around turn 7, as shown in the row for participant performance), though not as early as the optimal strategy (around turn 5). These two best-fitting strategies both end with thresholds below 50 (which is always inappropriate), but this would

have little effect on overall performance because very few explore choices are made in the last few turns.

Surprisingly, the cutoff strategy, which performs comparatively poorly on this type of search, fit participants' data well, close to the BIC fit of the threshold strategies[7]. To explore the possible distribution of strategies used in more depth, we analyzed which strategy best fit the data of each participant at the individual level (in terms of lowest BIC). As seen in Table 3, the model which achieves the best fit for all the participant data together, the linear decreasing threshold strategy, also fits the most individual participants best. But the cutoff strategy fits the second largest group of participants individually. The other threshold rules best describe a few more participants, while the other sample-based rules and the random rules fit almost nobody best. Thus, there appears to be variation in strategy use across individuals, with the majority following a decreasing threshold strategy that performs very well on this type of search problem (and enables them to achieve a mean performance of 1545 points), and others using a cutoff strategy more appropriate to optimal stopping problems like the Secretary Problem (which yields 1491 points in this task).

---------- Insert Table 3 about here ----------

[7] To examine the role of noise in models for these results, we also tested how well a version of the cutoff rule with stochasticity from the logistic function in Equation 1 could fit the participants' choices, using a threshold of 100 for the first $k-1$ steps and after that using a threshold of the highest value seen in the first $k-1$ steps. This model fit the data very poorly (BIC=599.6, k=1, s=.05; all values are medians), suggesting that the lower relative performance of the sampling models is not due solely to a trembling hand version of noise. Similarly poor fits to human data by this kind of probabilistic cutoff rule in an optimal stopping Secretary Problem were found by Baumann et al. (2019).

6. Learning effects

Although participants do not know or exactly follow the optimal strategy, their performance comes quite close. How does this happen—do participants start the task with a good strategy and use it consistently across the 30 games, or do they start with lower performance and learn and improve over time, perhaps based on the feedback they are given after each game?

To find out, we divided participants' data into three parts according to games. Data from games 1 to 10 form the first part; from games 11-20 are the second (middle) part; and from games 21-30 are the third (last) part. For each part, we analyzed the number of switches per game, the initial turn of persistent exploitation (i.e. the first exploitation turn that was not followed by an exploration turn within the game), and task performance. The means for each measure are shown in Table 4, along with the means of the measures for the optimal rule applied to the last 10 games of card value sequences that participants saw (as in section 4). The three sections of games were treated as three different levels of the independent variable in a repeated-measure one-way ANOVA for each of these three measures. As shown in Table 5, all of the measures—mean switching, start of exploitation, and performance—improved appreciably from the first to the last 10 games, more closely approximating the optimal strategy over time. The frequency distributions of the initial turn of persistent exploitation for the three parts are shown in Fig. 7, along with the optimal threshold's distribution. Over games, the participants' distributions become more similar to the optimal one. These results support the presence of learning in terms of reduced switching between exploring and exploiting, and exploiting high-valued cards earlier, yielding improved performance as well.

---------- Insert Table 4 about here ----------

---------- Insert Table 5 about here ----------

---------- Insert Figure 7 about here ----------

Finally, the best-fitting linear decreasing threshold model was found for each of the three parts of the results to determine how participants' thresholds change over games. The median fitted parameter values for the linear decreasing threshold models that best fit the first 10, middle 10, and last 10 games for all 191 participants are shown in Table 6 and the corresponding decreasing thresholds are plotted in Fig. 8.

---------- Insert Table 6 about here ----------

---------- Insert Figure 8 about here ----------

These parameter values show that participants on average acted as though they used a high threshold for card values at the beginning of each game (being as choosy as the optimal strategy, with intercept *b* around 83), and lowered their threshold as the game continued (negative slope *m*; both aspects are also seen in the best-fitting six-threshold model in Fig. 5). Participants also changed their behavior over the course of the experiment, lowering their threshold more quickly in later games (lower values of *m*). Further, they applied their changing thresholds more

consistently, more than doubling their strength parameter *s* from the first to the last games. Hence, overall these results show support for participants learning to adjust their behavior across games in this search task in a way that improves their performance to become closer to optimal levels. The 38 participants best fit by the cutoff model also showed considerable learning across games with respect to improving parameters and fit of the cutoff model—see Table S6 in the Supplemental Materials file.

