

Applying One Reason Decision-making: The Prioritisation of Literature Searches

Michael D. Lee, Natasha Loughlin, and Ingrid B. Lundberg
University of Adelaide, Australia

The prioritisation of literature searches aims to order the large numbers of articles returned by a simple search so that the ones most likely to be relevant are at the top of the list. Prioritisation relies on having a good model of human decision-making that can learn from the articles users select as being relevant to make predictions about which of the remaining articles will be relevant. We develop and evaluate two psychological decision-making models for prioritisation: A "rational" model that considers all of the available information, and a "one reason" model that uses limited information to make decisions. The models are evaluated in an experiment where users rate the relevance of every article returned by PsycINFO for a number of different research topics, with the results showing that both models achieve a level of prioritisation that significantly improves upon the default ordering of PsycINFO. The one reason model is shown to be superior to the rational model, especially when there are only a few relevant articles. The implications of the results for developing prioritisation systems in applied settings are discussed, together with implications for the general modeling of human decision-making.

When a researcher first does a literature search, they usually are only able to supply general search criteria, such as one or two keywords, to indicate their broad topic of interest. Typically, these initial searches will return a large number of potentially relevant articles. Faced with this information overload, one option for the researcher is to refine their search, and hope that a more manageable list of articles is returned. Often, however, this refinement is difficult, because the researcher is unsure exactly what sorts of materials are available, and there is a need to "sample" or "explore" the large initial list of articles before a more detailed search can be constructed with any confidence.

Prioritisation offers a different approach to dealing with the information overload. The basic idea is to begin presenting the articles, requiring the user to indicate whether or not that article is of interest. As each article is examined, prioritisation acts to re-order the remaining articles so that the relevant ones are placed at the top of the list. If prioritisation is effective, the problem of information overload is solved without the user ever having to construct a refined search. They only need to work from the top of the prioritised list until they reach the point where the articles are no longer of sufficient relevance to be worth pursuing.

While the prioritisation problem has been tackled in a variety of information retrieval contexts using machine learning techniques (e.g., Balabanovic, 1998; Macskassy, Dayanik, & Hirsch 1999; Mehran, Dumais, Heckerman, & Horvitz 1998), it has typically not been tackled from a cognitive modeling perspective. This is unfortunate, because prioritisation rests on the ability to predict whether or not a user will evaluate an article as a relevant one, and so requires an effective model of human decision-making to be successful.

In this paper, we develop and evaluate two cognitive models for the prioritisation of literature searches. One is a "rational" model, that performs exhaustive calculations, while the other is a "one reason" model, that requires only limited time by making assumptions about the nature of its environment. In the next section, we describe how literature searches 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 where 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 a literature search prioritisation system.

TWO MODELS OF DECISION-MAKING

Representation and Learning Assumptions

We follow Gigerenzer and Todd (1999), and a substantial body of other cognitive modelling (e.g., Medin & Schaffer, 1978; Tversky, 1977), in representing stimuli in terms of the presence or absence of a set of discrete features or properties, which we call cues. This means that each article is represented by information such as the authors of the article, the journal it appeared in, keywords in the title or the abstract, the language of publication, and so on.

As the user provides information, rating some articles as relevant and rejecting others, it is possible to learn how the individual cues are associated with the different judgments. At any stage, it is known how many times a cue has been associated with a previously presented article, and how many of those articles have been relevant. From these counts, it is possible to measure the evidence the presence of a cue provides for a new article being relevant. In effect, the cues correspond to our representational assumptions, while the adjustment of evidence values associated with the cues correspond to our learning assumptions.

Following Lee, Chandrasena and Navarro (2002), we adopt a Bayesian approach to learning. The basic idea is that we start with complete ignorance about how the cues relate to relevance, and each time a cue is observed to be associated with a relevant article, it comes to provide greater evidence that a new article with the cue will also be relevant. Similarly, as a cue is associated with more irrelevant articles, it provides greater evidence that a new article will also be irrelevant. The basic result from Bayesian statistics (Gelman, Carlin, Stern, & Rubin, 1995, p. 31) we use is that, if an event (such as a cue being associated with a relevant article) occurs k times out of n trials, the best estimate of its underlying probability is $(k+1)/(n+2)$. Lee, Chandrasena and Navarro (2002) provide

a more detailed explanation of this Bayesian approach to learning, with a particular focus on its theoretical advantages over other methods.

