Using Cognitive Decision Models to Prioritize E-mails

Michael D. Lee, Lama H. Chandrasena and Daniel J. Navarro
{michael.lee,lama,daniel.navarro}@psychology.adelaide.edu.au
Department of Psychology, University of Adelaide
South Australia, 5005, AUSTRALIA

Abstract

E-mail prioritization involves placing all of the 'useful' or 'good' unread e-mails at the top of the inbox, and all of the bad ones at the bottom. We use two cognitive decision models—a rational model, which considers all of the available information, and a fast and frugal model that uses one reason decision making—to prioritize e-mails. Experimental results, using real data obtained by unobtrusively logging e-mail user behavior, show that the fast and frugal model is just as effective as the rational model. The results also show that a Bayesian approach to learning is superior to the standard frequentist approach, because it balances the competing demands of exploration and exploitation in finding good e-mails. We use the results to draw some applied conclusions about the development of an e-mail prioritization system, and note some theoretical implications of the results for the cognitive modeling of human decision making in general.

Introduction

Anybody who has returned from holidays to be confronted with 600 unread e-mails appreciates the need for prioritization. Ideally, we would like an unread inbox to rank the e-mails, putting those that are the most 'important', 'urgent', 'useful' or 'good' at the top, and those that are less important at the bottom.

While machine learning methods have been applied to the problem of e-mail prioritization (e.g., Macskassy, Dayanik, & Hirsh 1999; Mehran, Dumais, Heckerman, & Horvitz 1998), it has typically not been treated as a cognitive modeling problem. Clearly, however, prioritizing requires an ability to predict whether or not a user is likely to evaluate a message as a good message, and so requires an effective model of human decision making to be successful.

Using cognitive models for prioritization does not only promise to provide an answer to an applied problem, but also has theoretical benefits for the more general study of human decision making processes. This is because, in the form of real-world e-mails, it deals with a richly structured stimulus domain. There are, of course advantages in studying decision making with artificial stimuli, as is often done in the categorization and classification literature (e.g., Shepard, Hovland, & Jenkins 1961), because of the experimental control that is achievable. A central argument of ecological approaches (e.g., Simon 1956; Gigerenzer & Todd 1999), however, is that it is also important to consider the role of non-arbitrary stimulus environments in supporting (or confounding) human decision making.

In this paper, we develop and evaluate two cognitive models for prioritization. One is a 'rational' model, that performs exhaustive calculations, while the other is a 'fast and frugal' model, that requires only limited time by making assumptions about the nature of its environment. In the next section, we describe how e-mails are represented by these models, and how information about them is learned. We then describe the two models in detail, before presenting the results of an experiment in which both are evaluated on real-world data. Finally, we draw some conclusions regarding the theoretical implications of the results for understanding human decision making, and the applied implications for building an e-mail prioritization system.

E-mail Representation and Learning

Cues and Cue Validities

We follow previous research in assuming e-mails are represented in terms of a set of binary features, which we call cues. These cues may relate to the content of the e-mail, such as a keyword in the message text, or metadata associated with the e-mail, such as the name of the sender. In this way, each e-mail may be defined by the set of cues that it contains.

Following Gigerenzer and Todd (1999), we associate a cue validity with each cue, which measures the probability that an e-mail will be regarded as good, given that it has the cue. Formally, this means that the validity, v_i of the i-th cue, c_i is defined as $v_i \equiv p\,(\mathrm{G} \mid c_i)$, where G denotes good. Notice that, because each e-mail is assumed to be either good or bad, $1 - v_i$ gives the $p\,(\mathrm{B} \mid c_i)$, the probability that an e-mail will be bad when it has the i-th cue.

Learning Cue Validities

Where the cues constitute the representational component of our decision making models, the way in which the validities are specified constitute the learning processes, in the sense that different validities apply in different environments, and are formed on the basis of information observed in those environments. We consider two methods for learning cue validities, arising from the alternative frequentist and Bayesian statistical approaches. In both cases, we assume that (in ways described later) every e-mail that has previously been processed by a user has been classed as a good e-mail or a bad e-mail. This means that the raw data for the i-th cue take the form of a count g_i, giving the number of good e-mails with the cue, and a count b_i, giving the number of bad e-mails with the cue.