7. Discussion

When people face the common search problem of needing to choose among a succession of reward-providing options in a given period of time, the best approach is to start off being quite selective, passing by any options that are not among the best until a good-enough one is found to stick with, gradually becoming less choosy if no satisfactory option is found, and more rapidly decreasing in choosiness near the end of the time period. Many people appear to adopt a simple linear decreasing threshold (choosiness) strategy in our laboratory version of this problem, and learn to improve their threshold strategy with experience, even if they report doing something different. Starting off choosy and dropping one's threshold once after some time has passed (or even maintaining the single choosy threshold) can do almost as well. People are thus able to search effectively in this setting by selecting between exploring for possibly better new resource options and exploiting a good-enough option that has already been found.

Why do these simple rules work well for this task? The most important component of an effective strategy for this kind of search problem seems to be to start exploiting a relatively good

(high-value) option early; most of the time (i.e., turns) can then be spent repeatedly accruing that good (exploited) value rather than collecting random returns from exploring. Thus, this type of search calls for a faster transition to exploitation than for many optimal stopping problems, because of the importance of accumulated payoffs here versus the one-time reward from a single chosen value as in the Secretary Problem. (For instance, the mean point at which the optimal strategy for the card search task switches to exploitation is within 6 turns, while for the equivalent Secretary Problem with 20 options the optimal cutoff strategy leads to stopping and exploiting after around 14 turns.) Accordingly, even a simple fixed one-threshold rule does very well on this task, typically leading to an effective early switch to exploiting. Only if a searcher gets to the rare case of still exploring in the last few turns should they rapidly lower their threshold for what values to exploit. Such cases of extended exploration become increasingly unlikely, though, for longer search lengths—as the task gets longer, there are more chances along the way for a good option to have appeared—so the possible advantage of decreasing one's threshold at the end of the search shrinks, again supporting the appropriateness of a fixed threshold. In contrast to this "get out quick" approach, the optimal exploration time for the Secretary Problem grows essentially linearly with the overall length of the search task.

This difference in best strategies arises from the distinct natures of the exploration/exploitation tradeoff for the two types of search: In the resource-accumulating search of this card task, more exploration slowly increases the highest value that could be exploited but also rapidly decreases the total amount achievable by exploitation, while in the optimal stopping Secretary Problem, more exploration rapidly increases the chance of setting an aspiration level that could select the highest value to exploit, but slowly decreases the likelihood of being able to exploit that value because it may have been passed by already (e.g., Todd & Miller, 1999, Fig.

13-1). A threshold rule that quickly switches to exploitation makes the appropriate tradeoff for the former accumulative search problems, while a cutoff rule that explores more extensively balances better against exploitation in the latter type of optimal stopping problems.

In line with the performance advantages of simple threshold rules in this resource-accumulating type of search problem, the behavior of a majority of participants in our experiment (60%) was best fit by the linear decreasing threshold strategy. This type of decreasing threshold strategy is also supported by the pattern of response times found for continuing, and switching between, exploration and exploitation. Furthermore, a decreasing threshold implies some rate of returning to exploit previously-explored values, which we saw in the data: Participants exploit earlier values on 27% of games. (In comparison, the optimal strategy goes back to exploit an earlier card rather less often, on 531 out of the 5730 games, or about 9% of the time; out of these "returning" games, participants also went back to the same card in 88 cases, or 1.5% of all games, but they missed the optimal point to return to a previous card in the other 443 cases.) The fact that most participants were best fit by a decreasing threshold rule rather than the simpler and nearly equally performing one-threshold rule may indicate a general expectation that one should get less choosy as a deadline approaches, as suggested for various search contexts including mate choice (e.g. Cohen et al., 2015.) In another version of this task with a known variable game length of 5-10 turns, Song, Bnaya, and Ma (2019) also found that participants' exploitation decisions were fit well by a linear decreasing threshold rule, and showed that in this case the thresholds declined in proportion to the length of the game (slower for longer games) in a reasonable way.

