A Bayesian Satisficing Model of Human Adaptive Planning

Wai-Tat Fu (wfu@andrew.cmu.edu)
Department of Psychology,
George Mason University
4400 University Drive, m/s 3f5
Virginia, VA 22030 USA

Abstract

This paper presents a Bayesian satisficing model of when a problem-solver stops planning and begins acting. Existing knowledge about the environment is incrementally updated by new observations, and performance improves as a consequence of better knowledge about the environment. The model aims at bridging the gap between machine learning and cognitive science by adopting the bounded rationality framework (Simon, 1956), which assumes that cognition tends to exploit the characteristics of the environment without engaging in psychologically implausible computations. Empirical studies were conducted when human subjects learned to find the fastest path in a simple map (only one was reported in this paper). The model fit the human learning and performance well and provided insights into the mechanisms behind learning and performance in problem solving.

Introduction

Problem-solving research shows that people seldom plan out complete sequences of actions before acting on the world (Hayes-Roth & Hayes-Roth, 1979; Roberson & Black, 1986). Instead, problem solvers construct partial plans, execute the actions, and plan on further as they act. Problem-solving behavior can therefore be cast somewhere along the planning continuum. At one extreme, a single action is decided and executed based purely on the current state of the problem. This results in highly reactive behavior that can respond quickly to the changing external world. However, this may lead to inefficient solutions to the problem. At the other extreme of the planning continuum, a complete sequence of actions is planned out before any action is executed. This results in highly efficient performance, as the best sequence of actions can be chosen and executed. However, the costs associated with planning often nullify the benefits of planning. Interleaving of planning and acting allows a balancing of problem-solving performance and planning cost involved in the search of actions.

This paper focuses on a class of problems that can be solved relatively easily, but requires sufficient planning to find an efficient solution path. For example, finding a route from Pittsburgh to New York City is relatively easy (there are many possible routes to choose from), but finding the fastest route may require some planning ahead (to find out traffic conditions in different routes).

A number of successful methods have been proposed to obtain the optimal level of planning given sufficient information about the problem-solving environment (Sun & Giles, 2001), but most of them adopt some variations of the dynamic programming approach that requires extensive backward search to find the optimal solution (e.g. Kopf, 1988). Although these methods are often able to lead to optimal solutions, the computations involved is clearly beyond the capacity of the human cognitive system. Psychological research, however, did find that people are able to change their level of planning in response to the characteristics of the environment to improve problem-solving performance (Kirsh & Maglio, Gunnzelman & Anderson, 2002). For example, Kirsh and Maglio found that expert players of the interactive video game Tetris outperformed novice players by better cost-benefit tradeoffs between mental planning and external planning (information-seeking actions). Gunzelmann and Anderson also found that their subjects increased the level of planning in the Tower-of-Hanoi task as they learned that it increased the efficiency of their solutions. In summary, the results suggest that (1) people are sensitive to the cost and benefit of planning, (2) people learn to perform better cost-benefit tradeoffs through experience, and (3) good cost-benefit tradeoffs are often critical to performance.

The studies of how people perform cost-benefit tradeoffs have been of great interests to researchers in the domains of problem solving and decision-making (Christensen-Szalanski, J. J. J., 1980; Beach, L. R., & Mitchell, T. R., 1978; Payne, Bettman, & Johnson, 1993). Recently, a growing number of researchers have adopted the rational approach to explain cost-benefit tradeoffs in decision making behavior and strategy selection (Anderson, 1990; Lovett & Anderson, 1996; O'Hara & Payne, 1998; Fu & Gray, 2000). The major assumption of the rational approach is that people are well adapted to statistical characteristics of their environment, and computations in the human cognitive system perform in ways that are optimal in response to the demand of the environment. Under the rational framework, planning should stop as soon as the cost of further planning exceeds the benefit that further planning could bring. Since continued search of the problem space takes place at increasing cost, searching should stop and execution should begin once the expected benefit of further search drops below a certain level. At this point, the problem solver "satisifices" on

the current level of planning with no guarantee that it is the global best (Simon, 1956; Russell & Wefald, 1991). Similarly, biologists have long entertained the hypothesis that animals forage for resources in a near-optimal manner, given the distribution and replenishment rate of the resources, and the energy cost to obtain the resources (Krebs & Davis, 1978). For example, hummingbirds have been shown to forage flowers in a region until the rate of return is below the average for all flowers, and then forage another regions with greater-than-average return (Pyke, 1978).