The rational and one reason models we consider both use cue representations and Bayesian evidence values, but differ in the way they calculate and combine the evidence values to reach a final decision. In this sense, the two models make the same assumptions regarding how people represent and learn about the articles, but make different assumptions regarding their decision processes in judging relevant articles.

The Rational Model

The “rational” approach assumes that people combine all of the available information is combined in some (near) optimal way. This means that the evidence provided by all of the cues must be weighted and integrated to arrive at a final decision. Because it uses all of the data, the rational approach is often regarded as a normative theory of decision-making, and is central to the decision and utility theoretic frameworks widely used in the physical sciences, and in behavioural sciences such as psychology and economics (see Doyle, 1999, for an overview).

Our version of a rational approach works by estimating the probability that an article is relevant, as opposed to irrelevant, on the basis of the cues it has. As it turns out, it is simpler to calculate this probability on a log-odds scale. This is a straightforward transformation: a probability of 0.75 means an event will occur three times out of four, which correspond to odds of 3:1, or log-odds of $\ln 3 \approx 1.10$. The log-odds scale has the advantages of being symmetric about the origin and additive: log-odds of zero mean that the probability of relevance is 0.5, and equal positive or negative increments represent equal amounts of evidence in favor of relevance or irrelevance. It should be emphasised, however, that making decisions based on the log-odds is identical to making decision based on the probabilities themselves.

Formally, the log-odds that an article is relevant (denoted G for “good”), as opposed to irrelevant (denoted B for “bad”), given their cue representations, is written as:

$$L_{GB} = \ln \frac{p(G | c_1, \dots, c_k)}{p(B | c_1, \dots, c_k)}$$

Using Bayes’ Theorem, this may be re-written as:

$$L_{GB} = \ln \frac{p(G)}{p(B)} + \ln \frac{p(c_1, \dots, c_k | G)}{p(c_1, \dots, c_k | B)}$$

The rational model we use here assumes that the evidence provided by each cue is independent when integrating them to give an estimate of the overall log-odds that an article is relevant, so the log-odds become:

$$L_{GB} \approx \ln \frac{p(G)}{p(B)} + \sum_{i=1}^k \ln \frac{p(c_i | G)}{p(c_i | B)}$$

Given the log-odds for every article, prioritisation involves simply sorting from the greatest log-odds value to the smallest. In this way, the articles most likely to be relevant are at the top, and those least likely to be relevant are at the bottom. Notice that, in doing this ordering, the prior odds, $p(G)/p(B)$ will be constant for every article, and so do not need to be calculated. Prioritisation is based entirely on the evidence provided by each of the cues associated with the articles.

The evidence values $p(c_i | G)/p(c_i | B)$ can be estimated on the basis of the user’s acceptance or rejection of previous articles. In this way, the evidence values are continually learned from the user, starting at uninformative prior values, but evolving over time to reflect the preferences implicit in user decisions. Formally, suppose at a given point there

have been g relevant articles, in which the i -th cue has been present x times, and b irrelevant articles, in which it has been present y times. Using the Bayesian approach to learning, we have:

$$\frac{p(c_i | G)}{p(c_i | B)} \approx \frac{(x+1)/(g+2)}{(y+1)/(b+2)}$$

It is the sum of these evidence values, for each of the cues belonging to a new article, that gives a rational estimate of the log-odds that it is relevant.

The One Reason Model

In developing their “fast and frugal” approach to modeling human decision-making, Gigerenzer and Todd (1999; see also Gigerenzer & Goldstein, 1996; Todd & Gigerenzer, 2000) challenge the rational approach. They argue that because human decision-making processes evolved in competitive environments, they need to be fast, and because they evolved in changeable environments, they need to have the robustness that comes from simplicity. Rational models usually do not meet these constraints, because they involve extensive and often complicated calculations in their decision-making processes.