Under the frequentist approach, the validity of a cue is estimated simply as the proportion of good e-mails with the cue:

$$\hat{v}_i = g_i / \left(g_i + b_i + \epsilon\right),$$

where ϵ is a small positive number that ensures cues have a defined validity of zero before they have been observed in any e-mail (i.e., the case $g_i = b_i = 0$).

Under the Bayesian approach, prior beliefs regarding the validity of the i-th cue are modified using the data provided by the counts g_i and b_i. As a cue becomes associated with more good e-mails, higher values for its validity become more likely. Conversely, as a cue becomes associated with more bad e-mails, lower values for its validity become more likely. Bayes' theorem describes the way in which the prior beliefs are modified by data to give a probability distribution over the range $[0, 1]$ of possible validities. Defining the validity of a cue as the mean of this distribution, and assuming a uniform prior, gives the result (see Gelman, Carlin, Stern, & Rubin 1995, p. 31):

$$\hat{v}_i = E\left[p\left(v_i \mid g_i, b_i\right)\right] = \frac{g_i + 1}{g_i + b_i + 2}.$$

As more e-mails with the i-th cue are processed the counts g_i and b_i increase, and the frequentist and Bayesian approaches converge towards the same value. When few data are available, however, we later show that the Bayesian approach has advantages for prioritization.

Decision Models for Prioritization

The 'Rational' Approach

Under the rational approach to decision making used here, the evidence provided by every cue associated with an e-mail is integrated to give an estimate of the overall log odds that the e-mail is good, as opposed to bad. Assuming that the evidence provided by each cue is independent, and that the prior probabilities of an e-mail being good or bad are equal, then Bayes' theorem gives:

$$\ln \frac{p\left(\mathrm{G} \mid c_1, ..., c_n\right)}{p\left(\mathrm{B} \mid c_1, ..., c_n\right)} = \sum_{i=1}^{n} \ln \frac{p\left(c_i \mid \mathrm{G}\right)}{p\left(c_i \mid \mathrm{B}\right)}.$$

The required evidence ratios $p\left(c_i \mid \mathrm{G}\right) / p\left(c_i \mid \mathrm{B}\right)$ can be estimated from the data in the same way as cue validities, or (with some manipulation) written in terms of the validities themselves. The rational approach has the attraction of considering all of the data, in the sense that it considers the evidence provided by every cue associated with every stimulus. For this reason, it is often considered a normative account of decision making, and has been used extensively (in one form or another) to model human decision making. As an (arbitrary) example, consider Kruschke's (1992) well known ALCOVE model, which uses a weighted sum of the evidence provided by each dimension of a stimulus in deciding whether or not that stimulus belongs to a category. The rational approach is also widely used in machine learning, and has been applied in previous research (Macskassy *et al.* 1999; Mehran *et al.* 1998) on prioritizing e-mails.

The 'Fast and Frugal' Approach

In developing their 'fast and frugal' approach to modeling human decision making, however, Gigerenzer and Todd (1999) challenge the rational approach. They argue that because human decision making processes evolved in a competitive environment, they need to be fast, and because they evolved in a changeable environment, they need to have the robustness that comes from simplicity. To meet these challenges, the fast and frugal approach adopts Simon's (1982) notion of 'bounded rationality', and models human decision making using simple algorithms that rely on an assumed structure in the stimulus environment to function effectively.

For example, in an environment where the validity of one stimulus cue is highly predictive of the validities of the remaining cues, and the examination of additional cues is an effortful process, it is sensible to consider only the first cue. Similarly, in an environment of diminishing returns, where the examination of each successive cue provides less information than previous cues, it makes sense to base decisions on a small number of cues. Gigerenzer and Todd (1999) show that many real-world stimulus domains have these sorts of structures, and develop a number of cognitive models—including the 'Take the Best' model of forced choice, the 'QuickEst' model of value estimation, and the 'Categorization by Elimination' model of categorization—that make inferences by assuming environmental regularities.

Unfortunately, none of these models is directly applicable to the problem of e-mail prioritization, and so we developed a new model using the basic 'fast and frugal' modeling approach. Gigerenzer and Todd (1999) argue that their models of human decision making are based on simple mechanisms that answer three fundamental questions:

- How should a stimulus environment be searched for information?
- When should this search for information be terminated?
- Once the search has been terminated, what decision should be made given the available information?