Learning and performance

One useful approach to study how people adapt their behavior to the contingencies of a situation is reinforcement learning. During reinforcement learning, the problem solver learns by observing the consequences of their actions over time, and improves their choice of actions with experience. One of the challenges that arises in reinforcement learning and not in other kinds of learning is the tradeoff between exploration and exploitation. In reinforcement learning situations, the learner has to take both learning and performance into account – the problem solver has to learn about the environment to improve performance in the long run and take advantage of the knowledge gained to improve immediate performance. A number of computational methods have been explored (e.g. Sutton & Barto, 1998) to investigate the efficiency of learning from interacting with the environment from the machine learning perspectives. The current paper attempts to bridge the gap between computational methods in machine learning and the rational approach in cognitive science by assuming that the human cognitive system is well adapted to the demands of the environments. By making these assumptions, complex computations can often be replaced by simple mechanisms that exploit the structure of the environment while maintaining the same level of performance (e.g. Anderson, 1990; Gigerenzer & Selten, 2000). Based on this assumption, the model attempts to characterize both learning and performance when human subjects adapt to the environment with some simple, psychologically plausible mechanisms.

The Bayesian Satisficing Model

The current analysis focuses on how optimal performance can be achieved through an adaptation process in which information is incrementally accumulated from the environment. The adaptation process has the dual goals of (1) learning the characteristics of the environment and (2) improving performance by choosing better actions to solve the problem. For learning, the adaptation process has a mechanism that combines new information obtained from the environment with existing knowledge; for performance, a decision criterion is used that allows the person to choose actions based on the existing knowledge of the environment. Specifically, when deciding how much planning one should do during problem solving, the model has a mechanism that updates the knowledge about the relationship between the amount of planning and the execution cost of actions, and a decision criterion on when to stop planning given the existing knowledge.

The model assumes that the problem-solver updates the existing knowledge about the environment through a Bayesian combination of new information from the environment and the existing knowledge. For simplicity, some assumptions about the environment are made. First, it is assumed that the problem is solved by the combination of general heuristics (such as hill-climbing) and special heuristics based on information obtained from planning. It is assumed that special heuristics are likely to be more efficient than general heuristics, and the more planning the problem solver does, the less effort will be required to accomplish the task. It is also assumed that the problem can always be solved (even with no planning). All other possible variables that may influence the perception of the amount of effort required to solve the problem are assumed to be constant.

Learning in the Bayesian satisficing model is concerned with estimating the amount of effort (the acting costs) required to solve the problem from information observed from the environment. The model assumes that the function describing the amount of effort required to solve the problem has an exponential relationship with the number of steps of planning (**n**). Mathematically, f(n,B) can be calculated by

$$f(n,B) = \frac{A}{B} e^{-\frac{n}{B}}$$ (eq1: function of the amount of effort required to solve the problem)

where A is a proportionality constant. B is related to the amount of effort saved per each step of planning. For example, for a given n, the higher the value of B, the lower the value of f(n,B) will be. The exponential distribution also has the characteristic that there is diminishing return in the amount of effort one can save per each additional step of planning[1].

To account for the randomness in the perception of the actual effort a noise term σ is added to f(n,B). σ is a random variable following a normal distribution with

[1] The exponential distribution implies unrealistically that with sufficiently large n, the execution cost will be close to zero. However, it is assumed that n will never be too large with the stopping criterion. Additionally, Figure 3 suggests that the exponential distribution is a good approximation at least to the specific task used in the experiment.

mean equals zero, and standard deviation equals t. The value of t can be considered a free parameter. The uncertainties of **B** in equation 1 are represented by a gamma distribution. The gamma distribution is a two-parameter general distribution that describes the uncertainties of **B** in a general environment. The gamma and exponential distribution are standard non-informative distributions in Bayesian analysis (e.g. see Berger, 1985), which make minimal assumptions on the structure of the environment. However, based on these assumptions, models of the mechanisms that respond optimally to the characteristics of the environment can be constructed.

