Information Foraging

Peter Pirolli
and
Stuart K. Card

UIR Technical Report
Funded in part by the Office of Naval Research
January 1999

This paper has been accepted for publication in *Psychlogical Review*.

Portions of this research were funded by the Cognitive and Neural Science and Technology Division, Office of Naval Research, under contract N00014-96-C-0097.

ABSTRACT

Information Foraging Theory is an approach to understanding how strategies and technologies for information seeking, gathering, and consumption are adapted to the flux of information in the environment. The theory assumes that people, when possible, will modify their strategies or the structure of the environment to maximize their rate of gaining valuable information. Field studies inform the theory by illustrating that people do freely structure their environments and their strategies to yield higher gains in information foraging. The theory is developed by (a) adaptation (rational) analysis of information foraging problems and (b) a detailed process model (ACT-IF). The adaptation analysis develops (a) *information patch models*, which deal with time allocation and information filtering and enrichment activities in environments in which information is encountered in clusters (e.g., bibliographic collections), (b) *information scent models* which address the identification of information value from proximal cues, and (c) *information diet models* which address decisions about the selection and pursuit of information items. ACT-IF is developed to instantiate these rational models and to fit the moment-by-moment behavior of people interacting with complex information technology. ACT-IF is a production system in which the information scent of bibliographic stimuli is calculated by spreading activation mechanisms. Time allocation and item selection heuristics make use of information scent to select production rules in ways that maximize information foraging activities.

INTRODUCTION

Humans actively seek, gather, share, and consume information to a degree unapproached by other organisms. Ours might properly be characterized as a species of *informavores* (Dennett, 1991). Our adaptive success depends to a large extent on a vast and complex tributary of cultural tasks that engage our physical and social environments. These tasks require increasingly sophisticated information-gathering, sense-making, decision-making, and problem-solving strategies. In this paper, we are interested in understanding these information-gathering and sense-making strategies from an evolutionary ecological perspective, treating adaptations to the flux of information in the cultural environment in much the same manner as biologists study adaptations to the flux of energy in the physical environment. Here, we propose an *Information Foraging Theory* that is in many ways analogous to evolutionary ecological explanations of food-foraging strategies in anthropology (Smith & Winterhalder, 1992) and behavioral ecology (Stephens & Krebs, 1986). The basic hypothesis of Information Foraging Theory is that, when feasible, natural information systems evolve towards stable states that maximize gains of valuable information per unit cost (see also, Resnikoff, 1989, p. 97). Cognitive systems engaged in information foraging will exhibit such adaptive tendencies. Rational analyses of the adaptive value of information foraging tasks can guide psychological theory just as they have in other domains (Anderson, 1991; Anderson & Milson, 1989).

Providing people with an independent and improved ability to access and understand available information has been a social aim for many movements at least since the Enlightenment, and it is also the aim of more mundane and practical efforts of improving modern-day productivity. Technological innovation has lead to an explosive growth of recorded information. The number of scientific journals has been growing by about a factor of 10 every 50 years since the 18th century (Price, 1963). The number of Internet hosts has been doubling about every year since 1992, and the number of pages accessible from a computer user's desktop has increased about five orders of magnitude in the last five years.[1] Computer users world-wide now have desktop access to more than 275 million publicly accessible World Wide Web (WWW) pages, growing at the rate of

[1] For purposes of rough calculation, we take a web page to be like a file and assume a user had about 1000 files in 1993 vs 275 million World-Wide Web pages in 1998. This calculation is actually conservative, since it does not take into account the very large increase in the number of Internet users during this period, the increases in CD-ROMs, and the increases in electronic mail.

7.5 pages every second (estimated in March 1998, Bharat & Broder, 1998). Similar, if less spectacular, observations could be made for other information sources.

Such growth triggers (and is triggered by) adaptations in human information technology, since human minds, although growing in number, are limited in their ability and available time to keep pace. Providing people with access to more information is not the problem. Rather, the problem is one of maximizing the allocation of human attention to information that will be useful, a point eloquently made by Simon:

> What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. (H.A. Simon as quoted by Hal Varian, Scientific American, Sept. 1995, p. 200).

For Information Foraging Theory, a central problem in information gathering and sensemaking is the allocation of attention. Information Foraging Theory could also provide the scientific basis from which we might engineer new innovations that enrich the information that people process.

The Task Environment of Information Foraging

It has been argued that most of our everyday tasks can be characterized as ill-defined problems (Reitman, 1964; 1965). Such tasks require substantial acquisition and integration of knowledge, typically from external sources (Simon, 1973), in order to better define goals, available courses of action, heuristics, and so on. Such tasks might include choosing a good graduate school, developing a financial plan for retirement, developing a successful business strategy, or writing an acceptable scientific paper. The structure of processing and the ultimate solution are, in large part, a reflection of the particular external knowledge used to structure the problem. Consequently, the value of the external information may often ultimately be measured in the improvements to the outcomes of an embedding task.

The structure of the interface between people and information repositories in the external world determines the time costs, resource costs, and opportunity costs of different information foraging and sensemaking strategies. Such costs include access, recognition, and handling costs, which can be weighed against the rate at which useful information is delivered to an embedding task. Our analyses will often concentrate on developing an understanding of the amount of valuable information per unit time that is yielded by an interface between people and information repositories. Our basic assumption is that people will modify their strategies, or modify the structure of the interface if it is malleable, in order to maximize their rate of gaining valuable information. A cognitive strategy will be superior to another if it yields more useful information per unit cost. Over time, we would expect strategies to evolve to improve returns on foraging. Alternative designs for information access systems may be compared on same grounds. We might assume that people prefer, and consequently select, designs that improve returns on information foraging. Over time, we would expect information technologies to evolve to improve foraging returns. In the face of selection pressures, natural and artificial information systems will evolve towards stable states that maximize gains of valuable information per unit cost

Structures that are so adapted may often be recognized in physical workspaces that are home to recurrent tasks. Research on office organization (Case, 1991; Malone, 1983; Soper, 1976) shows that action items associated with ongoing tasks are most readily at hand, often in stacks and piles on office surfaces, a personal archive is located in conventional shelves and office furniture, and other archival information is stored or available at further distances from the office, for instance in libraries. The close coupling between the access cost of information and propensity for being used is also noted in studies of what information is read or cited (Soper, 1976). Resnikoff (1989, pp. 112-117) presents a mathematical analysis showing that the particular hierarchical structure of the common library catalog card system minimizes manual search time. Faced with information foraging tasks, we expect that cognitive structures and strategies will also evolve to maximize information gains per unit cost, given the opportunity to evolve through learning and practice.