Somewhat surprisingly, though, the second-largest group of participants (20%) was best fit by the cutoff rule, better suited to the optimal stopping Secretary Problem (though in this case

the initial sample period was very short, just 2 turns, leading to a relatively early switch from exploration to exploitation). Perhaps these participants do not recognize or appreciate the differences between the current search task setting and that of an optimal stopping problem like the Secretary Problem and so they apply a type of rule that is appropriate and commonly used for the latter problem here as well (without suffering too much in performance in the current task). In contrast, Baumann and colleagues (Baumann, Gershman, Singmann, & von Helversen, 2019) found that a linear threshold rule also best fit participants' behavior in a version of the Secretary Problem with well learned distributions of actual-value rewards, while a cutoff model failed to fit anyone well. This points to the possible importance of knowledge of the distribution of rewards for what rules people use. Testing what rules best fit the behavior of people facing other types of search tasks will help to clarify these issues.

One of the most marked differences between our participants' behavior and the optimal rule was that participants explored more than was optimal in two senses: They too often chose an unknown card from the deck rather than a known card from the table, and they switched back and forth between exploration and exploitation too often (i.e., more than once). Both of these behaviors could be accounted for by participants treating our static environment paradigm as a dynamic environment, in which the distribution of card values might change over time, and consequently exploring more (Navarro et al., 2016; Tversky & Edwards, 1966; Zhang & Yu, 2013). In a dynamic environment, expected values for actions change over time. Greater exploration is predicted and observed (Speekenbrink & Konstantinidis, 2015) for dynamic environments because decision makers need to make periodic checks on uncertain resources to assess whether their value has changed. In fact, people are highly sensitive to the informational requirements of dynamic environments, approaching optimal behavior in some settings (Brown

& Steyvers, 2009). The possibility that our participants initially had the wrong assumption about the static versus dynamic nature of the experimental paradigm is consistent with their gradually improving performance with experience (including fewer switches from exploitation back to exploration—Table 4), and with similar improvements found in other related experiments (Navarro et al., 2016; Rakow, Newell, & Zougkou, 2010).

Indeed, we found that people can perform better (closer to optimal) over time via learning based on minimal feedback. Notably, people did not get any feedback about whether particular choices followed the optimal strategy, only whether they had done better, worse, or the same as the optimal strategy at the end of each 20-turn game. Participants' mean total points per game, number of switches between exploring and exploiting, and number of turns before initiating persistent exploitation became closer to those of the optimal strategy as they completed more games of searching. (Interestingly, they appeared to improve by learning to explore less, while in the Secretary Problem studies of Seale and Rapoport, 1997, participants improved by learning to search more—again pointing to a fundamental difference between these two tasks.) The best-fitting linear decreasing threshold model of participants' behavior also moved toward the optimal solution across games by becoming less choosy sooner, and more deterministic and consistent with respect to the specified thresholds. This learning leads to a final model that achieved a cumulative score quite close to that attained by the optimal strategy (linear decreasing threshold model performance = 1588 over the last 10 games seen by participants, while optimal model performance = 1595 for the same last 10 games)—even though that final, best-fitting model has a simple linear decreasing shape, different from the accelerating fall-off seen in the optimal threshold. It could be that the participants employ a learning process that is more adept at constructing a simple rule of this linear form, involving few parameters, than what optimal

performance calls for; however, in this setting at least, performance hardly suffers as a consequence. Whether or not participants are actually using a threshold to determine their choices, we can at least say that the threshold rule does an increasingly good job of predicting participants' choices as they gain experience with the card search task.

Participants themselves mostly did not report that they thought they were using a threshold that decreased over turns within a search game. When asked to explicitly state their thresholds, the majority reported that they changed in the opposite direction of the modeled and optimal thresholds, rising (slightly) across turns. This mismatch between reported and modeled thresholds may have arisen because participants were not using a threshold-based mechanism to make their decisions (though the linear decreasing threshold model did best fit most participants' behavior). The mismatch could also be a consequence of participants not remembering or incorrectly reporting whatever thresholds they might be using, or of reporting what value they expect to be able to obtain after a particular number of turns. Or participants may have introspected little and just reported that their threshold should increase as the turns increase. Such use of a linguistic frame in participants' responses, in which one variable (e.g., threshold) increases as another variable (e.g., turns) increases, could potentially be alleviated in future work by rewording the threshold-reporting question and asking it repeatedly between games. In any case, participants probably do not explicitly know what is optimal nor what they may actually be doing, as is often found in decision-making tasks (Nisbett & Wilson, 1977), but they still improve, getting closer to optimal, over the course of learning.