The optimal number of steps of planning (n_{opt}) with respect to the current distributions of B can be calculated. If w_n is the cost of a unit step of planning, the optimal decision rule to stop planning is when the cost of an additional step of planning exceeds its expected benefit. Mathematically, planning will stop as soon as n satisfies the following equation:

$f(n-1) - f(n) < w_n$ …..(eq2: the stopping criterion)

After the execution of actions, the acting cost can be obtained. Based on the information, the model updates the distribution of **B** using Bayes' theorem. Mathematically, if p(B') is the updated distribution, then

$$p(B') = \frac{f(n_{opt} \mid B)\,p(B)}{\int f(n_{opt} \mid B)\,p(B)\,dB} \quad \text{(eq3: Bayesian learning)}$$

which can be calculated for each n_{opt} chosen according to the stopping criterion stated above. p(B') can then be used as the prior distribution of **B** for the next cycle of selection of n_{opt}, and p(B') can be calculated, and so on. The updated distributions, with more information, will describe the environment better, and therefore generate a better value for n_{opt}. The adaptation process continues and n_{opt} will approach a value that is tuned to the characteristics of the environment (see Figure 1).

The above stopping criterion is a local decision rule that guarantees that performance is optimal given the existing knowledge about the environment. The Bayesian learning process updates the global information, f(n), with new information obtained from the result of the local decision rule. The Bayesian satisficing model therefore combines performance and learning through an incremental update of knowledge of the environment. Global representation of the environment is improved through Bayesian learning and performance is improved through a simple local decision rule, which greatly simplifies the computations involved in many machine-learning approaches.

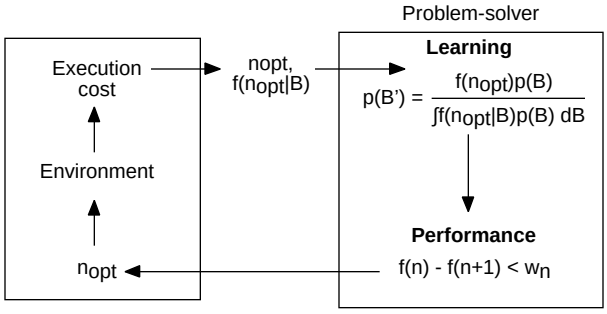

Figure 1. The Bayesian satisficing model – learning is through the Bayes' theorem, and performance is based on an optimal stopping criterion. Execution cost is sampled from the environment and used as estimate for $f(n_{opt}|B)$ for each cycle of learning.

The Bayesian satisficing model therefore nicely combines optimally learning and performance given the characteristics of the environment. The Bayesian learning equation combines optimally the uncertainties in existing knowledge and uncertainties in new observations. The stopping criterion also assumes that the decision on the level of planning is optimal given the existing knowledge. These two components of the model therefore provide a unified account of learning and performance in adaptive planning.

The Task

A simple map-navigation task was chosen as shown in Figure 2. Simple hill-climbing strategies (i.e. no planning) are always sufficient to finish the task, but it does not guarantee to yield the fastest path. With sufficient experience, one learns the speeds of different routes and turns, and will be able to improve performance by a better choice of solution paths. Subjects are given a start station (the blue dot) and a destination (the yellow dot) in each trial, and are asked to travel from the start station to the destination. Subjects can choose to go to any one of the adjacent stations directly connected to the current station. To go to one of the adjacent stations, subjects have to point the mouse cursor to the station, press and hold down the mouse key. A red line will be drawn from the current station (the red dot) to the station clicked. The speeds of the train lines and the transfers are indicated by the speed of the movement of the red line. When the red line reaches the station, the station turns red and becomes the current station.

Subjects can use the transfer at the intersection of the train lines to change direction. When subjects are at a transfer station, he/she can go to another train line or stay on the same train line. Subjects are told that there are two kinds of transfers, the pink transfers and the orange transfers, and one of them is faster than the other. However, they are not told which one is faster. When the trial starts, the colors of the transfers are covered (i.e. in black). The color of a transfer will be shown when the subject is at the transfer station or when the subject clicks on the transfer. As soon as the

experiment starts, the subject can check the color of any transfer in the map anytime before they reach the destination. Figure 2 shows the two kinds of transfers randomly located in the map. The colors are uncovered for illustration purpose only. At any time during the experiment, the subject can see at most one transfer uncovered.