Evolutionary-Ecological Models of Foraging

We have drawn heavily upon models and techniques developed in *optimal foraging theory* (Stephens & Krebs, 1986), which seeks to explain adaptations of organism structure and behavior to the environmental problems and constraints of foraging for food. Optimal foraging theory originated in attempts to address puzzling findings that arose in ethological studies of food seeking and prey selection among animals (Stephens & Krebs, 1986). It has had an enormous impact in anthropology (Smith & Winterhalder, 1992), where it has been used to explain dietary choice (Kaplan & Hill, 1992), variations in land tenure and food sharing (Smith, 1987), group size (Smith, 1981), habitat choice (Cashdan, 1992), time allocation (Hames, 1992), and many other aspects of hunter-gatherer culture. Independent of our own efforts, Sandstrom (1994) has suggested that optimal foraging theory may successfully address the complex empirical phenomena that arise in the library sciences. We would like to think that the information foraging adaptations we observe are *exaptations*[2] of the behavioral plasticity that humans evolved for food-foraging, but it is unlikely that we will be able to obtain data relevant to tracing this evolution.

Our working heuristic is to try to understand the degree to which information foraging behavior is adaptive given the environmental context in which it occurs. We call this *adaptation analysis*, noting that in biology "adaptation is demonstrated by observed conformity to *a priori* design specifications...The hand is an adaptation for manipulation because it conforms in many ways to what an engineer would expect, *a priori*, of manipulative machinery; the eye is an optical instrument because it conforms to expectations for an optical instrument." (Williams, 1992, p. 40, italics in original). Adaptation analysis is a kind of engineering analysis that is to be considered a proper component of evolutionary-ecology explanations (Winterhalder & Smith, 1992).

In this regard, optimization models[3] are a powerful tool for studying the design features of organisms and artifacts.[4] Optimization models in general include the following three major components:

- *Decision assumptions* that specify the decision problem to be analyzed. Examples of such information foraging decisions include how much time to spend processing a collection of information or whether or not to pursue a particular type of information content.

- *Currency assumptions*, which identify how choices are to be evaluated. Information foraging theory will assume information value as currency. Choice principles include maximization, minimization, and stability of that currency.

- *Constraint assumptions*, which limit and define the relationships among decision and currency variables. These will include constraints that arise out of the task structure, interface technology, and the abilities and knowledge of a user population. Example of constraints include the rate at which a person can

[2] An adaptation to one purpose that becomes adapted to another.
[3] We prefer the terminology of the natural selection theorist G.C. Williams (1992) who uses "optimization model" rather than "optimality model" to acknowledge a focus on the optimization process and corrective tendancies rather than the attainment of global optimal states.
[4]The use of optimization models in biology has had its controversies. See for example the famously critical "spandrels" paper of Gould and Lewontin (1979), the eloquent adaptationist stance of Mayr (1983), and the readable overview of these arguments by Dennett (1995).

navigate through an information access interface, or the value of results returned by bibliographic search technology.

In general, all activities can be analyzed according to the value of the resource currency returned and costs incurred, which are of two types: (1) *resource costs* and (2) *opportunity costs* (Hames, 1992). Resource costs are the expenditures of calories, money, etc. that are incurred by the chosen activity. Opportunity costs are the benefits that could be gained by engaging in other activities, but are forfeited by engaging in the chosen activity.

As argued above, information foraging is usually a task that is embedded in the context of some other task. The value and cost structure of information foraging is consequently defined in relation to the embedding task and this often changes dynamically over time. The *value* of information (Repo, 1986) and the *relevance* of specific sources (Saracevic, 1975; Schamber, Eisenberg, & Nilan, 1990) are not intrinsic properties of information-bearing representations (e.g., documents) but can only be assessed in relation to the embedding task environment.

The use of optimization models should not be taken as a hypothesis that human behavior is classically rational, with perfect information and infinite computational resources. A more successful hypothesis about humans is that they exhibit *bounded rationality* or make choices based on *satisficing* (Simon, 1955). However, satisficing can often be characterized as localized optimization (e.g., hill-climbing) with resource bounds and imperfect information as included constraints (Stigler, 1961). Also, "One does not treat the optimization principle as a formula to be applied blindly to any arbitrarily selected attribute of an organism. It is normally brought in as a way of expanding our understanding from an often considerable base of knowledge" (Williams, 1992, p. 62). Optimization models do not imply that animals or information foragers will necessarily develop so as to embrace the simple global optimum. Rather, they describe the possibilities of a niche, a possible advantageous adaptation if not blocked by other forces (for example, the consequences of another adaptation).

> Organisms are never optimally designed. Designs of organs, developmental programs, etc. are legacies from past and natural selection can affect them in only two ways. It can adjust the numbers of mutually exclusive designs until they reach frequency-dependent equilibria, often with only one design that excludes alternatives. It can also optimize a design's parameters so as to maximize the fitness attainable with that design under current conditions. This is what is usually meant by optimization in biology. An analogy might be the common wooden-handed, steel-bladed tool design. With different parameter values it could be a knife, a screw driver, or many other kids of tool–*many*, but

> not *all*. The fixed-blade constraint would rule out turning it into a drill with meshing gears. The wood-and-steel constraint would rule it out as a hand lens. (Williams, 1992, p. 56, italics in original)

Analogies Between Food Foraging and Information Foraging

Imagine a predator, such as a bird of prey, that faces the recurrent problem of deciding what to eat, and we assume that its fitness, in terms of reproductive success, is dependent on energy intake. Energy flows into the environment and comes to be stored in different forms. For the bird of prey, different types of habitat and prey will yield different amounts of net energy (energetic profitability) if included in the diet. Furthermore, the different food-source types will have different distributions over the environment. For the bird of prey, this means that the different habitats or prey will have different access or navigation costs. Different species of birds of prey might be compared on their ability to extract energy from the environment. Birds are better adapted if they have evolved strategies that better solve the problem of maximizing the amount of energy returned per amount of effort. Conceptually, the optimal forager finds the best solution to the problem of maximizing the rate of net energy returned per effort expended, given the constraints of the environment in which it lives. These constraints include the energetic profitabilities of different habitats and prey, and the costs of finding and pursuing them. This is the essence of conventional models in optimal foraging theory (Stephens & Krebs, 1986).

