Active Learning Strategies in a Spatial Concept Learning Game

Todd M. Gureckis (todd.gureckis@nyu.edu)
Doug Markant (doug.markant@nyu.edu)
New York University, Department of Psychology
6 Washington Place, New York, NY 10003 USA

Abstract

Effective learning often involves actively querying the environment for information that disambiguates potential hypotheses. However, the space of observations available in any situation can vary greatly in potential "informativeness." In this report, we study participants' ability to gauge the information value of potential observations in a cognitive search task based on the children's game *Battleship*. Participants selected observations to disambiguate between a large number of potential game configurations subject to information-collection costs and penalties for making errors in a test phase. An "ideal-learner" model is developed to quantify the utility of possible observations in terms of the expected gain in points from knowing the outcome of that observation. The model was used as a tool for measuring search efficiency, and for classifying various types of information collection decisions. We find that participants are generally effective at maximizing gain relative to their current state of knowledge and the constraints of the task. In addition, search behavior shifts between an slower, but more efficient "exploitive" mode of local search and a faster, less efficient pattern of "exploration."

Traditional experimental approaches to human learning tend to emphasize passive learning situations. For example, in a typical concept learning task, subjects are presented with examples one at a time, the order of which are selected by the experimenter (often at random and with exhaustive sampling of the training set). However, this procedure ignores the fact that real-world learning often requires learners to *actively* create their own learning experiences by constructing revealing queries or engaging in exploration of unknown contingencies (Nelson, 2005; Skov & Sherman, 1986; Sutton & Barto, 1998). For example, children might ask about particular objects in their environment and receive feedback from an adult (e.g., "What is that?", "What does that do?"). In order for such sampling behavior to be effective, queries should be directed to maximize the potential information that could be obtained from an answer. We need not ask about things that we already know, and, all else being equal, should prefer questions whose answers are expected to be most revealing (Klayman & Ha, 1987; Nelson, 2005; Oaksford & Chater, 1994).

In this paper, we present an initial study examining the mutual unfolding of information search and learning in a task modeled on the classic children's game *Battleship*. Participants attempted to learn a hidden "concept" (the shape and spatial configuration of three hidden rectangles) by sequentially uncovering points on a large grid (see Figure 1). Each observation cost points and participants were motivated to minimize the points accumulated in each game. As a result, they had to minimize the number of observations they made in order to correctly identify the hidden rectangle configuration.

In order to analyze participant's performance in such a complex and dynamic learning task, we develop a formal model of information search based on Bayesian learning principles. The model specifies how past observations should influence current beliefs, and how uncertainty should translate into information sampling behavior on a trial-by-trial basis. By comparing the utilities the model assigned to each possible observation with the selections that participants make, we are able to characterize the efficiency of participants' search strategies relative to an ideal learner who had experienced the same set of previous observations. In addition, the model allowed us to objectively classify particular information collection decisions as being either "exploitative" of known contingencies or "exploratory" of relatively unknown parts of the game board. We find interesting behavioral differences between these two modes of information search.

Active Sampling in Concept Acquisition

The ultimate goal of the present work is to understand how people actively seek information when acquiring new concepts. Previous work has shown that allowing learners to make their own decisions about what information to sample can have an impact on both the efficiency of learning (Castro et al., 2008) and what is learned (Fazio, Eiser, & Shook, 2004) in concept acquisition tasks. For example, Castro, et al. (2008) found that allowing learners to actively select training examples greatly improved the efficiency by which they learned a linear decision boundary. Fazio, Eiser, and Shook (2004) studied the impact of experiential sampling on category learning. Participants were presented with different "beans" and were asked on each trial if they would like to (virtually) eat the bean and find out if it was healthy or poisonous. Decisions to not eat a particular bean thus provided no information (i.e., feedback in the task was contingent on sampling decisions). Experiential learners were found to be risk averse, in that they showed a bias to think that novel beans were bad and were more accurate classifying bad beans than good beans. Interestingly, this asymmetry in learning only occurred in situations where learners made the sampling decisions themselves as opposed to conditions where full information was provided on each trial.