The emphasis of fast and frugal modeling on the role of the environment follows ecological approaches¹ to psychology (e.g., Brunswik, 1943; Simon, 1956, 1982), and suggests that understanding human decision-making requires understanding not just mental processes, but also the external task environment, and its interaction with mental processes. As Gigerenzer and Todd (1999) argue, the fact that environments are not arbitrary means that they can play a role in supporting (or confounding) human decision-making. For example, in an environment where one piece of information in a stimulus is highly predictive of the remaining pieces of information, and the search for additional information is an effortful process, it is adaptive to consider only the first piece of information. Similarly, in an environment of diminishing returns, where each successive piece of information provides less information than previous pieces, it makes sense to base decisions on the first few pieces of information. Gigerenzer and Todd (1999) show that many real-world stimulus domains have these sorts of information structures, and develop a number of cognitive models that make inferences by assuming the presence of environmental regularities.

Unfortunately, none of these models is directly applicable to prioritisation, and so we developed a new model using the basic fast and frugal 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 relevant articles, as required for prioritisation, it is not difficult to provide answers to these questions:

- Unread articles should be searched in terms of cues, looking for articles with cues that provide strong evidence that they are relevant.
- The search should be terminated as soon as a candidate relevant article has been identified. Since users read articles serially, there is no benefit in seeking to sort the unread articles beyond attempting to ensure that at any time the top-most article is the one most likely to be good.
- The best available article should be placed at the top of the list, as the next one to be read by the user.

These answers suggest a simple fast and frugal decision model for prioritisation. The cues are ordered in terms of the evidence they provide in favor of an article being relevant. As with the rational model, these evidence values are easily estimated on the basis of the user's acceptance or rejection of previous articles, and so start at uninformative prior values, but are continually updated over time by learning from the cumulative information provided by all of the user decisions. Formally, if the i -th cue has been associated with n articles, k of which were relevant, then the ratio $(k + 1)/(n + 2)$, corresponding to what Gigerenzer and Todd (1999) call the validity of the cue, provides an appropriate measure. Starting with the highest validity cue, a search is made for an unread article with that cue. If this search is successful, the process terminates without considering any further cues. If no article is found, the search continues using the next best cue, and this process is repeated until an article 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 evidence cue, is ever required to find the next article for presentation.

EXPERIMENT

Data Collection

To compare the rational and one reason models, we tested their ability to prioritise literature searches from the PsycINFO (2001) database. Our data set contained 10 different literature searches, done by people with experience in using the system, but without detailed knowledge of the models being evaluated. For each of the ten searches, a topic was chosen, and a small set of keywords was chosen for an initial search. Every one of the articles returned by PsycINFO was then evaluated independently, assessing whether or not it was relevant to the topic.

Table 1 details the ten literature searches, giving a description of the topic, the initial search keywords, the number of articles returned by the initial search, and the number of articles relevant to the topic found by exhaustive evaluation. The first five topics all relate to different keywords searches, while the remaining five relate to only two different searches. In this way, we are testing prioritisation not just of different searches, but also of different topics within the same search. A range of topics are covered, most falling within the discipline of psychology, but with some (e.g., the foreign policy topics) extending into the social sciences more generally.

All of the initial searches returned a large number of articles, ranging from 327 to 606. Importantly, the relative number of relevant articles varies significantly, ranging from a very small fraction (e.g., 3 out of 327), to a significant minority (e.g., 127 out of 342). This variation allows us to test the effectiveness of the two models for different base-rates of relevant articles.

Within the datasets, the returned articles were represented using cues defined by standard PsycINFO fields. For each field, the entire text entry for the article was considered, and common English words (such as "the" and "a") were removed using a stoplist. All of the remaining words were used to generate a cue by pairing it with the field name. For example, an article authored by Robert Goldstone would have the cues "Author = Robert" and "Author = Goldstone". The field within which a word appeared was regarded as establishing a different meaningful context for that word, and so distinct cues were created for repeated words in different fields. This means, for example, that if the word "study" appeared in both the title and the abstract of an article's entry, its representation would include both "Title = study" and "Abstract = study" cues. A complete list of the fields used to create cues is given in Table 1, together with a concrete example of a cue for each field. Across the 10 datasets, the number of cues used to represent all of the articles ranged from 4346 to 10,712, with a mean of 7447.

Results

Both the rational and one reason models were applied to the datasets by simulating their impact on the order in which articles would have been presented to users. This was done by presenting the first article in the dataset, and then using the information regarding whether or not it was relevant to update the evidence values for the cues. Using one or other of the decision models, the next prioritised article was then presented, its relevance noted, and evidence scores updated again. This process continued until all of the articles had been presented, and a record was kept of the order in which they had been seen.