In the context of finding good unread e-mails, as required for priorization, it is not difficult to provide answers to these questions:

- Unread e-mails should be searched in terms of cues, looking for e-mails with high validity cues.
- The search should be terminated as soon as at a candidate good e-mail has been identified. Since users process e-mails serially, there is no benefit in seeking to sort the unread e-mails, beyond attempting to ensure that at any time the top-most e-mail is the one most likely to be good.
- The best available e-mail should be placed at the top of the inbox, as the next one to be read by the user.

These answers suggest a simple fast and frugal decision model for prioritization. The cues are ordered in terms of their estimated validity and, starting with the highest validity, a search is made for an unread e-mail that has this cue. If this search is successful, the process terminates without considering any further cues. If no e-mail is found, the search continues using the next highest validity cue, and this process is repeated until an e-mail is found. This model is closely related to Take the Best, and belongs to the class of what Gigerenzer and Todd (1999) term 'one reason decision making' models. Only one reason, in the form of the presence of a high validity cue, is all that is required to find the next e-mail for presentation.

Experiment

Data Collection

We developed a macro for the Microsoft Outlook e-mail application that unobtrusively logged the behavior of one user for a period of 76 consecutive days. This logging involved recording the actions made by the user in reading, responding to, and organizing the e-mails in their inbox. Every time the user replied to, forwarded, saved, moved or deleted an e-mail in their inbox, an entry in a log file was made. Often, a particular e-mail was subjected to several processing actions over time, such as being forwarded, and then deleted. We used an operational definition of what makes an e-mail a good or bad one, based on these processing sequences, to enable each e-mail to be labeled as either good or bad. If an e-mail was *ever* printed, forwarded, replied to, copied to another folder, or saved, it was regarded as a good e-mail. Otherwise, when the e-mail was only deleted, it was regarded as bad. Of course, this sort of operational definition is not unproblematic, but we believe it represents a reasonable first-order approximation for identifying those e-mails that are more 'important', 'useful' or 'urgent'.

Table 1: The logged e-mail properties used to generate cues, together with a sample cue.

Property	**Sample Cue**
Attachments	"AttachmentCount=2"
CC list	"CC=mark@adelaide.edu.au"
Flag status	"FlagStatus=0"
Importance	"Importance=1"
Sender's e-mail	"SendersEmail=jdl@mbox.com"
Sender's name	"SendersName=John Lee"
Subject keyword	"Subject=upgrade"
Addressee list	"To=Ben Stamley"

The properties of the e-mails being manipulated were also recorded, providing the subject text of the e-mail (for ethical reasons no message content was recorded), as well as metadata identifying the sender of the e-mail, whether it had an attachment, and so on. Those properties taking discrete values were used to generate the cues for representing e-mails by pairing the property with each possible value. Table 1 details the eight properties used in this way, together with an example of a cue for each. The only pre-processing used in generating these cues was to remove common English words from the subject text using a 'stopword' list. The final data set contained 886 e-mails, 362 of which were good, defined in terms of 3,112 binary cues.

Effectiveness of Prioritization

Both the rational and the fast and frugal models were applied to the e-mail data, using the Bayesian learning approach. Each day's e-mails were prioritized in sequence, to simulate the effect that prioritization would have if it were implemented on-line. Figure 1 summarizes the results of 10 independent applications of each method using an effort-reward graph. The performance curves relate hypothetical levels of 'effort', which describe the proportion of available e-mails processed by the user, to the resultant level of reward, as measured by the proportion of available good e-mails that are found. Mean performance levels are shown by the curve, with best- and worse-case perfor-

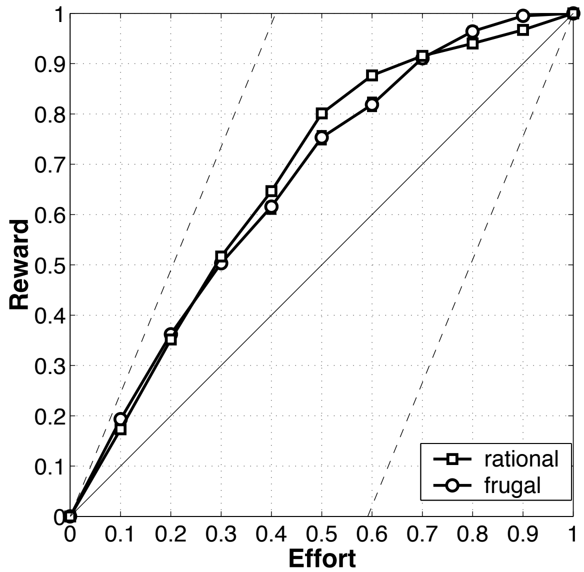

Figure 1: Effort-reward performance of the rational and 'fast and frugal' decision models.