While allowing learners to make decisions about what they want to learn about can have interesting consequences for what is learned, these studies leave aside the question of exactly how people decide which observations to make. However, assessing the "usefulness" of potential observations has a long history of study in psychology, particularly in hypothesis testing situations (Skov & Sherman, 1986; Klayman &

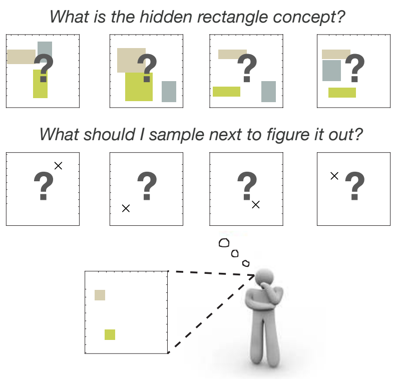

Figure 1: The information search problem studied in this report. Three rectangles of unknown sizes, shapes, and positions are hidden in a 10x10 grid. On each trial, participants make selections of which grid cell to turn over (revealing either a miss or a part of a hidden rectangle). The goal is to discover the true hidden concept by actively choosing which new observations to make. Efficiency is encouraged via a cost structure imposed on the task.

Ha, 1987; Oaksford & Chater, 1994). A classic finding is that participants often show a bias towards using confirmatory test strategies when testing between alternative hypotheses (Wason, 1960). While these original findings raised concerns about the intuitive reasoning abilities of humans, other researchers have developed probabilistic approaches which calculate the expected *information gain* (among other norms) of potential questions given an appropriate set of prior beliefs and find that such utilities can often account for participants' choices (Nelson, 2005; Oaksford & Chater, 1994).

Our work is inspired by these previous probabilistic approaches to hypothesis testing and information search as well as research on "active learning" algorithms in the machine learning literature which seek to optimize data selection for training artificial learners (Mackay, 1992). However, most studies in the hypothesis testing literature have focused on simple, one-trial judgments and reasoning tasks with a relatively constrained hypothesis space (often times there is only a small set of possible hypotheses participants are attempting to distinguish). In contrast, our goal is to understand how people search for information in complex, ongoing learning tasks where they must continually updated expectations based on past observations and use these expectations to drive new information-seeking behaviors.

The Rectangle Search Game

In the rectangle search game (see Figure 1), the player is presented with a 10x10 grid that contains three, non-overlapping hidden rectangles of different colors (i.e., the unknown concept). On each trial, players can choose to turn over one square in the grid, revealing either part of a hidden rectangle or a blank space. Each game is divided into two phases: an *information collection phase*, where participants make selections of grid points to uncover and receive feedback, and a *test phase*. In the first phase, participants are told that their goal is to discover the identity of all three hidden rectangles with the fewest number of observations possible. In order to formalize the costs of information collection, participants start each game with zero points, and each observation (i.e., choice to turn over a point on the grid) added one point to their score. Subjects could choose to end the sampling phase at any point by clicking on a button at the bottom of the screen which would begin the test phase.

In the test phase, subjects were presented with an empty grid, and were asked to "paint in" the correct position, shape, and color of each of the three rectangles using the computer mouse. Participants were informed (prior to the start of the task and at the beginning of the test phase) that each incorrectly colored grid point would cost two (2) additional points. Thus, errors were overall more costly than collecting additional samples. Following the painting phase, participants were shown their final score for that game which was a combination of the points incurred due to making observations in the first phase, and the points incurred for making errors in the test phase.

A Bayesian Search Model

In the following section, we describe a simple Bayesian model of the task. It is important to point out that the model we describe is not meant to mimic the specific cognitive strategy that participant's use while learning. Instead, our goal is to formally specify the behavior of an "ideal" learner who searches for information under the cost structure imposed by the task, and to use this model as a tool for understanding human performance[1].

In formal terms, players in the game are presented with a NxN grid of squares and are asked to sequentially make observations in order to learn the identity of the hidden game board concept, $g_{hidden} \in G$, where G is the universe of legal game boards. Each individual game board is defined by a set of three, non-overlapping rectangles, $\{r_1^g, r_2^g, r_3^g\}$, and each individual rectangle r_n is denoted by a quadruple (x_n, y_n, w_n, h_n) where x_n and y_n are the coordinates of the top left corner of the rectangle in the grid and w_n and h_n are the width and height, respectively.