An analogous situation in information foraging theory might be an office worker or academic researcher facing the recurrent problems of finding task-relevant information. Information flows into the environment to be represented in different types of external media, such as books, manuscripts, or on-line documents. The different information sources (or repositories) will have different profitabilities, in terms of the amount of valuable information returned per unit cost in processing the source. In addition, the different kinds of sources will be distributed in the task environment in different ways. Some will be more prevalent, or less effortful to access, than others. Conceptually, the optimal information forager is one that best solves the problem of maximizing the rate of valuable information gained per unit cost, given the constraints of the task environment. These constraints include the profitabilities of different sources, and the costs of finding and accessing them.

Information Patches: Problems of Time Allocation to Activities

Patch models in optimal foraging theory concern situations in which the environment of some particular animal has a "patchy" structure. For instance, imagine a bird that forages for berries found in patches on berry bushes. The forager must expend some amount of *between-patch* time getting to the next food patch. Once in a patch, the forager engages in *within-patch* foraging, and faces the decision of continuing to forage in the patch or leaving to seek a new one. Frequently, as the animal forages within a patch, the amount of food diminishes or depletes. For instance, our imaginary bird would deplete the berries on a bush as it ate them. In such cases there will be a point at which the expected future gains from foraging within a current patch of food diminish to the point that they are less than the expected gains that could be made by leaving the patch and searching for a new one. Quantitative formulations of patch models in optimal foraging theory determine the optimal policies for allocating time to foraging within a food patch vs searching for new patches.

By analogy, the task environment of an information forager often has a "patchy" structure. Information relevant to a person's information needs may reside in piles of documents, file drawers, office book shelves, libraries, or in various on-line collections. Information patches could be relatively static on-line collections such as WWW sites, or temporary collections constructed by a WWW search engine in response to user queries. Often the information forager has to navigate from one information patch to another—perhaps from one pile to another, from one on-line collection or WWW site to another, or from one search engine result to another. Often the person is faced with decisions much like our imaginary bird: how should time be allocated among between-patch foraging tasks and within-patch foraging tasks?

Conceptually, the empirical examples analyzed in this paper exhibit two kinds of between-patch activities that we discuss next: (a) *enrichment activities* and (b) *scent-following* activities. Typically, these activities will be intertwined in observed foraging activities. We will present models addressing these activities. These models assume that information foragers allocate their time to between-patch vs within-patch foraging activities in ways that optimize their overall rate of gaining valuable information per unit cost.

Problems of Enrichment vs Exploitation

The traditional patch models of optimal foraging theory deal with an unmoldable environment. The forager must optimize its selection of feasible strategies to fit the constraints of the environment. The information forager, however, can often mold the environment to fit the available strategies. We call this process *enrichment*.

One kind of environmental enrichment is to reduce the average cost of getting from one information patch to another. That is, the forager can modify the environment so as to minimize the between-patch foraging costs. As we noted above, office workspaces tend to evolve layouts that seem to minimize the between-patch search cost for needed information. Such enrichment activities create the trade-off problem: should one invest in reducing between-patch foraging costs, or should one turn to exploiting the patches?

A second kind of environmental enrichment involves making information patches that yield better returns of valuable information. That is, the forager can modify the environment so as to improve within-patch foraging results. For example, one may invest time in constructing and refining keyword queries for a search engine so that it returns lists with higher proportions of potentially relevant document citations. One may also enrich information patches by using filtering processes. For instance, people often filter their readings on a topic by first generating and filtering bibliographic citations and abstracts. Many computer systems for electronic mail, news, and discussion lists now include filters. Such enrichment activities create the trade-off problem: should one continue to enrich patches to improve future within-patch foraging or should one turn to exploiting them?

Information Diet and Scent-following: Problems of Selecting and Pursuing Items to Process

Information foraging often involves navigating through spaces (physical or virtual) to find high-yield patches. For instance, imperfect information at intermediate locations is used by the forager to decide on paths through a library or an on-line text database to target information. Such intermediate information has been referred to as “residue” by Furnas (1997). In keeping with foraging terminology, we have elsewhere called this *scent* (Pirolli, 1997). Information scent is the (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues, such as bibliographic citations, WWW links, or icons representing the sources.

Diet models in optimal foraging theory deal with situations in which an organism lives in an environment containing a number of potential kinds of food sources. Conceptually, the organism faces the problem of constructing the diet that optimizes its gain of energy per unit cost.[5] For instance, a species of predator may exist in

[5] Foraging theorists have also dealt with the more complex problem of optimizing mixtures of nutrients (Stephens & Krebs, 1986).

environments with several species of potential prey. These prey may have different amounts of *prevalence*, or may be more or less faster to find. The prey may differ in the amounts of energy they provide (perhaps because of their size), and may differ in the amount of time they take to *handle* (e.g., in pursuing and consuming them). If *profitability* of prey is defined as the energy returned per unit of handling time, then clearly less profitable prey should be ignored if they would prevent the predator from the opportunity to pursue a more profitable prey. For example, a predator that relentlessly pursued small hard-to-catch prey while large easy-to-catch prey were equally available would have a suboptimal diet. It has been noted in biology that predators will often ignore potential low-profitability prey in order to seek out higher-profitability prey. It has also been noted that diets broaden to include more prey species, or narrow to include less, depending on the prevalences and profitabilities of prey species in their habitats.

By analogy one may think of an information forager as an information predator whose aim it is to select information prey so as to maximize the rate of gain of information relevant to their task. These information prey might be relevant documents or document collections. Different sources will differ in their access costs or prevalences, and they will differ in profitability. The profitability of an information source may be defined as the value of information gained per unit cost of processing the source. For instance, physical and electronic mail may come from a variety of sources that have different arrival rates and profitabilities. Clearly, low-profitability junk mail should be ignored if it would cost the reader the opportunity of processing more profitable mail. We might also expect the diet of an information forager to broaden or narrow depending on the prevalences and profitabilities of information sources.