Figures 1 and 2 summarise the results of 50 independent applications of both models to all ten datasets. They take the form of effort-reward graphs, relating hypothetical levels of effort (i.e., the proportion of the articles read by the user) to the resultant level of reward (i.e., the proportion of relevant articles found). Figure 1 shows the curve representing the

Table 1
The Search Keywords, Topic, Number of Relevant Articles, and Total Number of Articles for the 10 Datasets

Initial Search	Topic	Relevant	Total
"drug abuse and delinquency"	The relationship between family and teenage delinquency or drug abuse	103	450
"effects of abuse"	Child physical and sexual abuse including incestuous and sibling abuse but not spousal abuse	127	342
"social facilitation"	Social intelligence in people with intellectual disabilities	59	606
"teamwork and teams"	Studies of teamwork and team training that operated in an adventure-based setting	3	327
"extroversion and introversion"	Differences between extroverted and introverted people's ability to deal with noise disturbances	7	464
"eyewitness testimony"	The use of line-ups for identifying suspects	9	379
—	The use of polygraph procedures to determine the credibility of eyewitness testimony	4	379
"foreign policy"	The role of prime ministerial leadership styles in foreign policy decision-making	4	384
—	Foreign policy using propaganda or the impact of propaganda on foreign policy	4	384
—	Foreign policy with nuclear implications	17	384

Table 2
The PsycINFO Fields Used to Define Cues, and an Example of a Cue from Each Field

Field	Sample Cue
Document Type	"Document Type = Journal-Article"
Title	"Title = conditioning"
Author	"Author = Jones"
Journal Name	"Journal Name = Psychonomic"
Language	"Language = English"
Abstract	"Abstract = findings"
Key Phrase	"Key Phrase = group"
Age Group	"Age Group = Adulthood"
Population	"Population = Human"
Population Location	"Population Location = US"
Publication Type	"Publication Type = Empirical-Study"

mean performance of both models averaged over the datasets. For example, once a user has read the first 60% of the prioritised articles using the one reason model, they have seen about 90% of the relevant articles. Because the models have a stochastic element, arising from breaking ties when two or more articles have equal evidence, the best- and worst-case performance is indicated by error bars (where large enough to be visible). The chance level of effort-reward performance, where each extra 10% of reading yields another 10% of the relevant articles, is shown by the thin solid line, and the performance obtained by reading the articles in the default reverse chronological order used by PsycINFO is shown by the thick broken line. The thin broken lines show the best and worst

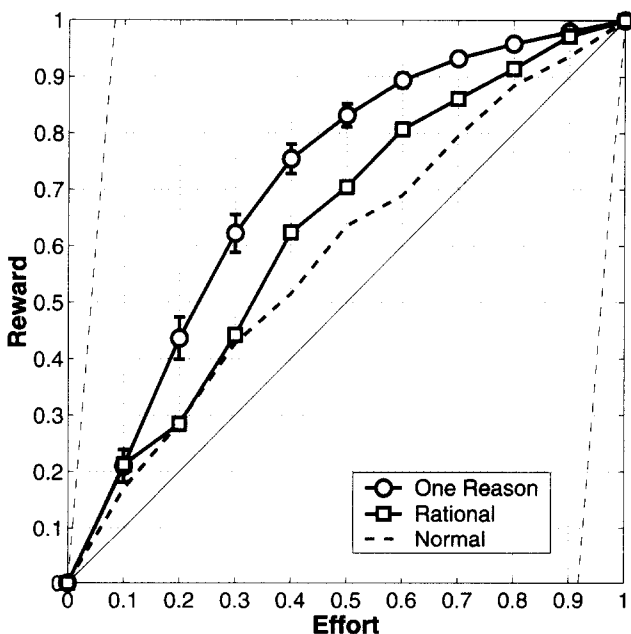


Figure 1
Effort-reward graph showing the average prioritisation performance of the Rational and One Reason models across all 10 datasets. Each dataset has been given equal weight in forming the average, and the error bars represent best- and worst-case performance across 50 independent applications of each model. The performance of the normal reverse chronological ordering, and chance performance, are also shown for comparison

possible performance corresponding, respectively, to the cases where all relevant articles are presented first, and all relevant articles are presented last.