Learning On each trial, the player selects a single square in the grid, x_{ij}, and receives feedback about if it belongs to r_1, r_2, or r_3, or isn't part of any rectangle. We denote the feedback (or observed label) as l_n where l_0 means that the observed point is empty, l_1 means it falls within rectangle r_1, and so on (for short-hand, we simply denote a sampled location and its associated label as $x_{ij} = l_n$). Since each point in the grid is assigned to either one or zero rectangles and this

[1] In this sense our model provides a "rational analysis" of the task. However, we are unable at the current stage of this work to call our model the complete rational solution. One reason is that (for computational reasons) the current model assumes that participants always choose the option with the highest expected saving on each trial (i.e., they assume the game will end on the next trial). It is possible that participants engaged in multi-step planning which may change the model's valuation of particular observations, an issue we hope to evaluate in future work.

assignment is deterministic, we assume that the likelihood of a particular observation and associated label given a particular game board configuration is given by:

$$p(x_{ij} = l_n|g) = \begin{cases} 1 & \text{if } x_{ij} \in r_n^g, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

for $n > 0$ (see Tenenbaum, 1999 for a similar formulation in a similar task). Alternatively, if $x_{ij} = l_0$ then,

$$p(x_{ij} = l_0|g) = \begin{cases} 0 & \text{if } x_{ij} \in r_1^g \text{ or } x_{ij} \in r_2^g \text{ or } x_{ij} \in r_3^g, \\ 1 & \text{otherwise.} \end{cases} \qquad (2)$$

This captures the basic intuition that empty squares provide support for hypotheses that *don't* place the observation within a rectangle, while hits provide evidence for hypotheses that *do*.

The prior belief about the likelihood of each individual game board is represented by $p(g)$. In our experiments, participants were instructed that rectangles were chosen at random and that each possible game board configuration was equally likely (i.e., $p(g) = 1/|G|$ for all g, a uniform prior). This prior, along with a piece of new data, $(x_{ij} = l_n)$, gives us the following posterior belief about the identity of the hidden board according to Bayes rule:

$$p(g|x_{ij} = l_n) = \frac{p(x_{ij} = l_n|g)p(g)}{\sum_{g_h \in G} p(x_{ij} = l_n|g_h)p(g_h)} \qquad (3)$$

On each trial, the posterior belief $p(g|x_{ij} = l_n)$ following from the last observation is used as the new prior allowing for incremental updating of our beliefs with each new observation. After each update, the new prior is also used in predicting the label associated with any point y_{ij} in the grid. The predicted probability that location $y_{ij} = l_k$ (for $k \in 0,1,2,3$) given our current belief $p(g)$ is:

$$p(y_{ij} = l_k) = \sum_{g_h \in G} p(y_{ij} = l_k|g_h)p(g_h) \qquad (4)$$

Assessing the Value of Future Observations We now consider how agents might use their beliefs at any point to select the best new samples to learn about (i.e., active learning). In our experiment, participants were given the explicit goal of minimizing the number of points they accumulated during each game, where each individual observation cost them C_{obs} point, each error during painting/recall cost C_{miss} points, and each correctly colored square cost C_{hit} points. Given these costs, we can quantify the value of particular observations with respect to the overall objective of minimizing accumulated points.

Formally, we assume that the goal of the learner on each trial is to select the observation x_{ij} that minimizes the expected points that would be incurred during the recall phase if the game were to end after that observation was made. With costs C_{miss} and C_{hit} as defined above, we compute the expected cost $EC(G)$ given the current beliefs as:

$$EC(G) = \sum_i \sum_j \sum_{n=0}^{4} p(x_{ij} = l_n|G) \cdot [C_{hit} \cdot p(x_{ij} = l_n|G) + C_{miss} \cdot (1 - p(x_{ij} = l_n|G))] \qquad (5)$$

which simply says that the cost associated with painting square $x_{ij} = l_n$ is related to the probability that we currently believe that square x should be painted color l_n (or not) times the cost associated with being either correct or incorrect. These expected costs are then weighted by our overall estimate that the true state of affairs is that square x_{ij} actually is colored with label l_n.