There are two major manipulations of the experiment – planning and acting costs. The task is constructed so that increasing the amount of planning will decrease the cost of acting, and vice versa. Cost is measured by the time to completion of the task, which is the sum of the planning and acting costs. Given the design of the task, subjects have to trade-off planning costs and acting costs to maximize performance.

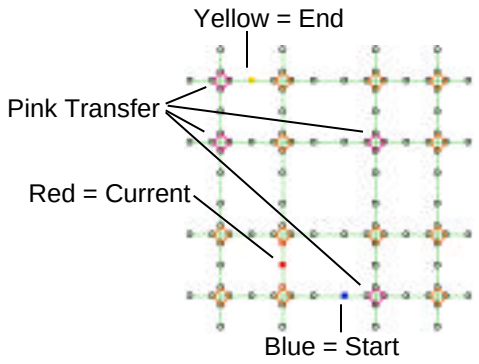

Figure 2. The map used in the experiment. There are two kinds of transfers, one in orange and the other in pink. In the actual experiment, the colors of the transfers are covered. There are 4 pink transfers and 12 orange transfers in each map.

Planning is measured by the number of mouse clicks that check the colors of the transfers. The cost of planning is manipulated by adding a lockout time after the transfer station is clicked. For example, a 1-second lockout time requires the subject to hold down the mouse button for one second before he/she can see the color of the transfer. The cost of acting is manipulated by the speed of the slow transfers. When the slow transfer is much slower than the fast transfer, the effect of planning will be larger because using the fast transfer allows the subject to solve the map much faster than when using the slow transfer.

The Experiment

The experiment was a 2x3 between-subject design. The two between-subject independent variables were Planning Cost (information access cost) and Acting Cost (difference between the speeds of the fast and slow transfers). (In the rest of the paper, the high Planning Cost and high Acting cost condition will be referred to as Hi-Hi, the low Planning Cost and medium Acting Cost condition will be referred to as Lo-Med, etc). The dependent variable was the amount of planning, as measured by the number of mouse clicks to uncover the transfer (a more comprehensive set of analyses can be found in Fu, 2003). Each subject solved 8 blocks of 8 maps. The first seven maps of each block had 12 slow transfers and 4 fast transfers, and there was always one and only one path that contained only fast transfers, and it was always the fastest path. Planning was necessary to find the fastest path (which uses only fast transfers) on these maps, and the shortest path was never the fastest path. These maps were called the round-about-fastest maps. The locations of the start and end stations were randomized in all maps. The eighth map had no fast transfer and will be called the all-slow maps. The all-slow maps showed how much planning the subjects would do when they could not find any fast transfer, and thus allowed the measure of how subjects' decisions to stop planning differed in different conditions and how they changed through experience.

Before the experiment begins, each subject was given a practice trial. The Planning Cost was set according to the condition the subject was in (i.e. either with or without the 1-second lockout time). Subjects were told that the task was to go from the start station (the blue dot) to the end station (the yellow dot) as fast as possible, and they were timed during each map. Subjects were told that there were two kinds of transfer, one was orange and the other was pink, and that one kind was faster than the other. However, they were not told which one was faster. This was not to bias the subject on the use of any one kind of the transfer. Half of the transfers (eight) in the practice trial were orange and the other half (eight) were pink, but the actual speed of the orange and pink transfers were the same in the practice trial. They were shown how to go from one station to another, as well as how to uncover the color of the transfers. Subjects were then asked to solve the map by themselves. Subjects were instructed to solve each map as fast as possible.

Parameters for the environments

Simulations were conducted to estimate the relationship between the amount of planning (n) and the execution cost (C) for map-navigation task (Figure 3). When no planning was done (i.e. n=0), the execution cost was estimated to be the time to go from the start station to the end station using the shortest path (i.e. hill-climbing). For n>0, the following simulation was conducted to obtain the curves in Figure 3. First, one of the transfers on the same train line of the start station would be randomly selected. This would be counted as one planning step, and n will be incremented. For n=1, planning would stop here. If the transfer selected was fast, the transfer would be used; and from the selected transfer, the hill-climbing heuristic would be used to find the path from the selected transfer to the end station (i.e the shortest path from the selected transfer to the end station would be used). The execution cost could

then be calculated. If the selected transfer was slow, then the shortest path from the start station to the end station would be selected. For n = 2, planning will continue depending on whether the first transfer was fast or slow. If the first selected transfer was fast, then one of the transfers would be selected on the train line connected by the selected transfer. If the first selected transfer was slow, another transfer would be selected on the same train line of the start station. This would continue for higher values of n and the corresponding execution costs could be obtained. The simulations were run 100 times because of the stochasticity involved in the selection of transfer. The curves in Figure 3 were used to approximate the function f(n) in the model fits presented below.