8. Conclusion

In this paper we have investigated situations in which people searching for resources that they accumulate over time must decide how to allocate their effort between exploring for high-value resources and exploiting those resources once they have been discovered. We found evidence for people using a simple linear threshold mechanism to determine when to switch between exploration and exploitation in a card search task incorporating such accumulating resources. With experience, our participants improved their search behavior and approached the performance of the optimal strategy, apparently by adjusting their linear threshold strategy appropriately.

Previous work on search has typically focused on only one aspect of the explore/exploit tradeoff or a single transition between exploration and exploitation. For example, studies of optimal stopping problems including the Secretary Problem have mainly looked at how long to explore before stopping (and making the single switch to exploitation). In the field of decision making, the bandit problem (e.g., Steyvers, Lee, & Wagenmakers, 2009; Lee, Zhang, Munro, & Steyvers, 2011) and the repeated-choice paradigm for studying decisions from experience (Gonzalez & Dutt, 2011) have been argued to rely on related cognitive processes that make a transition from an exploration to an exploitation phase. Future research using a search task that instead calls for repeated transitions between exploration and exploitation (as in foraging among patches) may present a different picture of the search mechanisms people typically use. Such a task involving multiple explore/exploit tradeoff decisions over time could also make individual differences in search behavior easier to observe, along with their correlations with other measures such as working memory and impulsivity.

In the current card search task, participants made relatively few transitions from

exploitation back to exploration in part because the exploited card values were non-depleting—once participants found a sufficiently high card value, they could have been sufficiently satisfied with that value to exploit it until the end of that game (which is also the pattern of behavior that the optimal strategy dictates). To induce more exploit-to-explore transitions in this task, we can make the card values depleting so that every time a particular card gets exploited, its value will decrease by a certain amount. Moreover, changing the number of turns in each game from a known fixed length to a random length may have a similar effect of increasing participants' switches back and forth between both phases of search (see Sang, 2017, for examples of both of these manipulations of the card search task). New search models may be needed to predict these switches, including mechanisms combining stochasticity with inertia or momentum, which can lead to alternating stretches of exploration and exploitation.

Open questions beckon regarding how individual differences in the tendency to explore versus exploit play out across different search settings, including social search and information search on the Web, as well as measuring priming effects between settings (Hills et al., 2015; Hills, Todd, & Goldstone, 2008). Finally, different clinical populations may make the explore/exploit tradeoff in different ways, emphasizing one aspect of search over the other (Hills, 2006), and fMRI could also be useful in exploring these differences and providing insights into the neural mechanisms used in search across different domains. By stripping search down to a setting where exploration and exploitation are most prominent, the card search task and its possible variations may help us elucidate the decision strategies underlying search more effectively.

Acknowledgments

We thank Ross Branscombe, Jerome R. Busemeyer, and Woo-Young Ahn for their help with this research. We also acknowledge the support of National Science Foundation REESE grant 0910218. Part of this work was supported by the John Templeton Foundation grant, "What drives human cognitive evolution," to the second author.

Figure captions

Figure 1. A screen shot of the card search experiment. In the lower-left corner is the deck of face-down cards that can be explored, while the previously found cards (here, four so far out of up to 20 possible) that can be further exploited for their points are displayed in the upper portion of the screen. The highest score face-up card is highlighted in red. Also displayed near the bottom of the screen are the number of cards left in the deck, number of turns taken and left, total points accrued so far, and the values of all cards already selected at each turn (in brackets)—here, the value 91 has been exploited several times.

Figure 2. Threshold curve for the optimal strategy: For any turn (on *x*-axis), if the highest card value found so far exceeds the optimal threshold for that turn, then that card should be exploited for all remaining turns in the game.

Figure 3. General framework of the experiment, showing a single game (trial) consisting of 20 turns of exploring or exploiting, with 30 games overall, followed by a strategy questionnaire regarding the thresholds that participants used.

Figure 4. Frequency distributions of initial turns for final persistent exploitation phase, for participants (left) and the optimal strategy applied to the same data participants saw (right).