We can also calculate the saving (S) from making any observation as:

$$S(G, x_{ij} = l_n) = EC(G) - [EC(G|x_{ij} = l_n) + C_{obs}] \qquad (6)$$

and the expected savings (ES) from observation x_{ij} is found by calculating a weighted average of the savings based on our current belief about the possible labels associated with x_{ij}:

$$ES(G, x_{ij}) = \sum_{n=0}^{4} p(x_{ij} = l_n|G)S(G, x_{ij} = l_n) \qquad (7)$$

The choice that maximizes expected savings is predicted to be the best choice on any trial according to the "greedy" strategy[2].

Classifying Search Behavior as "Exploration" or "Exploitation" In addition to predicting which observations participants should make on each trial, the model provides a simple way to classify each observation as either `Exploit`ing local evidence for a rectangle or `Explore`-ing relatively unknown regions of the board. Prior to making a "hit" (i.e., an observation that reveals a rectangle rather than empty space), all observations are treated as `Explore`. Following a hit for rectangle r_n, we can compute the posterior probability that each grid location belongs to rectangle r_n using Equation 4. Observations where the actual sample, x_{ij}, match the constraint $0 < p(x_{ij} = l_n) < 1.0$ suggest targeted attempts to decrease uncertainty about rectangle r_n and are considered `Exploit`. In contrast, actions where $p(x_{ij} = l_n) = 0$ and $0 < p(x_{ij} = l_k) < 1.0$ for any rectangle r_k that has not yet been discovered are `Explore`. Additionally, observations may be classified as errors if they fail to resolve any uncertainty according to the model (i.e., sampling where $p(x_{ij} = l_n) = 1$ for one rectangle, or $p(x_{ij} = l_n) = 0$ for all rectangles).

The Experiment

Participants and Apparatus Six undergraduates at New York University participated in the study to fulfill part of a class requirement. The experiment was run on standard Macintosh computers over a single session.

[2]Our model ties sampling behavior to the cost structure imposed on the task. However, another solution we considered but do not report (due to space) assumes that agents to make observations that provide the greatest reduction in uncertainty about the game board in information theoretic terms (Kruschke, 2008; Nelson, 2005; Oaksford & Chater, 1994).