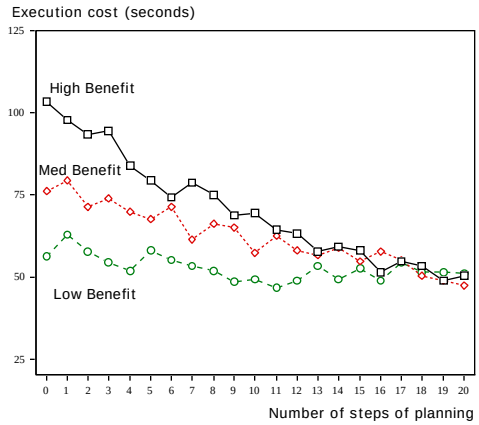

Figure 3. The relationship between the number of steps of planning and the execution cost obtained from the simulations for the map-navigation task. Each point in the figure represents the mean of the results from 100 simulations.

Results

Since modeling the results from the round-about-fastest maps require more elaborate strategies construction, this paper focuses on the results from the all-slow maps. Since there was no fast transfer in the all-slow maps, the major factor affecting the decision on how much to plan would be affected solely by subjects' perception of the expected cost and benefit of planning. The dependent measure was the amount of planning (measured by the number of mouse clicks that uncovered the color of the transfers). The initial parameters for the prior gamma distribution were set to a = 10, b=2. The value of t (the standard deviation of the noise term in f(n|B)) is set to be 2.5. These were free parameters chosen to better fit the data. The same set of parameters was used to fit two other sets of data, providing constraints to the model. However, due to space limitation, only the first set of data was presented here. The model fit is shown in Figure 4.

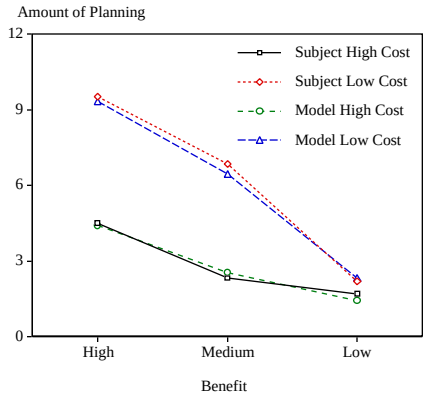

Figure 4. The amount of planning in the all-slow maps done by the human subjects and the model in the map-navigation task.

The main effects of Cost and Benefit were significant ($F(1, 84) = 37.95$, $MSE = 2645.53$, $p < 0.01$ and $F(2, 84) = 34.39$, $MSE = 2397.386$, $p < 0.01$ respectively). Subject planned more when the cost was low and when the benefit was high. The results showed that subjects' decisions on when to stop planning were sensitive to the cost and benefit of planning. The fit of the model to the empirical data was good, $R^2 = 0.92$, $RMSE = 0.21$, suggesting that the model captured subjects perception of the expected cost and benefit of planning well. It also shows that the model captured the cost-benefit tradeoffs well.

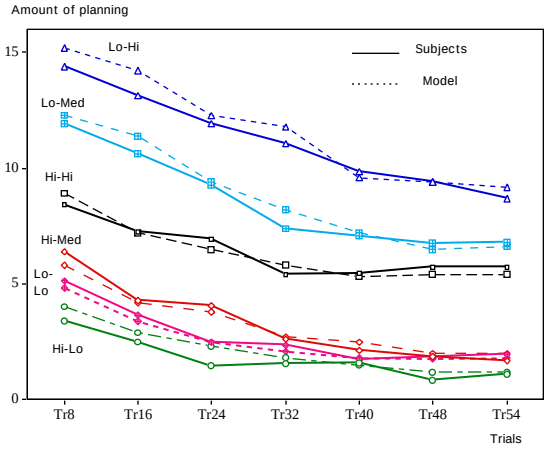

Figure 5. The amount of planning in the all-slow maps across trials done by the human subjects and the model in the map-navigation task. $R^2 = 0.78$, $RMSE = 2.64$.