It is clear from Figure 1 that the one reason model outperforms the rational model, and that both approaches to prioritisation are superior to either the default ordering or a random ordering. Using the one reason model, for example, the first 30% of articles contain more than 60% of the relevant ones, compared to 40% for the rational model. It is also clear, however, that neither of the models achieves anything approaching the best possible performance, and that it is necessary to read all of the articles to guarantee finding all of the relevant ones.

Figure 2 shows the weighted average performance across all of the datasets, taking into account the number of relevant articles. This means, for example, that the dataset with nine relevant articles is weighted three times as much as the dataset with three relevant articles in forming the average performance curves. In effect, this aggregation treats all of the datasets as if they were one large multi-faceted search, whereas Figure 1 treats the datasets as a series of separate searches. Under the weighted average, Figure 2 shows the one reason and rational models now have similar performance, and that they remain superior to both the default and random orderings. Once again, however, both models fall short of optimality, and all of the articles must be read to find the relevant ones.

The similar performance of the two models under a weighted average in Figure 2 suggests that the better performance of the one reason model in Figure 1 is due to data sets with small numbers of relevant articles. To test this idea, we used a measure of prioritisation effectiveness based on effort-reward performance that considered the level of the performance curve in relation to chance performance, averaged across all possible levels of effort. Geometrically, this measure

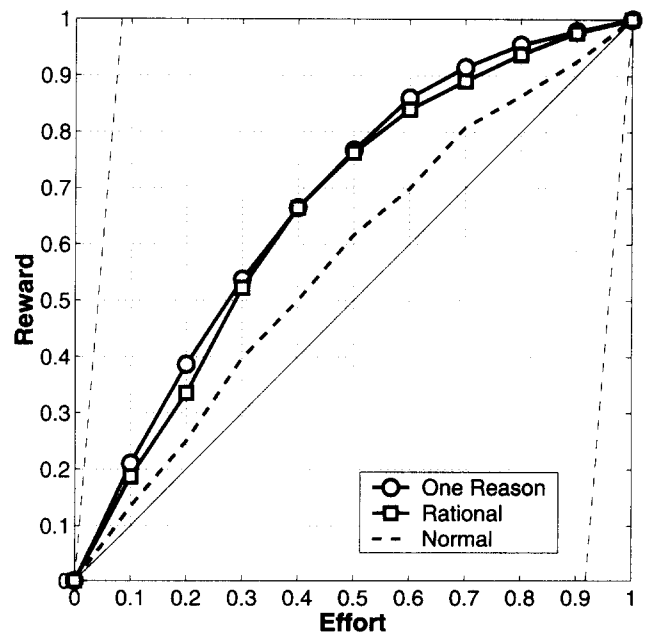


Figure 2
Effort-reward graph showing the average prioritisation performance of the Rational and One Reason models across all 10 datasets. Each dataset has been weighted according to the number of relevant articles in forming the average, and the error bars represent best- and worst-case performance across 50 independent applications of each model. The performance of the normal reverse chronological ordering, and chance performance, are also shown for comparison.

is basically the area between a model's performance curve and the chance line on an effort-reward graph. After normalisation, this measure takes the value one for best-case performance, zero for worst-case performance, and 0.5 for chance performance. Figure 3 shows the prioritisation performance measure, for both rational and one reason models, as a function of the proportion of relevant articles in the data set. This confirms that, for those data sets with a small proportional of relevant articles, the prioritisation performance of the one reason model is generally superior to the rational model. Indeed, for many of these data sets, the rational model does not perform much better than chance. For data sets with a larger proportion of relevant articles, both the models seem to perform similarly, as would be suspected from Figure 2.

The prioritisation performance measures shown in Figure 3 allow statistical inferences to be made about the differences between the rational and one reason models. In particular, it is possible to examine whether the distribution of performance scores for the two models have significantly different means and/or variances. Most psychological research tackles these problems using Null-Hypothesis Significance Testing (NHST), despite long-standing and authoritative demonstrations (e.g., Edwards, Lindman, & Savage 1963; Lindley 1972) that it is an inconsistent, incoherent, and irrational methodology for statistical inference. Accordingly, we use the Bayesian approach to model selection (e.g., Kass & Raftery, 1995; Lindley, 1972; Pitt, Myung, & Zhang, 2002; Sivia, 1996) to examine the two distributions, relying on the theory and software described in Lee (2002a).