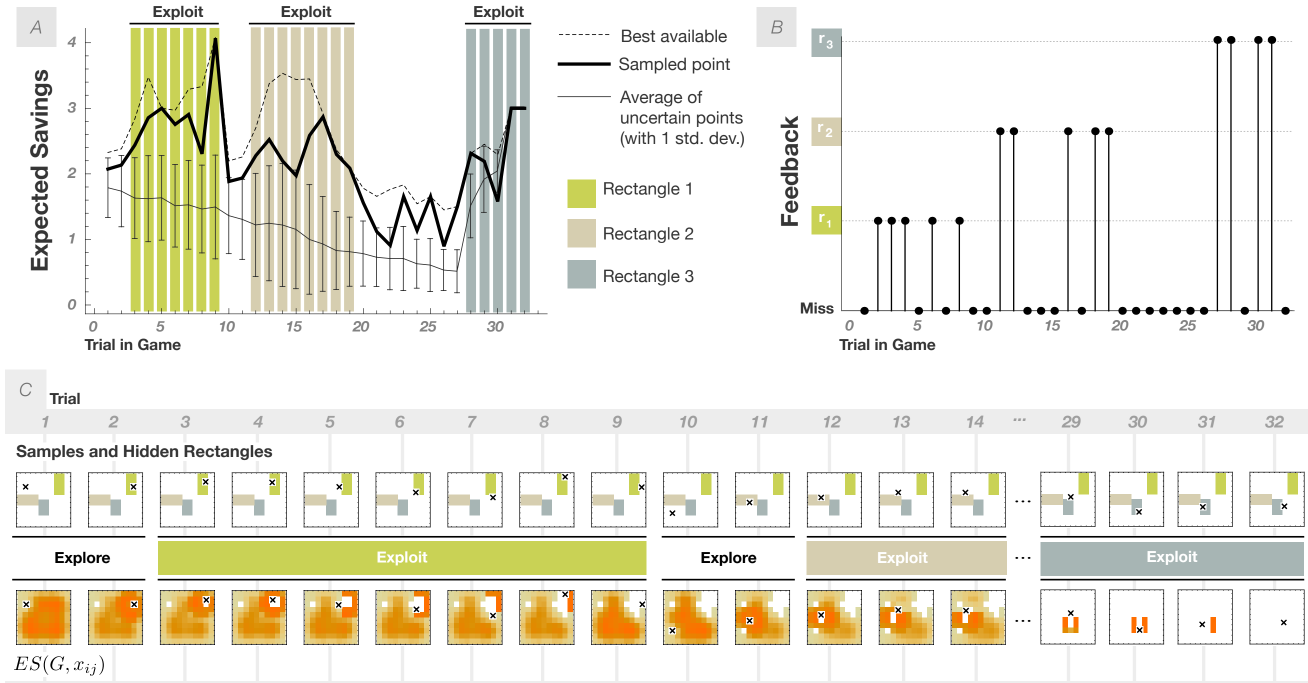

Figure 2: Comparison of a typical human search pattern in a single game along with the predictions of the model (subject 3, game 1). The solid line in **Panel A** reflects the expected savings assigned to the participant's sampling choices (Eqn. 7). The top dashed line demonstrates the maximum expected savings that could be obtained from any choice on a given trial. The participant's choices were maximally informative if the solid line matches this upper bound. In order to determine a lower bound, we computed the average expected savings assuming that the participant chose randomly among the remaining uncertain grid points (lower line, along with one std.dev. of that mean). The shaded regions indicate choices the model classified as `Exploit`, while the unshaded regions reflect `Explore` trials. Regions are color coded according to which rectangle is being exploited. **Panel B** shows the feedback the participant received on each trial. Finally, **Panel C** plots individual observations super-imposed on the hidden game board (top row) along with the dynamic evolution of $ES(G,x_{ij})$ over the entire grid as evidence accumulates (bottom row). In both sequences, the small black 'x' shows the participant's sample. The shaded regions in the top row show the hidden rectangles for this game. In the bottom row, darker colors indicate that the model predicts higher expected savings from selecting that point. Note how the model's estimate of the most useful information accumulates in the regions surrounding hits (e.g., trials 2-8 or 11-14). See http://smash.psych.nyu.edu/projects/activelearning/ for some supplementary information including movies.

Procedure At the start of the task, participants were given on-screen instructions detailing the rules and objectives, followed by a single practice game. In addition, participants were shown a set of 50 randomly generated legal gameboards on the screen to help develop an appropriate prior expectation about what they might encounter in the task. Participants also completed a questionnaire which tested for understanding of the game. After the experimenter was confident that participants had a complete understanding of the task, the experiment began.

For the remainder of the session, participants played a sequence of games at their own pace. In order to facilitate between-subject comparisons, the sequence of games experienced by each player was identical. Each game-board contained three hidden rectangles that were drawn randomly from a fixed set of eight possible sizes (1x3, 3x1, 1x4, 4x1, 2x2, 3x2, 2x3, 3x3, 2x4, 4x2, 3x4, 4x3) and were placed on the grid with the constraint that the entire rectangle laid inside the boundaries of the grid, and there were no overlapping rectangles As a result, there were 563,559,150 possible game boards that the participant had to distinguish in order to identify the true game board.

During the information collection phase, participants made observations by clicking on a grid point, after which the point changed color according to its category membership (rectangle 1=red, rectangle 2=blue, rectangle 3=green, or no rectangle=grey). Throughout the entire information collection phase, a visual reminder of the possible shapes and size of the rectangles was provided on screen to the right of the current gameboard. After each observation, the subject's total number of samples was incremented and displayed at the top of the screen as a reminder. All previous observations remained on the screen throughout the information collection phase. There were no time constraints on sampling, and subjects could choose to end the first phase at any point by clicking on a button at the bottom of the screen which would begin the test phase (described above).