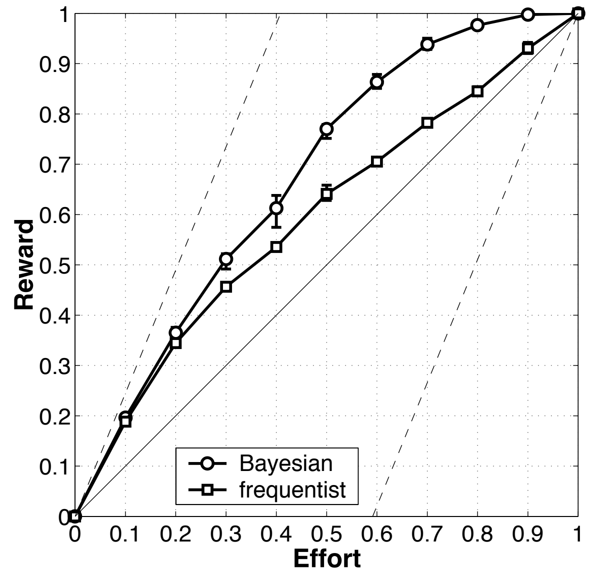

Figure 2: Effort-reward performance of the 'fast and frugal' model using Bayesian and frequentist learning.

mance, arising from the stochastic process of breaking ties, indicated by error bars (where large enough to be visible).

Without prioritization, good e-mails are evenly distributed according to their base-rate of occurrence, which corresponds to the diagonal line in Figure 1. The best- and worse-case possible effort-reward performance of prioritization are shown by the dotted lines, which correspond, respectively, to the cases where all good e-mails are presented first, and where all bad e-mails are presented first. Figure 1 shows that the rational and the fast and frugal model perform very similarly. They are close to optimal for the first 10-20% of good e-mails, but then perform less impressively, although they continue to provide a significant advantage over non-prioritized presentation. Reading the first 50% of e-mails, for example, results in finding approximately 75-80% of the good e-mails available.

Figure 1 suggests two important conclusions. Firstly, it shows that prioritization is effective, which suggests that human decisions in processing the e-mails have some level of systematic relationship with the various cues by which the e-mails are represented. Second, the fast and frugal approach is approximately as effective as the rational approach, which suggests that the human decision making process can be understood in terms of the identification of key features of the e-mails, rather than the exhaustive integration of all of their properties.

Bayesian and Frequentist Learning

An important theoretical problem for prioritization relates to the balance between exploitation and exploration processes. In the context of e-mail prioritization, exploitation involves using cues that are known to have some validity to find good e-mails, while exploration involves learning more about cues for which little or nothing is known, in the hope of find new sources of good e-mails. Prioritization algorithms are of limited use if they achieve their results by exploitation at the expense of exploration, particularly in dynamically changing environments. For this reason, there has been some considerable effort in the machine learning literature (see Sutton and Barto 1998) to balance the competing demands of exploitation and exploration, usually by introducing some stochastic element into the search process.

As it turns out, the Bayesian approach to learning validities addresses this problem. Figure 2 shows the effort-reward performance of 10 runs of the fast and frugal model using both the Bayesian and the frequentist approaches. To assist in the exposition of our subsequent analyses, only a limited set of cues, consisting of all of those generated from the easily understood 'Senders Name' field were used. As Figure 2 shows, the Bayesian approach performs better, particularly for effort levels greater than about 0.5.

The reason for the superiority of the Bayesian validity estimate can be demonstrated through a concrete example. On day 43, a (small) total of five e-

Table 2: Sender cues, good (G) and bad (B) counts, and estimated Bayesian and frequentist validities for day 43.