Our notion is that the proximal perception of information scent is used to assess the profitability and prevalence of information sources. These scent-based assessments inform the decisions about which items to pursue so as to maximize the information diet of the forager. In contrast to conventional diet models in optimal foraging theory (Stephens & Krebs, 1986), which are static, our model of scent-following is dynamic. As the state of the forager changes through the foraging process, the forager must make search decisions based on imperfect proximal information. Examples of these imperfect proximal cues include bibliographic citations or abstracts, or the text snippets found on WWW pages that represent linked documents. Scent-following is very much like heuristic search studied in human problem solving and in artificial intelligence. If scent is sufficiently strong, the forager will be able to make the correct choice at each decision point. If there is no scent, the forager would perform a random walk, either literally in physical space or metaphorically in abstract search space. These two extreme search regimes have different characteristic cost functions.

We first present data from two field studies to illustrate the general phenomena of interest in information foraging. These are then used as context for presenting optimization models adopted from theories of food-foraging strategies. From these models, we develop a cognitive model, called ACT-IF, that is matched to data collected form people using a system for browsing and searching large collections of electronically stored documents.

EXAMPLES OF INFORMATION FORAGING

We begin with descriptive analyses of information foraging "in the wild." In these examples, we have examined information foraging as embedded in knowledge intensive work we call *knowledge crystallization tasks*. A knowledge crystallization task is one in which a person gathers information for some purpose, makes sense of it, then packages it into some form for communication or action. The results could be a briefing, a short paper, or even just a decision. Knowledge crystallization tasks are characterized by the use of large amounts of heterogeneous information, ill-structured problem-solving, but a relatively well-defined goal having to do with a selection and compaction of information relative to some purpose. Information foraging occurs as part of these tasks, interleaved with activities for making sense of the information obtained and activities for generating some action or product.

As examples, we present field studies of two knowledge crystallization tasks: (1) an individual task of writing a business intelligence newsletter and (2) a group task in which MBA students do research for a strategic business analysis report. Both cases are descriptions of information-intensive work, sampled from ongoing activities of the participants. Both consist of the ill-structured activities required to sift through information available from a set of heterogeneous sources and develop a product that crystallizes the information into a more easily assimilated form. These field observations provide some sense of information foraging in the messy real world and motivate the analyses and models presented later.

Example 1. Business Intelligence Newsletter

We studied the task of a professional technology analyst who, among other duties, writes a set of monthly newsletters, each covering a specific topic in material science or computer science. These newsletters go to a select set of subscribers to this premium service. Other analysts at the firm write similar newsletters on other topics, and the newsletters are only one of an integrated set of business intelligence services.

Method

The analyst was interviewed in his office and described his work in detail, making specific reference to the materials in evidence for newsletters in progress or recently completed as well as the other activities of his job. The analyst was asked to give the interviewer an annotated tour of all of the materials in the office, including filing systems and their organization. Later, the analyst was asked questions that elicited his way of organizing the knowledge content of his job. The interview and the analyst's office were videotaped for later analysis. We also collected samples of the analysts products and videotaped as much of his working materials as feasible.

Results and Discussion

We present an analysis of the information flow in this analyst's task environment in order to illustrate foraging activities that involve (a) scent-detection processes that serve to judge the potential relevance of information sources and (b) enrichment activities that successively filter information sources to improve the future rates of return of relevant information per unit cost. We also present the layout of the analyst's workspace and the arrangement of work piles, which are one kind of information patch. This layout appears to minimize foraging costs. The evolution of the workspace to this layout would have been one kind of enrichment activity. These analyses will be used to motivate formulations of patch foraging models.

<u>Information Needs</u>

Before presenting the analysis of information flow and physical layout it is worth sketching the information needs of the analyst. We noted in the interviews and in the analysts' information products that a set of concepts were used repeatedly. The concepts can be used to describe the information of relevance to the production of the analyst's newsletters. These concepts form a knowledge schema used to forage and make sense of incoming information. Informally (for details see, Pirolli & Card, 1997), the analyst produces newsletters that identify: (a) the *players* in a research and development field, (b) the *markets* for the technology, (c) *applications* of technological innovations and technology-based application opportunities, and (d) *timing* issues concerning items to watch, recent developments, industry and market trends, and indicators of future changes.

The newsletters produced by the analyst were organized by these categories, and his filing system was organized by these categories. The categories were used for clustering and annotating articles within work piles. We assume that the analyst's sensemaking and foraging activities largely proceed by recognizing instances of the categories in the materials scanned (e.g., the entry of a new player into the industry). The analyst is an information forager with multiple information needs defined by this schema.

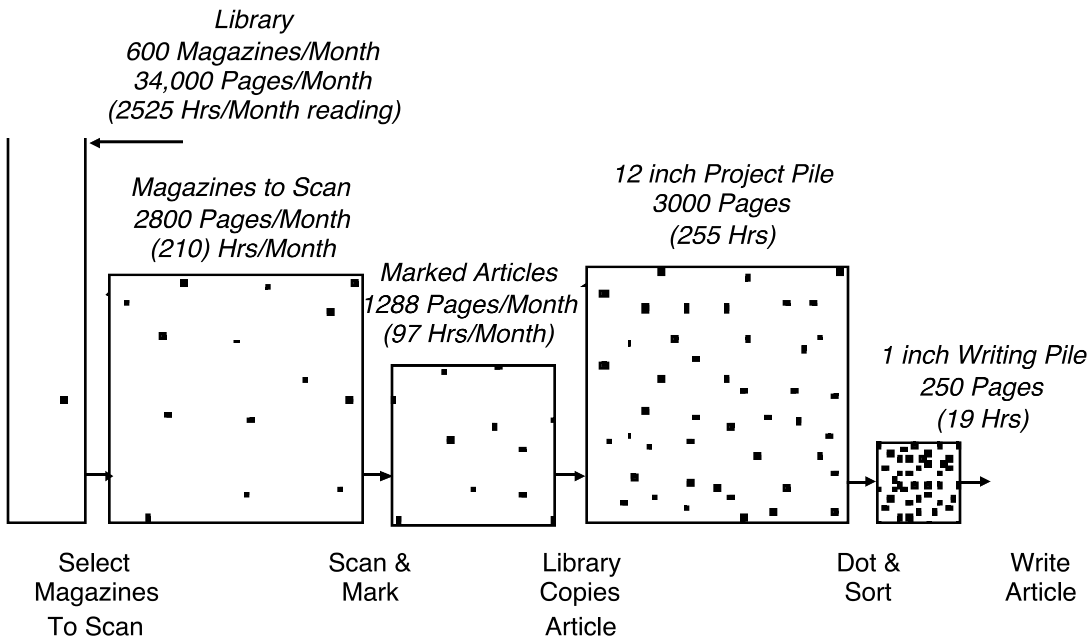

Figure 1. Condensed information flow for Business Intelligence Newsletter example. Width indicates time investment in activities, height indicates total documents, dark fill indicates relevant documents, and white fill indicates irrelevant documents.