The memory demands of the test phase were significant. However, the game was self-paced and thus, prior to terminating search participants could spend as much time as they wanted memorizing the layout of the uncovered rectangles. Overall, we were less interested in participant's performance during the test phase, as much as using that phase to set up an incentive for participants to actually discover the true board during the search phase.

Results

Participants completed a variable number of games before the end of the session; however, in our initial analysis (and due to the computational complexity of computing the full model solution of each trial of every game) we considered only the first twenty games that everyone completed (a total of 120 unique games). On average, participants made 34.4 (SD=3.6) observations and 1.8 sampling errors (observations that were redundant given their current knowledge) per game.

Figure 2 presents the data from a typical game and gives some intuition for the dynamics of the model (see the caption for a full description). In particular, Figure 2A shows the overall expected savings of a participant's choices compared

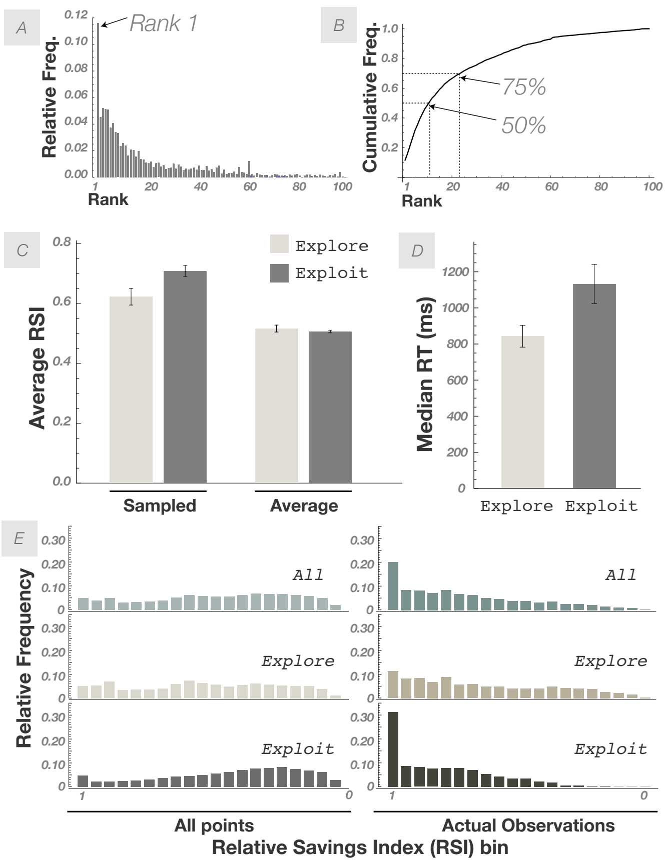

Figure 3: **Panel A** shows the frequency of participants choices as a function of the rank-order $ES(G,x_{ij})$ on each trial. **Panel B** shows the same data in cumulative distribution form. **Panel C** shows an increased relative savings index (RSI) for participants' observations (left) as compared to a random strategy (right) considered separately for `Exploit` and `Explore` trials. **Panel D** shows a corresponding increase in median reaction time for `Exploit` trials over `Explore` trials. In **Panel E**, left, the distribution of RSI values that occur across all games is shown for all trials (top), explore trials (middle), and exploit trials (bottom), with values counted in intervals of 0.05 RSI. At right is shown the distribution of RSI values of participants' selections for the same groups of trials.

with the model, as well as the average expected savings for remaining unknown points at each trial. Qualitatively, this participant performed better than would be expected on average, and frequently selected the best possible observation as scored by the model.

How effective were people's sampling decisions? In order to characterize the overall match of participant's choices relative to the model, on each trial of each game, we calculated the $ES(G,x_{ij})$ for each possible choice and rank ordered these choices. Then, we counted the number of trials in which each subject actually made an observation at each rank (given by the model). As Fig. 3A & B shows, participants chose the optimal choice (rank 1) on more than 20% of the trials, and chose an option in the top 10 approximately 50% of the time (remember that participants made approximately 30 samples per game, thus there were usually more than 70 possible choices on any given trial).