Sender's Name	G	B	Bayes	Freq.
ABC News Online	1	140	0.01	0.01
Scott Brown	1	0	0.67	1.00
Tapes Subliminales	0	0	0.50	0.00
Virtual Florist	0	1	0.33	0.00
W. Paul Malcolm	0	0	0.50	0.00

mails required prioritization, coming from five different senders. Of these senders, three had previously sent e-mails: "Scott Brown" had sent one good e-mail, "Virtual Florist" had sent one bad e-mail, and "ABC News Online" had sent 141 e-mails, only one of which had ever been good. These patterns of good and bad counts, together with their Bayesian and frequentist cue validity estimates, are shown in Table 2.

Under the frequentist approach, the "Scott Brown" e-mail will be presented first, because it has been associated with the highest proportion of good e-mails. The next e-mail presented will be the "ABC News Online" e-mail, because it has the next highest estimated validity, by virtue of being the only other sender ever to provide a good e-mail. The remaining two unknown senders have estimated validities of zero, and so their e-mails will be presented in random order. As it happens, one of these e-mails, from the new sender "W. Paul Malcolm" is a good one, and so prioritization will be ineffective. Fundamentally, this is because frequentist validity estimation favors the exploitation of sources with very limited returns over the exploration of unknown sources.

Using the Bayesian approach, the "Scott Brown" e-mail will again be presented first, because it has the highest estimated validity. However, "Virtual Florist" and (especially) "ABC News Online" e-mails will not be presented until after those from the senders about whom nothing is known, because their validities are below the 0.5 prior. In this way, the potential new sources of good e-mails will be explored before those that are known to have limited returns are exploited. Notice also that the "Virtual Florist" e-mail will be presented before the "ABC News Online" e-mail, because less data are available for estimating the validity of the former, and so it has more scope to achieve a higher estimate as more observations are made (i.e., it is more worthy of further exploration). Finally, we note that the situation with many 'spam' e-mails is naturally handled within the Bayesian approach by changing the prior on good e-mails.

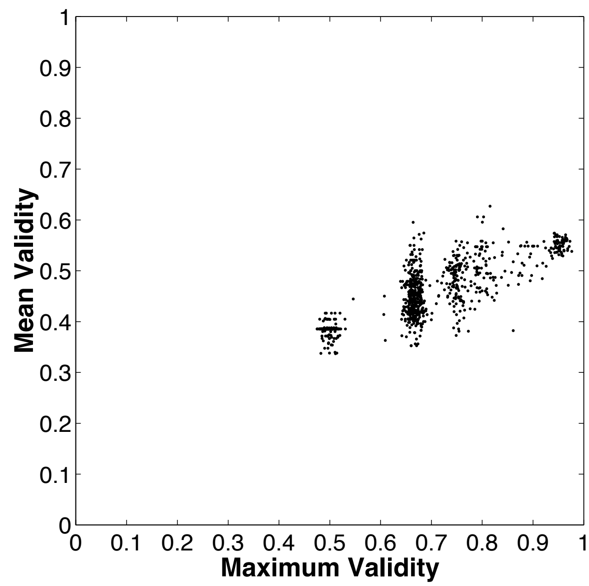

Figure 3: The relationship between the maximum cue validity for an e-mail, and its mean cue validity.

The Structure of the Environment

An analysis of the e-mail stimulus domain explains why the fast and frugal approach performs similarly to the rational approach. Figure 3 shows the relationship between the mean estimated validity of the cues associated with each message (using Bayesian learning), and the maximum estimated validity. There is a positive correlation of $r = 0.80$ between these measures, indicating that the maximum cue validity, as used by the fast and frugal method. is highly predictive of the validities of the remaining cues considered by the rational method. This environmental regularity is the reason for the success of the fast and frugal model: By finding the unread e-mail with the greatest cue validity, it does not need to consider further cues, because their validities are largely already determined by the maximum value.