Figure 5. Mean reported and modeled thresholds (interpolated between the 6 indicated turns)

along with optimal thresholds for switching from exploration to exploitation, across turns. Error bars are 1 SEM.

Figure 6. Mean probability at each turn that participants would exploit the highest-available card if its value fell in various ranges (indicated by different lines). Error bars omitted for clarity; values for later turns are based on fewer data points.

Figure 7. Frequency distributions of initial turns for final persistent exploitation phase, divided into first, middle, and last 10 games for participants, along with the distribution for the optimal strategy applied to the same data participants saw across all games (i.e., the same distribution as in Fig. 4, here plotted with one third of the data so that the y-axis matches the other graphs). Participants' persistent exploitation commences earlier as they complete more games, and approaches the optimal distribution.

Figure 8. Best-fitting modeled threshold curves found for the first, middle, and last 10 games from the linear decreasing threshold model applied to all 191 participants.

Tables

Table 1

Mean and Standard Deviation for Response Times of Different Decision Types

Type of decision (and number)	Mean residual response time, s	*SD*
Continue to explore (22,335)	1.41	2.35
Continue to exploit (76,021)	0.59	0.41
Switch to explore (2,538)	1.14	1.34
Switch to exploit (7,976)	1.24	0.88

Table 2

Comparison of Parameters for Best Performing and Best Fitting Models Across Strategies

	Best performing model				Best fitting model to data			
Strategy	Score per game	Best parameter values	Switch turn (and % switching)	Exploited card value	Best fit parameter values	Best fit error parameter	Number of parameters	BIC
Participant performance	1528.0	NA	7.64 (94.9%)	86.12	NA	NA	NA	NA
Optimal	1601.8	NA	5.51 (100%)	88.86	NA	$s = 0.12$	1	431.5
Epsilon-greedy	1312.0	$\varepsilon = 0.34$	NA	74.84	$\varepsilon = 0.21$	NA	1	596.5
Random switch	1318.9	NA	7.62 (95.3%)	73.84	NA	$s = 0.21$	1	454.8
One-threshold	1599.0	$T = 79$	5.50 (99.0%)	89.03	$T = 68.3$	$s = 0.132$	2	378.3
Linear decreasing threshold	1601.7	$m = -0.58$ $b = 81$	5.47 (99.9%)	88.80	$m = -1.78$ $b = 80.65$	$s = 0.12$	3	**326.4**
Two-threshold	1601.1	$T_1 = 80$ $T_2 = 75$ $k = 8$	5.44 (99.7%)	88.81	$T_1 = 77.1$ $T_2 = 57.7$ $k = 7$	$s = 0.126$	4	335.7
Six-threshold	1601.7	$T_1 = 82$ $T_2 = 81$ $T_3 = 79$ $T_4 = 76$ $T_5 = 71$ $T_6 = 58$	5.55 (100%)	88.86	$T_1 = 82.0$ $T_2 = 77.7$ $T_3 = 69.9$ $T_4 = 62.5$ $T_5 = 54.1$ $T_6 = 44.1$	$s = 0.13$	7	346.8
Fixed sample	1495.7	$k = 6$	7.0 (100%)	85.41	$k = 4$	$h = 0.42$	2	445.6
Cutoff	1391.6	$k = 2$	6.60 (90.4%)	80.07	$k = 2$	$h = 0.095$	2	389.5
Successive non-candidate count	1469.9	$j = 3$	7.29 (99.99%)	84.29	$j = 1$	$h = 0.46$	2	592.7

Notes: See text for explanation of all parameters. All values for the best performing models are means calculated over 50,000 random games. Switch turn shows the mean turn on which the model switched to final persistent exploitation as in Fig. 4 (followed by the percent of games in which a switch to exploitation was made). For the best fitting models, the parameter values are medians over the best fitting values for each participant using the same sequence of card values that each participant saw. Larger *h* (trembling hand) indicates more stochasticity, while larger *s* (strength) indicates less stochasticity. BIC values for best fitting models are means; the linear decreasing threshold model has the lowest BIC (shown in bold).