Figure 5 shows the model fit to the amount of planning across trials. The fit to the empirical data was good, $R^2 = 0.78$, $RMSE = 0.64$. It shows that not only did the model fit the average performance of the subjects well; it also fit the learning of the subjects well across different experimental conditions. Given the free parameters of the model were mainly the parameters for the prior gamma distribution (which were the same in all experimental conditions) and the noise parameter, the fit for both performance and learning suggest that the Bayesian satisficing model did a good job

characterizing the behavior of the subjects. The same model (with the same parameters) was run to fit two more sets of data (see Fu, 2003), which provides further constraints to the model. However, due to space limitations, they were not reported in this paper.

Conclusions and discussions

The above results provided evidence that the interleaving of planning and acting is adaptive. Subjects were willing to plan more when the benefit was high, and plan less when the cost was low. Subjects also showed continued adaptation to the environment with experience. Overall, the Bayesian satisficing model provides a good account of the cost-benefit tradeoffs in planning, and it adapts to similar amount of planning as the subjects in different environments. The model also generated learning curves similar to those of the subjects.

The Bayesian satisficing model is similar to the reinforcement learning approach in machine learning. However, it is rested on the rational assumption, under which the human cognitive system is assumed to be well adapted to the characteristic of the environment. Specifically, both the Bayesian learning computations and the stopping criterion were optimal given the existing knowledge and new observations of the environment. The Bayesian learning takes into account the uncertainties in both the existing knowledge and new observation. The stopping criterion takes advantage of the existing knowledge by choosing a level of planning so that further planning does not justify its cost. The overall model therefore nicely combines both learning of the characteristics of the environment and immediate performance.

Similarly to other models based on the rational assumption, the human cognitive system does not necessarily perform the computations involved as specified in the model. Instead, the computations involved should reflect what the cognitive system should do if the system is well adapted to the characteristics of the environment. For the same reason, the distributions used in the current model may not be general to all problem-solving situations. However, the current endeavor seems to suggest that with minimal assumptions of the environment, the Bayesian adaptation approach seem to be able to characterize human adaptation to new environments well.

Acknowledgments

The author would like to thank Wayne Gray, John Anderson, Lael Schooler, and Kevin Burns for their valuable comments on an earlier version of this paper.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Beach, L. R. & Mitchell, T. R. (1978). A contingency model for the selection of strategies. *Academy of Management Review, 3*, 439-449.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analyses*. New York: Springer-Verlag.

Christensen-Szalanski, J. J. J. (1978). Problem-solving strategies: A selection mechanism, some implications and some data. *Organizational Behavior and Human Performance*, 22, 307-323.

Fu, W.-T. (2003). *Adaptive planning in problem solving.* Dissertation. George Mason University, Fairfax, VA.

Fu, W.-T. & Gray, W. D. (2000). Memory versus perceptual-motor tradeoffs in a blocks world task. In *the proceedings of the 22nd annual conference of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Gunzelmann, G., & Anderson, J. R. (2003). Problem solving: Increased planning with practice. *Cognitive systems research, 4* (1), 57-76.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3*, 275-310.

Kirsch, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive science, 18*, 513-549.

Kopf, R. E. (1988). Optimal path finding algorithms. In Kanal, L. N. and Kumar, V. (eds), *Search in Artificial Intelligence*, pp. 223-267, Springer Verlag, Berlin

Krebs, J. R. & Davies, N. B. (1978). *Behavioural Ecology: An Evolutionary Approach.* Oxford: Blackwell.

Lovett, M. C., & Anderson, J.R. (1996). History of success and current context in problem solving: combined influences on operator selection. *Cognitive psychology*, 31 (2), 168-217.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. NY: Prentice Hall.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.

Pyke, G. H. (1978). Optimal foraging in hummingbirds: testing the marginal value theorem. *American Zoologist,18*, 739.

Robertson, S. P., & Black, J. B. (1986). Structure and development of plans in computer text editing. *Human-Computer Interaction, 2*, 201-226.

Russell, S. & Wefald, E. H. (1991). *Doing the right thing: studies in limited rationality*. Cambridge, MIT Press.

Simon, H. A.(1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138.

Sun, R., & Giles, L. (2001). *Sequence Learning: Paradigms, Algorithms, and Applications.* Heidelberg, Germany: Springer-Verlag.

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