Future Work

The outstanding problem relates to adaptation. If the characteristics of the external e-mail environment change (e.g., people send different types of e-mails), or the user changes the way they regard e-mails as good or bad, prioritization needs to reflect the new situation. The learning processes used in our study will be slow to adapt to these sorts of changes, as demonstrated for the Bayesian approach by the pattern of change of the five cues shown in Figure 4. Validities for the "To=Mike Lee" and "Subject=newmail" cues are learned effectively, because

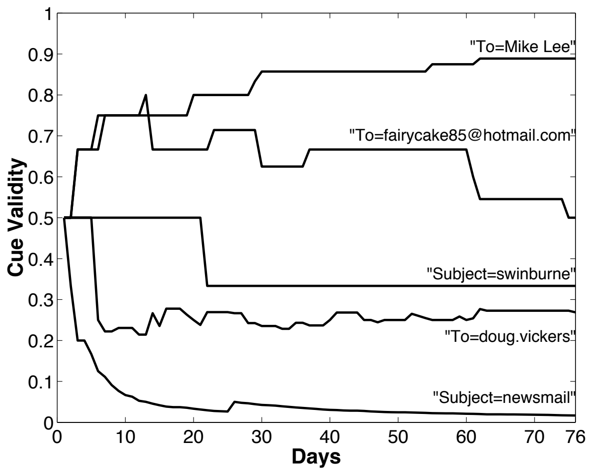

Figure 4: The pattern of change over all processing days for five cues, using Bayesian learning.

they are consistently evaluated by the user. The "To=fairycake85@hotmail.com" cue, however, is evaluated as good in the first two weeks, but its change to a bad cue is learned slowly. Meanwhile, the cues "Subject=swinburne" and "To=doug.vickers" have similar estimated validities at day 76, yet there are grounds to be more confident about the accuracy of the latter, since it is based on a significant volume of recent data, while the former has not been seen since about day 22.

The ability to adapt requires that memory processes be introduced into the cognitive decision models. By replacing old information in the counts g_i and b_i with new information, giving greater weight to new information, or forcing information to decay over time, validity estimates will be based on data that reflects the current state of affairs. A variety of memory mechanisms have been developed for simple psychological decision models (e.g., Pietsch & Vickers 1997), and their detailed empirical evaluation is a priority for future research. The other necessary area of future research is to extend our evaluation to a larger number of users.

Conclusion

We argued in the Introduction that using cognitive decision models to prioritize e-mails provided a way to address an applied problem, and also advance our theoretical understanding of human decision making. We conclude by suggesting some implications of our results on both the applied and theoretical fronts.

In terms of developing an e-mail prioritization application, the fast and frugal model has significant potential. The data required to drive the algorithm, in the form of user evaluations of good and bad e-mails, is done entirely unobtrusively, does not require any additional user effort, and provides a continual on-line data source that should allow for adaptation. The balance between exploration and exploitation is handled naturally by the Bayesian approach to validity estimation, and the the fast and frugal algorithm scales well to large problems. Only one e-mail with one cue needs to be found at each stage of prioritization, as compared with the rational approach, which examines every cue of every e-mail at every stage.

Theoretically, our results suggest that human decision making in processing e-mails can be understood in terms of a one reason decision making process that is tuned to regularities in its environment, and so supports Gigerenzer and Todd's (1999) fast and frugal approach to cognitive modeling. The Bayesian approach to validity estimation also provides a theoretical tool for any learning or decision making situation where exploration must be balanced with exploitation, and could be used in other cognitive decision models.

Acknowledgments

This work was supported by the Australian Defence Science and Technology Organisation. We thank Helen Braithwaite, Brandon Pincombe, Kenneth Pope, Florian Sollich, Douglas Vickers, Michael Webb, and Chris Woodruff.

References

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.

Gigerenzer, G., & Todd, P.M. (1999). *Simple Heuristics that Make Us Smart*. New York: Oxford University Press.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Macskassy, S. A., Dayanik, A.A., and Hirsh, H. (1999). Emailvalet: Learning user preferences for wireless email. In *Proceedings of Learning about Users Workshop, IJCAI'99.*

Mehran, S., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A Bayesian approach to filtering junk e-mail. In *AAAI-98 Workshop on Learning for Text Categorization.*

Pietsch, A., & Vickers, D. (1997). Memory capacity and intelligence: Novel techniques for evaluating rival models of a fundamental information processing mechanism. *Journal of General Psychology, 124*, 229–339.

Shepard, R.N., Hovland, C.L., & Jenkins, H.M. (1961). Learning and memorization of classification. *Psychological Monographs, 75*(13), 517.

Simon, H.A. (1982). *Models of Bounded Rationality.* Cambridge, MA: MIT Press.

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