Table 3

Number of Participants Whose Choices are Best Fit by Each Strategy

Strategy	Number of participants	Participants' mean score
Linear decreasing threshold	117	1545
Cutoff	38	1486
Two-threshold	24	1531
One-threshold	9	1532
Random switch	2	1415
Fixed sample	1	1306
Six-threshold	0	NA
Epsilon-greedy	0	NA
Successive non-candidate count	0	NA

Table 4

Mean Switching, Start of Exploitation, and Performance Across Games (with SD)

Games	Number of switches	Initial turn of persistent exploitation	Performance
First 10	2.57 (1.96)	8.77 (3.63)	1492 (112.3)
Middle 10	1.54 (1.20)	6.90 (2.80)	1539 (102.6)
Last 10	1.39 (1.00)	6.39 (2.13)	1553 (91.5)
Optimal	1.0 (0.0)	5.08 (1.04)[a]	1595 (70.6)[a]

[a] These values differ from those in Table 2 because here they are calculated over the last 10 games of card value sequences that participants saw, while in Table 2 they are calculated over 50,000 random sequences.

Table 5

Repeated-measure One-way ANOVA of Mean Switching, Start of Exploitation, and Performance Across Games

	df	*F*	*p*	η^2
Number of switches	(2, 380)	78.78	< 0.001	0.29
Initial turn of persistent exploitation	(2, 380)	84.14	< 0.001	0.31
Performance	(2, 380)	22.44	< 0.001	0.11

Table 6

Median Fitted Parameter and BIC Values (with 95% Confidence Intervals) for Best-fitting Linear Decreasing Threshold Model Across Games

Games	Slope *m*	Intercept *b*	Strength *s*	BIC
First 10	−1.29 [−1.59, −1.00]	82.75 [79.31, 84.71]	0.11 [0.09, 0.13]	133.36 [120.36, 147.82]
Middle 10	−1.68 [−1.96, −1.22]	82.65 [81.02, 84.29]	0.25 [0.21, 0.28]	68.53 [60.31, 79.82]
Last 10	−1.93 [−2.27, −1.54]	83.10 [81.29, 84.84]	0.28 [0.23, 0.34]	55.56 [49.56, 64.16]

References

Baumann, C., Gershman, S.J., Singmann, H., & von Helversen, B. (2019). A linear threshold model for optimal stopping behavior. Preprint: https://doi.org/10.31234/osf.io/cbn6t

Beachly, W. M., Stephens, D. W., & Toyer, K. B. (1995). On the economics of sit-and-wait foraging: Site selection and assessment. *Behavioral Ecology*, *6*(3), 258-268. doi: 10.1093/beheco/6.3.258

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology*, *58*, 49–67.

Charnov, E. L. (1976). Optimal foraging: The marginal value theorem. *Theoretical Population Biology*, *9*(2), 129-136. doi: 10.1016/0040-5809(76)90040-x

Christian, B., & Griffiths, T. (2016). *Algorithms to live by: The computer science of human decisions*. New York: Henry Holt and Company.

Cohen, S.E., and Todd, P.M. (2018). Relationship foraging: Does time spent searching predict relationship length? *Evolutionary Behavioral Sciences*, *12*(3), 139-151. http://dx.doi.org/10.1037/ebs0000131

Cohen, S. E., Todd, P. M., Garcia, J., & Fisher, H. (2015). Temporal reproductive pressures on human sexual strategies in a large, representative dataset. Poster presented at the Human Behavior and Evolution Society conference. Columbia, MO.

Ferguson, T.S. (n.d.). *Optimal stopping and applications*. Downloaded 4/1/2014 from http://www.math.ucla.edu/~tom/Stopping/Contents.html

Ferguson, T. S. (1989). Who solved the secretary problem? *Statistical Science*, *4*(3), 282-289. doi: 10.2307/2245639

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review, 118*(4), 523-551. doi: 10.1037/a0024558

Hills, T. T. (2006). Animal foraging and the evolution of goal-directed cognition. *Cognitive Science, 30*(1), 3-41. doi: 10.1207/s15516709cog0000_50

Hills, T. T., Kalff, C., & Wiener, J. M. (2013). Adaptive Levy processes and area-restricted search in human foraging. *PLoS One, 8*(4). doi: 10.1371/journal.pone.0060488

Hills, T. T., Todd, P. M., & Goldstone, R. L. (2008). Search in external and internal spaces: Evidence for generalized cognitive search processes. *Psychological Science, 19*(8), 802-808. doi: 10.1111/j.1467-9280.2008.02160.x

Hills, T. T., Todd, P. M., & Goldstone, R. L. (2010). The central executive as a search process: Priming exploration and exploitation across domains. *Journal of Experimental Psychology: General, 139*(4), 590-609. doi: 10.1037/a0020666

Hills, T.T., Todd, P.M., Lazer, D., Redish, A.D., Couzin, I.D., and the Cognitive Search Research Group (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Science, 19*(1), 46-54.