<u>Information Flow</u>

As mentioned above, the analyst worked for an organization that publishes business intelligence reports. New trade magazine issues are received by the organization's library and physically circulated via routing slip to staff members. The analyst scans the new issues, marking articles (using an accompanying form) to be copied for himself. He also marks articles to be copied and sent to other analysts who would likely be interested in them, but do not receive that particular magazine. In return, the analyst will receive copies of articles routed to him by other analysts. The issues go to the next person on the routing slip and eventually back to the library, where the indicated articles are copied and distributed to the appropriate analysts. When the analyst receives a pile of articles, he adds them to 12" high piles, one for each of the four newsletters or reports that he regularly authors. From time to time, each pile is cleaned up by filing some articles and discarding others so that it stays about constant size. As part of his work, he sorts the articles in the pile, marking with a dot those of special interest or those that can be grouped around a theme and collecting a subpile about 1" thick to use in the next issue. From this subpile, he begins to shape his report. At this point he telephones various contacts and people suggested by the articles and also uses their information for the report. Throughout much of this process the analyst judges the relevance (or scent) of articles by scanning titles and skimming, rather than by fully reading them. In general, people often perform enrichment activities on representations, such as titles or citations, that are less informative but also less costly to process than the full source document. People seem willing to accept some imperfection in their assessment of information value in return for lower costs of processing.

From the interview and by estimating the sizes of the piles, it is possible to estimate the search space reduction as a consequence of these activities. Figure 1 is a simplified version of the analyst's workflow. The analyst's organization receives about 600 magazines. Of these, the analyst receives about 50 magazines/month for an estimated 500 articles (2800 pages/month). As a way of stating this load in terms of reading time equivalents, if he were to read these at 200 words/min it would require around 210 hrs. Based on the number of articles marked at the time of our visit, we estimate that about 230 articles/month (1288 pages/month or 97 equivalent reading hrs) are marked. The projects piles on the analysts' desk serve as buffers holding about 3000 pages, but the 1" writing pile on which he bases a newsletter holds about 45 articles (250 pages or 19 hrs equivalent reading). Thus the scanning and culling activities appear to produce the following kinds of enrichment of within-patch foraging:

- The enrichment activities yield a total time cost reduction (in terms of potential reading time) of a factor of 12 over the 600 journals subscribed to by the group.
- Enrichment activities increase the proportion of relevant articles among those under consideration, (although it is impossible to assess this quantitatively in this case). This means that when the analyst turns to working with a particular pile that the rate of processing relevant articles per unit time has been increased by prior enrichment activities.

- The sorting of articles within piles to produce the 1” subpiles at the top of the 12” piles serves to further reduce foraging costs over and above the layout of piles discussed next.

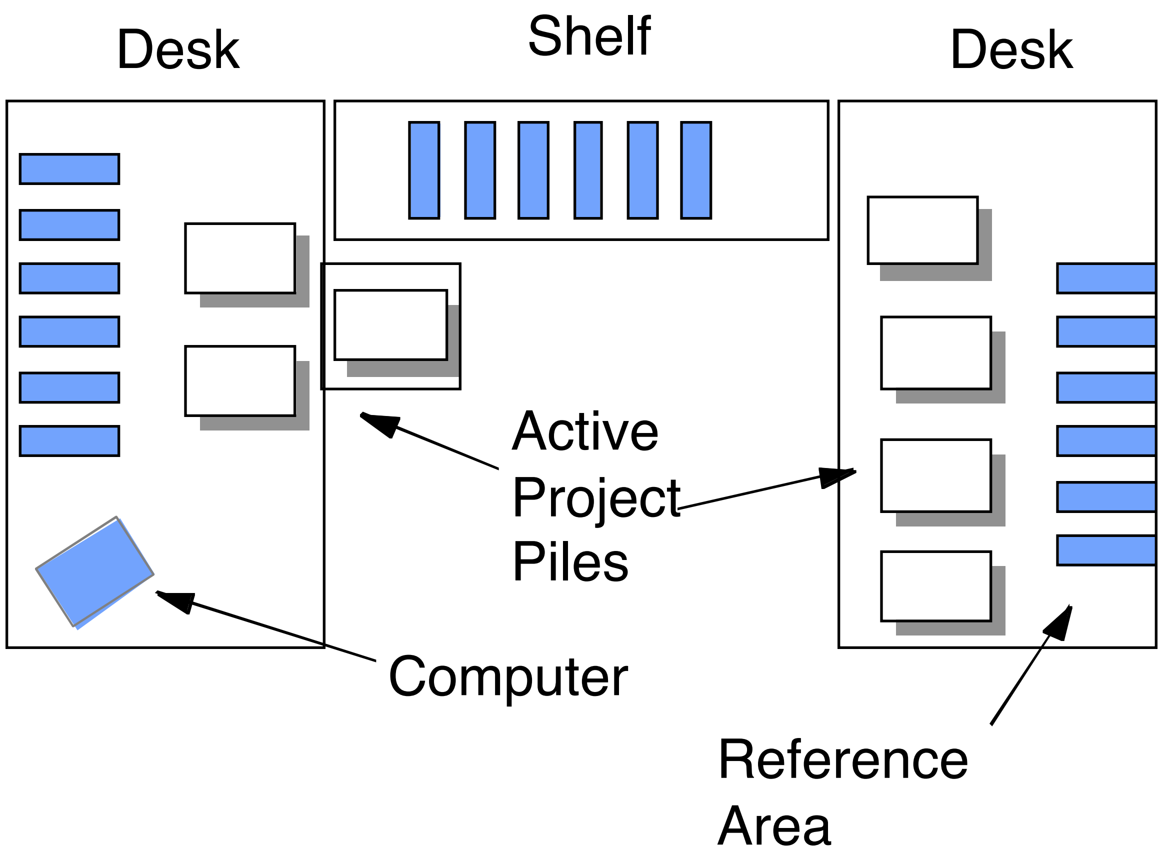

Figure 2. Schematic layout of the Business Intelligence office.