Search strategies The preceding analysis might suggest that participants' decisions were generally well calibrated to the structure of the task and reasonably followed the predictions of the model, but there were also systematic deviations between the model and human data (e.g., the example game in Figure 2A shows that choices during `Exploit` trials were somewhat more effective compared to `Explore` trials). In order to quantify this effect, we computed a "relative savings index" (RSI) which compares the expected savings of the participant's observation on each trial to the maximum expected savings available on the current gameboard ($RSI = \frac{ES(G,x_{ij})}{Max[ES(G,y_{ij})]}$, where the Max[] is computed over all possible observations y_{ij}). An RSI of 1.0 denotes selection of the maximally informative sample available, and a RSI of 0.0 denotes a completely uninformative, redundant observation.

As seen in Figure 3C, average RSI was significantly greater during trials the model classified as `Exploit` (M=0.72, SD=0.05) relative to trials that were classified as `Explore` (M=0.63, SD=0.07), ($t(5) = -4.36, p < 0.01$), and both types of trials were significantly better than would be expected from a random sampling strategy (`Explore`: $t(5) = 4.36, p < 0.01$; `Exploit`: $t(5) = 9.47, p < 0.001$). In addition, we found that mean reaction time (RT) increased for `Exploit` (M=1126, SD=265) trials relative to `Explore` (M=841, SD=147) trials, $t(5) = -5.03, p < 0.005$ (see Fig. 3D), suggesting that choices made during `Exploit` trials were not only more effective, but also more effortful, than in `Explore` trials.

To check that greater RSI of `Exploit` trials is a genuine improvement of performance and does not simply reflect aspects of the distribution of $ES(G,x_{ij})$ scores available on different trials, we computed the full distribution of RSI and compared this distribution to the distribution of participant's choices. Figure 3E shows the relative frequency of RSI values across the entire gameboard (left) and the the frequency of the RSI of participants' choices (right). Particularly on `Exploit` trials participants select points that fall within the top 5 percent RSI at a much greater frequency than would be expected given the overall distribution.

Learning to Learn We were also interested if participants would show evidence of learning-to-learn (i.e., improvement in sampling efficiency in novel games as a function of experience with other games). A 2 way ANOVA on reaction time using game and search mode (`Explore`/`Exploit`) as repeated factors found a main effect of search mode ($F(1,195) = 96.5, p < 0.001$), main effect of game ($F(1,195) = 3.87, p < 0.001$), but no interaction ($F(1,195) = 0.32, p = 0.99$) (see Figure 4A). A similar 2 way ANOVA on RSI using game and search mode as repeated factors also found a main effect of search mode ($F(1,195) = 45.2, p < 0.001$), but no main effect of game ($F(1,195) = 1.42, p = 0.12$) and no interaction ($F(1,195) = 1.0, p = 0.46$), Figure 4B. Thus, although there is no evidence of systematic improvements in RSI over games, the general decrease in RT suggests that participants

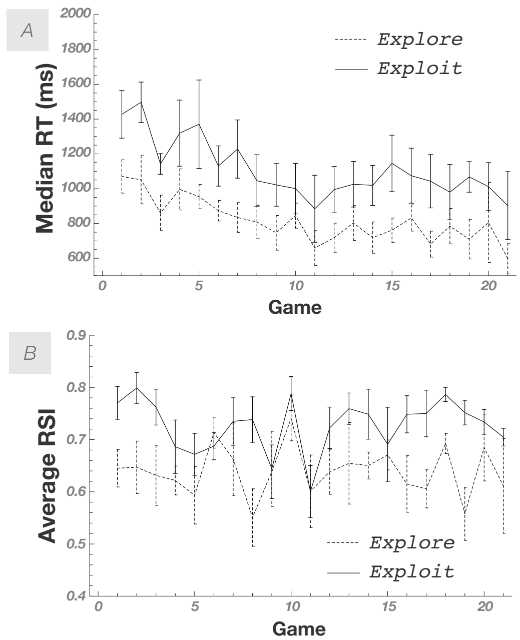

Figure 4: **Panel A** shows changes in average median reaction time (RT) over the course of the games. RT is consistently lower for `Explore` trials than `Exploit` trials, but there is a reduction in RT for both types with practice. In contrast, the average RSI of subjects' observations (an index of maximizing utility) shows no significant change over time (**Panel B**).

were able to make the same "quality" of choices in less time as task experience increased.