The basic idea behind the Bayesian analysis is to consider four possibilities for the two distributions: (a) that they have the same mean and the same variance (1m1v); (b) that they have different means but the same variance (2m1v); (c) that they have the same mean but different variances (1m2v); and (d) that they have different means and different variances (2m2v). The more complicated of these accounts, such as assuming different means and variances, will of course always fit the observed data better than a simpler account, such as assuming the same mean and same variance. Bayesian

model selection, however, is not based on goodness-of-fit, but instead considers which model is *most likely* given the data, in a way that naturally balances goodness of fit with model complexity (Roberts & Pashler, 2000). In this way, the Bayes Factor between any two models can be estimated, quantifying how much more likely one is than the other (Kass & Raftery, 1995).

The results of applying these ideas to prioritisation performance is summarised in Figure 4. The four top panels show the performance measures for both the rational (black dots) and one reason (white dots) models. Each panel shows the best-fitting Gaussian distribution or distributions corresponding to the four possible assumptions about the equality of means and variances. Where different distributions are assumed, the darker lines correspond to the rational model, while the lighter lines correspond to the one reason model. Under the Bayesian analysis, the most likely account was found to be the one that assumed different means and different variances (2m2v). The Bayes Factors of the other possibilities in relation to this account are shown in the bottom panel. It can be seen that the 1m2v, 2m1v and 1m1v accounts are, respectively, 1.5, 2.6 and 4.2 times less likely.

Within the Bayesian framework for statistical inference, what constitutes a "significant difference" is a question of the standard of scientific evidence for the problem at hand, and is not automated by reference to some critical value. Bayes Factors are naturally interpreted on a meaningful scale defined by betting, so that saying one account is twice as likely as another means that we should be willing to gamble twice as much money on the first account being correct. Against this background, we draw two basic conclusions from the analysis shown in Figure 4. The first is that there is considerable evidence the one reason model has less variable (i.e., more consistent) performance, because those possibilities that assume the same variance (1m1v and 2m1v) are the least likely. The second is that there is some evidence the one reason model has better prioritisation performance, because this assumption (2m2v) is 1.5 times more likely than mean performance being the same (1m2v).

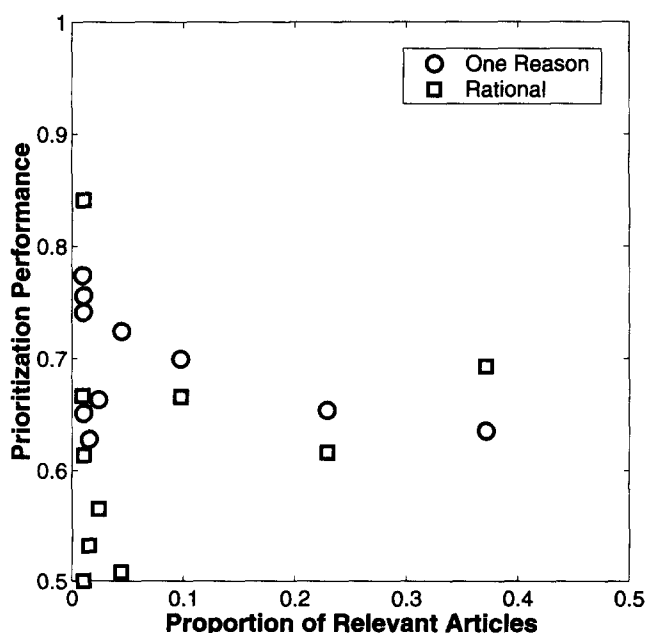


Figure 3
The prioritisation performance of both models across all of the data sets, shown as a function of the proportion of relevant articles in the data sets.