Physical Workspace

The analyst’s workspace was set up so as to allow his tasks to proceed with efficiency. Figure 2 gives a schematic picture of the space. There is a primary workspace where the analyst sits and can work on his computer together with place for several piles or pages to be placed. On the surfaces surrounding this primary place is a secondary area of surfaces where other piles can be placed. Several projects have more than one file; in each case, there is a main pile associated with some other information, such as related books or articles that should get filed. We identified four such pile groups plus a small set of piles on shelves related to social activities. These open pile groups allow the analyst to switch quickly among his major tasks. Surrounding this area is a set of reference materials such a directories, handbooks, text books, or supplies. In the desk is a set of file drawers, forming a third tier of project storage. Finally, several steps away from the chair is a set of filing cabinets, a fourth level of storage. Overall, the office held an impressive amount of paper relative to its volume.

The layout seems to exhibit structures that one would expect if it were evolving to maximize the rate of information foraged per unit time cost. In particular, the following between-patch enrichment seems to have taken place:

- The arrangement of task- or topic-related information into physically localized clusters reduces the overall costs of accessing items when engaged in the relevant task. When engaged in a particular task, or working on a particular topic, the analyst can localize foraging to patches of relatively high proportions of relevant documents, and minimized costs of access. This optimization policy is well-known in the design of virtual memory systems for computers (Denning, 1980).
- Clusters of task-related information, such as piles, files, and books, seem to be arranged such that those with higher frequency of access are placed in areas that have lower cost of access. This is, in general, a necessary condition for the optimal arrangement of information over storage media with different costs.

Example 2. Strategic Management Analysis[6]

Our second example involves analyses of one of two teams of MBA students who were studied while researching and writing a strategic analysis report.

Method

MBA students in a strategic management course at a state university were asked to participate in this research as they were working on a regular assignment for the course. Our investigation followed the work of two teams made up of two and three individuals each. The results reported are from one of the teams although data from both groups are similar. Each team was observed by two researchers. Each team had negotiated to congregate and meet with the observers at the university library. Within the library, meeting rooms were set aside for the teams to work in. While team members worked in the library proper (i.e., in the stacks, reference areas, microfiche room, etc.), the observers recorded activities in semi-structured field notes. Generally, the participants

[6]This work was conducted in collaboration with John van Gigch of the California State University, Sacramento.

made clear what they were doing, otherwise the observers would ask for clarification. Work done in the reserved meeting rooms was video taped: group discussions were recorded as were individual think-aloud protocols when people worked solo. All printouts, copied articles, notes, and final reports were collected by the investigators.

The students had been asked to write a report containing a strategic management analysis of a familiar food and beverage company of their choosing from a list of such companies. In order to carry out the assignment, students had to gather library material and references. Subjects started with materials that included a description of the assignment, materials from the strategic management course (notes, handouts, etc.), and library information sheets. Additional details are contained in Pirolli and van Gigch(1995).

Results and Discussion

Information Needs

As with the business analyst, the MBA students appeared to have had a well-developed schema for judging the relevance of information. In this case the schema identified the key elements of a strategic analysis. Indeed, the pedagogical point of the exercise was to indoctrinate the students into using these schemata for just this purpose. Our evidence for this schema comes from the students' notes, handouts, textbook, and our analysis of their protocols. The schema of information needs (for details, see Pirolli & Card, 1997) included many elements relating to: (a) *external company environment* (the industry, market, competition, etc.), (b) *internal company environment* (strengths, weaknesses, opportunities, threats, etc.), and (c) *strategic plans* (problem definitions, strategies, etc.). Each of the 18 articles that were eventually retrieved and processed by the students were found to have annotations corresponding to the schema. On average each article had a mean of 5.6 such annotations and the mean length of articles was 1190 words. So, there was an average of one annotation per 214 words read that encoded the source material into the analysis schema.

Information Flow

The assignment was due on a Monday. Participants started work in the library on the prior Saturday. Saturday morning activity was split between foraging in the library vs. meeting for discussions. Library activity was divided among collecting bibliographic references, finding information sources, printing from microfiche, and copying documents. Following adjournment on Saturday, team members went home to read and summarize the collected materials. These were discussed in a meeting lasting less than one hour on Sunday. Following that meeting, the team went home to write parts of the report.

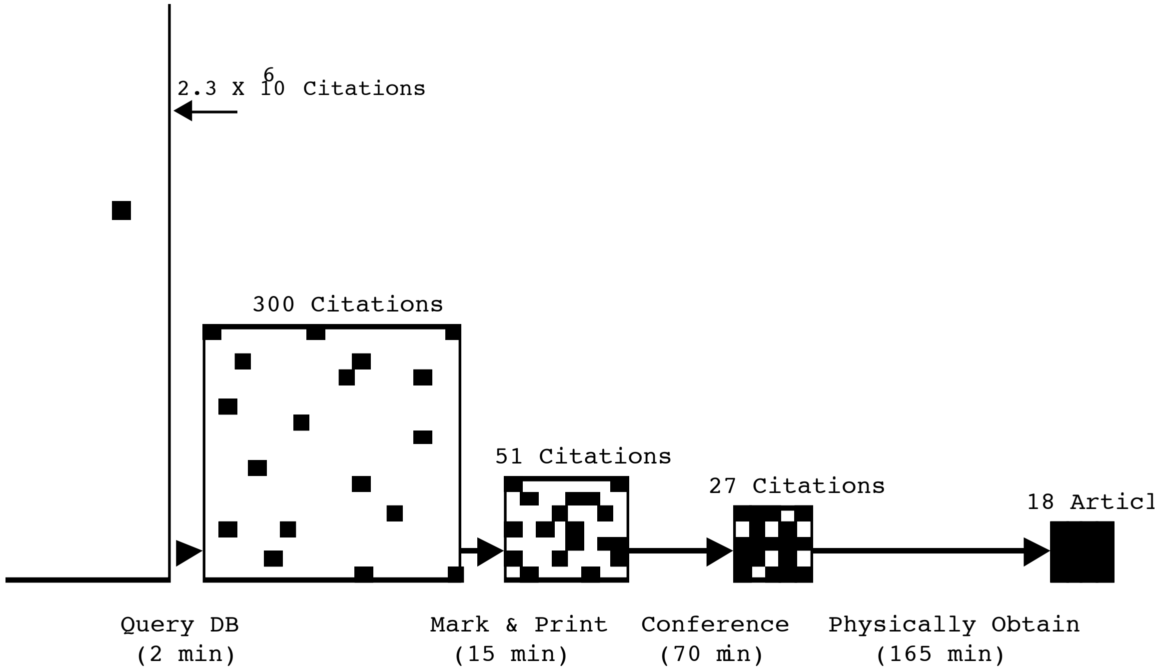

Figure 3. Information from the Strategic Management Analysis example. Width indicates time investment in activities, height indicates total documents, dark fill indicates relevant documents, and white fill indicates irrelevant documents.