Discussion

In this paper, we present a quantitative analysis of human behavior in a learning task which requires participants to continually update expectations based on past observations and to use these expectations to drive new information-seeking decisions. One interesting feature of our task is that it effectively separates information seeking actions from the exploitation of that information (which happens during the separate test phase). By doing so, we were better able to measure how these information generating actions relate to ongoing learning. In addition, unlike other sequential learning and decision making tasks like the n-armed bandit or foraging tasks, there are better and worse choices to be made even when exploiting a single "patch."

One interesting aspect of our results is the fact that both participants (and the model) generally adopt a sequential "hunt and kill" strategy when trying to disambiguate a complex hypothesis space (often disambiguating one rectangle at a time). This search strategy is a natural consequence of information search in a highly "clustered" environment and parallels search patterns adopted by biological organisms in spatial foraging tasks where costs of traveling between patches favors local exploitation (Hills, 2006). However, since the costs for switching between spatially disparate patches in our task were negligible, local search here is more an emergent consequence of adaptive information acquisition and the structure of the hypothesis space.

More importantly, our results suggest a number of interesting facts about human search behavior. First, the ability to devise efficient queries – those that return the most information or are most useful in reaching the learning goal – has previously been shown to vary across task domains, such that people perform highly efficient search in some perceptual tasks (Najemnik & Geisler, 2005) but exhibit biased search strategies in more abstract or conceptual tasks (Klayman & Ha, 1987). The rectangle game combines both these elements: reasoning about the next observation may require hypothesis testing, but it is also possible to "perceive" certain information affordances directly (such as the relative density of previously uncovered locations on the board). The levels of efficiency observed in our task may reflect the confluence of both perceptual and conceptual factors guiding search behavior. Second, our modeling allowed us to objectively classify individual trials as either exploiting a local information "patch" or exploring relatively unknown regions of the game board. We find that participants' response time and search efficiency differs between these two "modes." In contrast to theories which effectively equate information seeking behaviors with random exploration (such as the soft-max rule or epsilon-greedy rule often used in modeling reinforcement learning tasks), we suggest that information generating behaviors may naturally take two forms: one is relatively fast and undirected while the other is slower, more effortful, and efficiently exploits local information constraints.

Finally, note that in this preliminary report, we focused on how people select new observations in learning about a unknown "concept," but an interesting line of future work is to consider how allowing learners to create their own learning experiences can impact the acquisition and retention of new concepts.

Acknowledgements We thank Larry Maloney, Matt Jones, Art Markman, Noah Goodman, Nathaniel Daw, Jason Gold, and the Concepts and Categories (ConCats) group at NYU for helpful discussion.

References

Castro, R., Kalish, C., Nowak, R., Qian, R., Rogers, T., & Zhu, X. (2008). Human active learning. In *Advances in neural information processing systems* (Vol. 21). Cambridge, MA: MIT Press.

Fazio, R., Eiser, J., & Shook, N. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology*, *87*(3), 293-311.

Hills, T. (2006). Animal foraging and the evolution of goal-directed cognition. *Cognitive Science*(3-41).

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211-228.

Kruschke, J. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning and Behavior*, *36*(3), 210-226.

Mackay, D. (1992). Information-based objective functions for active data selection. *Neural Computation*, *4*, 590-604.

Najemnik, J., & Geisler, W. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*, 387-391.

Nelson, J. (2005). Finding useful questions: On bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, *112*(4), 979-999.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608-631.

Skov, R., & Sherman, S. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology*, *22*, 93-121.

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

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. Kearns, S. Solla, & D. Cohn (Eds.), *Advances in neural information processing systems* (Vol. 11, p. 59-65). Cambridge, MA: MIT Press.

Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129-140.