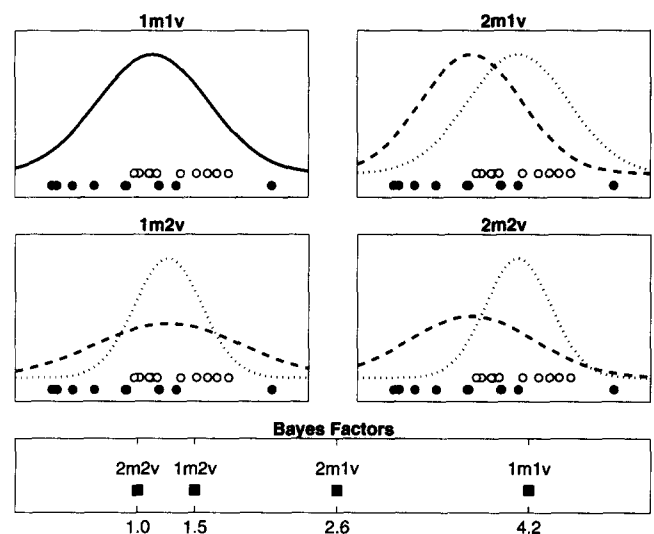


Figure 4
Bayesian analysis of the difference in prioritisation performance between the rational (black dots, darker lines), and one reason (white dots, lighter lines) models. The top four panels show the best-fitting distributions for the four possible Gaussian assumptions, and the bottom panel shows the estimated Bayes Factors in relation to the different means and different variances (2m2v) account, which is the most likely given the data.

DISCUSSION

We start our discussion by considering the applied implications or our results, before turning to their message for modeling human decision-making. Our experimental results demonstrate the potential of prioritisation in applied settings, but also show that there is some way to go before a useful applied system can be developed. While both the one reason and rational models clearly outperform the default ordering currently provided by PsycINFO, neither consistently manages to find all of the relevant articles before the user has had to do a significant amount of work. For an applied system, we suspect that something like 90% of the relevant articles would need to be found in the first 20–30% of those presented for almost every topic search. Our conclusion is that, while we believe we are on the right path, some improvement is necessary before the real-world problem of information overload is resolved without the need for refined searches.

This improvement must come from building better models of human decision-making, and it is here that our results have some clear lessons. Most importantly, the superior performance of the one reason model over the rational model suggests that more complicated decision processes are probably not the answer. The one reason model uses a single piece of information about an article in deciding it is the best available, whereas the rational model considers every piece of information about every article to make the same decision. This makes the one reason model much more computationally efficient, and so it has the potential to scale to the large volumes of information that characterise real-world problems. Given that statistical analysis of the prioritisation performance measures showed the one reason model is clearly more consistent and probably more effective than the rational model, it seems unlikely that adding more complications to the decision mechanisms will yield an improved model. In this sense, our results support Gigerenzer and Todd's (1999) claim that much of human decision-making is based on applying a simple heuristic to limited information. The relatively good performance of the one reason model suggests that, at least for some articles, users based their decisions on a single piece of information.

The real deficiency of our decision-making models, we suspect, lies in their representations. While PsycINFO users may well regard an article as relevant as soon as they see the author is "Robert Goldstone", the same probably does not apply if they can only see that the first name of the author is "Robert". This means that the important cue is something like "Author = Robert Goldstone", but our current representational approach creates only the separate cues "Author = Robert" and "Author = Goldstone". As has often been noted (e.g., Brooks, 1991; Pinker, 1998), a model of a cognitive process is only as good as the model of stimulus representation on which it relies. Using word boundaries to define cues, as we have, may be appropriate sometimes, but many of the cues to which users are sensitive seem likely to be more complicated.

For this reason, we believe that improved prioritisation performance will result from developing representational techniques that identify the cues actually used by people, and applying the one reason model to representations of the articles based on these cues. These sorts of representational techniques will, of course, also be required for developing better models of human cognition more generally within the fast and frugal framework. There are (at least) two complementary ways in which this research problem could be tackled. The machine learning community has developed a number of methods for extracting "key phrases" from textual information (e.g., Cohen, 1995) that could be tested as the basis for specifying representational cues. Alternatively, there is a large literature

within psychology that focuses upon making inferences about human mental representations from empirical measures of similarity (e.g., Shepard, 1974, 1980; Tversky, 1977; see Goldstone, 1999 for an overview). Within this framework, cues could be found by applying representational techniques such as multidimensional scaling (e.g., Cox & Cox, 1994) or additive clustering (e.g., Lee, 2002b; Shepard & Arabie, 1979; Tenenbaum, 1996) to data measuring the similarities between article summaries. It would be straightforward to evaluate both the machine learning and experimental psychology approaches, by examining the level of prioritisation performance they provided under the one reason and rational models. Both approaches offer the promise of contributing to the solution of an applied problem by improving prioritisation performance, as well as furthering our understanding of basic cognitive representational structures and decision-making processes.