We will concentrate here on the information foraging work Saturday morning. The first activity in the morning was devoted to collecting citations using a commercial on-line bibliographic system containing an estimated 2.3 million citations[7]. A query yielded about 300 citations that were then rapidly scanned while still on the computer; 51 of these were marked and printed over a span of 12 minutes. We estimate that the printing took three of the 12 minutes, so the average citation was scanned and marked in 1.8 sec. In the next three minutes, the company annual report was found and printed. These citations and the company report were the focus of the first Saturday morning meeting. In about a half-hour, the content represented by these materials was categorized into topics, evaluated for potential pursuit, and the main nine topics of the report decided upon. The meeting reduced the citations to be pursued down to 27. The students then spent the next 2 hours 45 minutes gathering the materials from around the library. At the end of this

---

[7] Matt Petrik, Information Access Company, personal communication. The estimate is for the company's General BusinessFile.

task, 18 references were retrieved (some dropped out because they were not useful once found, and some could not be found).

Thus the participants retrieved a fairly large number of reference citations and passed them through a set of filters (Figure 3). The set of filters appeared to be ordered such that those with higher rates for processing items occurred earlier in the overall foraging process. It also seems that the earlier filters may produce less accurate judgments than later ones. The early filters processed citations at the rate of 16.6 items/min, the group discussion processed .8 items/min, and physical retrieval and elimination processed .05 items/min. The overall rate of return, taking the final set of 18 articles retrieved over the 212 min foraging period was .086 items/min.

We can see the advantage of the method used by looking at the alternative cost of just reading the material. It would have taken 29.7h to read the 300 articles received based on an assumed 200 words/min reading speed. Marking the items reduced this to 5.1h, group discussion to 2.7h, and by actual retrieval to 1.8h, more than a factor of 16 overall improvement in expected reading time. It also appears that the proportion of relevant information increased among the documents under consideration (Pirolli & Card, 1997). In sum, the students showed definite sensitivity to the costs of alternate methods of obtaining the information (in so far as they could anticipate them) and arranged their methods so as to minimize those costs and maximize returns of relevant information. They acted so as to maximize information gain per unit time by using a set of filtering processes to enrich their future foraging within patches of information.

<u>Information Diet</u>

Students did not choose to read documents just because there were relevant. Instead, they selected a diet of certain types of information. In particular, they looked for short articles and the company report. For example, in one four minute interval, one team member was observed rejecting the printing of a 512 line full-text report found on-line, commenting that he was looking for short articles. He then found a 2431 line article and refused to print that as well. Finally he found a shorter 334 line article and printed it. It appears that the students evaluated the profitability of documents, in terms of potential value weighed against expected processing costs. The availability of short, more profitable articles drove out longer, less profitable articles from the information diet.

<u>Process Structure</u>

The analysis of the videotape of the Saturday morning meeting of the students reveals the processing structure they used to evaluate of the information they were gathering.

This process structure observed in the field bears strong similarity to the processing structure we studied in the laboratory, discussed below.

At the point of this meeting, the students had 51 references and the company's business plan printed on paper. Their goal was to decide which articles to retrieve and to understand how these would be used in their report. The video tape of this session was transcribed and coded using a scheme based on GOMS (Card, Moran, & Newell, 1983). The GOMS process operations were coded as well as the content generated (see Pirolli & van Gigch, 1995, for details). Team members then went through the citation printout, which was organized (by the retrieval system) into subcategories such as "advertising," "joint-ventures," etc., and judged the relevancy of category clusters and citations within the categories to a kernel plan for the strategic management report.

A goal to process the search output involved setting goals to process each of the labeled citation clusters. Processing the clusters involved goals of characterizing and evaluating the relevance of cluster topics, and characterizing and evaluating each of the citations listed under the cluster topic. Topic evaluations were largely based on (a) number of articles listed under a topic (more articles indicated importance of the topic, but only a sample of the articles were needed), and (b) relevancy to the kernel plan or the analysis outline handout. Citations were often evaluated by recency of publications. Interspersed with this processing we observed opportunistic creation and refinement of goals and policies for foraging activities and the writing of the final report. These evaluations and foraging decisions about the costs and benefits of specific items structured most of the remaining weekend activity and the structure of the final report.

General Discussion of the Examples

The two examples illustrate that: (a) there is an interleaved set of activities devoted to foraging, sensemaking, and knowledge product construction, (b) process and structure traces can be used to identify recurring activity loops, (c) schematic representations are used to judge the utility or relevance of information sources, and (d) information structures and flows in the environment can be analyzed to identify the costs and benefits that determine the rate-of-return on foraging activities. Both cases exhibit processes or structures that enrich rates of return on foraging, and processes that indicate judgments of the profitability of sources. In both field studies, people seemingly expend considerable energies in getting as much valuable information in as little time as possible.

CONVENTIONAL MODELS OF FORAGING

We now present some conventional models of optimal foraging and discuss their relation to the foraging activities illustrated in our field examples. This will serve as the grounds for our development of the ACT-IF cognitive model of foraging. Although the conventional models rest on strong assumptions, they provide a number of basic qualitative results that appear to be broadly applicable even when the assumptions are relaxed (Stephens & Charnov, 1982). Below, we apply the conventional models and the ACT-IF model to data obtained from an experiment on human information foraging.

Let us begin with a simplifying assumption that a forager's activities may be divided into two mutually exclusive sets: (1) between-patch (e.g., searching for the next item) and (2) within-patch (e.g., exploiting an item). For now, we leave the notion of patch undefined: it may be a collection of documents or an individual document viewed as a collection of content. Later, we discuss a model that applies when between-patch and within-patch activities can be carried out in parallel.