Hutchinson, J. M. C., Fanselow, C., & Todd, P. M. (2012). Car parking as a game between simple heuristics. In P. M. Todd, G. Gigerenzer, & the ABC Research Group, *Ecological rationality: Intelligence in the world* (pp. 454-484). New York, NY: Oxford University Press.

Hutchinson, J. M. C., Wilke, A., & Todd, P. M. (2008). Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches? *Animal Behaviour, 75*, 1331-1349. doi: 10.1016/j.anbehav.2007.09.006

Knox, W. B., Otto, A. R., Stone, P., & Love, B. C. (2012). The nature of belief-directed exploratory choice in human decision-making. *Frontiers in Psychology, 2,* 398. doi: 10.3389/fpsyg.2011.00398

Lee, M.D., Zhang, S., Munro, M.N., & Steyvers, M. (2011). Psychological models of human and optimal performance on bandit problems. *Cognitive Systems Research, 12,* 164-174.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., Hausmann, D., Fiedler, K., and Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision, 2*(3), 191-215. doi: 10.1037/dec0000033

Navarro, D. J., Newell, B. & Schulze, C. (2016). Learning and choosing in an uncertain world: An investigation of the explore-exploit dilemma in static and dynamic environments. *Cognitive Psychology, 85,* 43-77.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259. doi: 10.1037//0033-295x.84.3.231

Pirolli, P. (2005). Rational analyses of information foraging on the Web. *Cognitive Science, 29*(3), 343-373. doi: 10.1207/s15516709cog0000_20

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information.* New York, NY: Oxford University Press.

Rakow, T., Newell, B. R., & Zougkou, K. (2010). The role of working memory in information acquisition and decision making: Lessons from the binary prediction task. *The Quarterly Journal of Experimental Psychology, 63,* 1335–1360.

Sang, K. (2017). *Modeling exploration/exploitation behavior and the effect of individual*

differences (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses. (Accession NO. 10259864)

Sang, K., Todd, P.M., and Goldstone, R.L. (2011). Learning near-optimal search in a minimal explore/exploit task. In *Proceedings of the Thirty-third Annual Conference of the Cognitive Science Society* (pp. 2800-2805). Boston, MA: Cognitive Science Society.

Seale, D. A., & Rapoport, A. (1997). Sequential decision making with relative ranks: An experimental investigation of the "secretary problem". *Organizational Behavior and Human Decision Processes, 69*(3), 221-236. doi: 10.1006/obhd.1997.2683

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory, 4*(1), 25-55. doi: 10.1007/bf01766400

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*(1), 1-20.

Song, M., Bnaya, Z., & Ma, W.J. (2019). Sources of suboptimality in a minimalistic explore-exploit task. *Nature Human Behaviour, 3*, 361-368. DOI: 10.1038/s41562-018-0526-x

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science, 7*, 351–367.

Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology, 53*(3), 168-179. doi: 10.1016/j.jmp.2008.11.002

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA: MIT Press.

Todd, P. M., Hills, T. T., & Robbins, T. W. (2012). *Cognitive search : Evolution, algorithms, and the brain*. Cambridge, MA: MIT Press.

Todd, P. M., & Miller, G. F. (1999). From pride and prejudice to persuasion: Satisficing in mate

search. In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 287-308). New York, NY: Oxford University Press.

Tversky, A., & Edwards, W. (1966). Information versus reward in binary choices. *Journal of Experimental Psychology, 71,* 680–683.

Walton, M. E., Devlin, J. T., & Rushworth, M. F. S. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nature Neuroscience, 7*(11), 1259-1265. doi: 10.1038/nn1339

Zhang, S., & Yu, A. J. (2013). Cheap but clever: Human active learning in a bandit setting. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society* (pp. 1647–1652). Austin, TX: Cognitive Science Society.