FOOTNOTE

- 1 The phrase "ecological approaches" is intended here to extend well beyond Gibsonian ideas of ecological optics, and encompasses more general psychological theorising that emphasises the role the environment plays in guiding cognitive processes.

ACKNOWLEDGEMENTS

This research was supported by the Australian Defence Science and Technology Organisation, and by the University of Adelaide Faculty of Health Sciences B3 Scheme. We thank an anonymous reviewer for helpful comments.

REFERENCES

- Balabanovic, M. (1998). Exploring versus exploiting when learning user models for text recommendation. *User Modeling and User-Adapted Interaction*, 8, 71–102.
- Brooks, R.A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Brunswik, E. (1943). Organismic achievement and environmental probabilities. *Psychological Review*, 50, 255–272.
- Cohen, J.D. (1995). Highlights: Language- and domain-independent automatic indexing terms for abstracting. *Journal of the American Society for Information Science*, 46(3), 162–174.
- Cox, T.F., & Cox, M.A.A. (1994). *Multidimensional scaling*. London: Chapman and Hall.
- Doyle, J. (1999). Rational decision-making. In R.A. Wilson & F.C. Keil (Eds.), *MIT encyclopedia of the cognitive sciences* (pp. 701–703). Cambridge, MA: MIT Press.
- Edwards, W., Lindman, H., & Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review* 70(3), 193–242.
- Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (1995). *Bayesian data analysis*. London: Chapman & Hall.
- Gigerenzer, G., & Goldstein, D.G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103(4), 650–669.
- Gigerenzer, G., & Todd, P.M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Goldstone, R.L. (1999). Similarity. In R.A. Wilson & F.C. Keil (Eds.), *MIT encyclopedia of the cognitive sciences* (pp. 763–765). Cambridge, MA: MIT Press.
- Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Lee, M.D. (2002a). *Are these two groups of scores significantly different? A Bayesian approach*. Manuscript submitted for publication.
- Lee, M.D. (2002b). Generating additive clustering models with limited stochastic complexity. *Journal of Classification*, 19(1), 69–85.
- Lee, M.D., Chandrasena, L.H., & Navarro, D.J. (2002). Using cognitive decision models to prioritise e-mails. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, 478–483.
- Lindley, D.V. (1972). *Bayesian statistics: A review*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Macskassy, S.A., Dayanik, A.A., & Hirsh, H. (1999). Emailvalet: Learning user preferences for wireless email. *Proceedings of Learning About Users Workshop, IJCAI'99*.

- Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification. *Psychological Review*, 85, 207–238.
- Mehran, S., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A Bayesian approach to filtering junk e-mail. *AAAI-98 Workshop on Learning for Text Categorization*.
- Pinker, S. (1998). *How the mind works*. Great Britain: The Softback Preview.
- Pitt, M.A., Myung, I.J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109(3), 472–491.
- PsycINFO. (2001). Norwood, MA: SilverPlatter International.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107(2), 358–367.
- Shepard, R.N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 39(4), 373–422.
- Shepard, R.N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210, 390–398.
- Shepard, R.N., & Arbie, P. (1979). Additive clustering representations of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86(2), 87–123.
- Simon, H.A. (1956). Rational choice and the structure of environments. *Psychological Review*, 63, 129–138.
- Simon, H.A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.
- Sivia, D.S. (1996). *Data analysis: A Bayesian tutorial*. Oxford: Clarendon Press.
- Tenenbaum, J.B. (1996). Learning the structure of similarity. In D.S. Touretzky, M.C. Mozer, & M.E. Hasselmo (Eds.), *Advances in neural information processing systems*, (Vol. 8, pp. 3–9). Cambridge, MA: MIT Press.
- Todd, P.M., & Gigerenzer, G. (2000). Precis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23, 727–780.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327–352.