Let R be the rate of gain of valuable information per unit cost. We can characterize this (Holling, 1959) as the ratio of the total net amount of valuable information gained, G, divided by the total amount of time spent between-patches, T_B, and exploiting within patches, T_W,

$$R = \frac{G}{T_B + T_W} \quad \text{information-value-units/cost units.} \tag{1}$$

(The Appendix lists the definitions of variables used in models throughout this paper.) We may make some assumptions that allow us to construct a version of Equation 1 based on averages. We assume that (a) the number of patches processed is linearly related to the amount of time spent in between-patch foraging activities, (b) the average time between processing patches is t_B, (c) the average gain per item is g, and (d) the average time to process patches is t_W. Then,

$$\lambda = 1/t_B, \tag{2}$$

is the average rate of encountering patches. The total amount of information gained can be represented as a linear function of between-patch foraging time as,

$$G = \lambda T_B g. \tag{3}$$

Likewise, the total amount of within-path time can be represented as,

$$T_W = \lambda T_B t_W. \tag{4}$$

Equation 1 may be re-written as

$$\begin{aligned} R &= \frac{\lambda T_B g}{T_B + \lambda T_B t_W} \\ &= \frac{\lambda g}{1 + \lambda t_W}. \end{aligned} \tag{5}$$

This is what's known as Holling's Disc Equation (Holling, 1959). It serves as the basis for deriving other foraging models. Stephens and Charnov (1982) have shown that broadly applicable stochastic assumptions lead asymptotically to Equation 5 as foraging time grows large.

Using Equation 5 as context, we can now state more precisely the meaning of prevalence and profitability, as well as their impact on the overall rate of gaining valuable information per unit cost. The *prevalence* of information patches in the environment is captured by λ (the rate of encountering patches) and the *profitability*, π, of patches is the ratio of net value gained per patch to the cost of within-patch processing,

$$\pi = g / t_W. \tag{6}$$

Increasing the profitability of within-patch activities increases the overall rate of gain, R. Decreasing the between-patch costs, t_B (or equivalently, increasing prevalence λ) increases the overall rate of return R towards an asymptote equal to the profitability of patches, $R = \pi$.

Time Allocation Within and Between Information Patches

The conventional patch model of optimal foraging theory (Stephens & Krebs, 1986) is an elaboration of Equation 5. It addresses the optimal allocation of total time to between-patch activities vs within-patch activities, under certain strong assumptions. Rather than having a fixed average gain per patch and a fixed average within-patch cost the patch model assumes that (a) there may be different kinds of patches and (b) that the total gains from a patch depend on the within-patch foraging time, which is under the control of the forager (this is the decision variable for the patch model).

The patch model assumes that there may be different kinds of information patches, which we may index using $i = 1, 2, \ldots P$. The patch model assumes that the forager must expend some amount of time going from one patch to the next. Once in a patch, the forager faces the decision of continuing to forage in the patch or leaving to seek a new one. For a particular type of patch, the function $g_i(t_{Wi})$ in Figure 4, represents the cumulative amount of valuable information returned as a function of within-patch foraging time t_{Wi}. In this example, there is a linear increase in cumulative within-patch gains up to the point at which the patch is depleted. This might occur, for example, for an information forager who collects relevant citations from a finite list of citations returned by a search engine, where the relevant items occur randomly in the list. As the forager processes the items, the cumulative gain function increases linearly, and when the end of the list is reached the patch is depleted and the gain function plateaus.

We assume that patches of type i are encountered with a rate λ_i as a linear function of the total between-patch foraging time, T_B. Now imagine that the forager can decide to set a policy for how much time, t_{Wi} , to spend within each type of patch. The total gain could be represented as,

$$
\begin{aligned}
G &= \sum_{i=1}^{P} \lambda_i T_B g_i(t_{wi}) \\
&= T_B \sum_{i=1}^{P} \lambda_i g_i(t_{wi}).
\end{aligned} \tag{7}
$$

Likewise, the total amount of time spent within patches could be represented as,

$$
\begin{aligned}
T_W &= \sum_{i=1}^{P} \lambda_i T_B t_{wi} \\
&= T_B \sum_{i=1}^{P} \lambda_i t_{wi}.
\end{aligned} \tag{8}
$$

The overall average rate of gain will be

$$
\begin{aligned}
R &= \frac{G}{T_B + T_W} \\
&= \frac{T_B \sum_{i=1}^{P} \lambda_i g_i(t_{wi})}{T_B + T_B \sum_{i=1}^{P} \lambda_i t_{wi}}. \\
&= \frac{\sum_{i=1}^{P} \lambda_i g_i(t_{Wi})}{1 + \sum_{i=1}^{P} \lambda_i t_{Wi}}.
\end{aligned} \tag{9}
$$

Equation 9 is the *patch model* for information foraging—our first variant of Equation 5. Figure 4 illustrates graphically how the average rate of gain, R, can be seen to vary with different time allocation policies. Figure 4 shows three possible within-patch time allocation policies, t_1, t_2, and $t*$. To see graphically the average rate of gain R that would be achieved by different policies, we plot lines, such as R_1, R_2, and $R*$, from the origin and intersecting with the gain function, g_i, at a particular within-patch time policy, such as t_1,$t_{2,}$, or $t*$. The slope of these lines will be the average rate of gain because the slope will correspond to the amount of value gained from patches $g_i(t_{Wi})$ divided by the time spent in between-patch activities t_{Bi} and the time spent within patches t_{Wi}. For cases such as Figure 4 (linear but finite gains), a line, $R*$, tangent to g_i, and passing through the origin gives a slope equal to the optimal average rate of gain, and an optimal within-patch time allocation policy of $t*$. A forger should stay in such linear gain patches until the patches are exhausted (and no longer than that). When the patch is exhausted the forager should move on the next patch. All of this fits our intuitions. As we shall discuss in the next section, complexity is added when one considers patches that yield other kinds of gain functions.

We want to emphasize the graphical reasoning used in Figure 4, since it can be used again in the next section. Imagine drawing a series of lines as follows. First imagine drawing a line from the origin to the leftmost point of the gain curve, which would correspond to zero within-patch time. Such a line would be just a horizontal overlay of the x-axis. It would have zero slope, indicating zero rate of gain. Next, imagine drawing a series of lines intersecting the gain curve at points corresponding to successively larger investments of within-path time. The slopes of these lines would at first progressively increase until they were equal to $R*$, then decrease. In general, this holds true for linear gain functions that plateau, such as Figure 4, or the diminishing returns curves discussed in the next section: As one increases the time allocated to within-patch foraging there is, at first, an progressive increase in the overall gain rate, up to an optimal point, then there is a decrease in the overall gain rate. This is the result of opportunity cost: Rather than continuing to spend time in a patch that is now prodicing low yields, one should be moving on to find another patch.

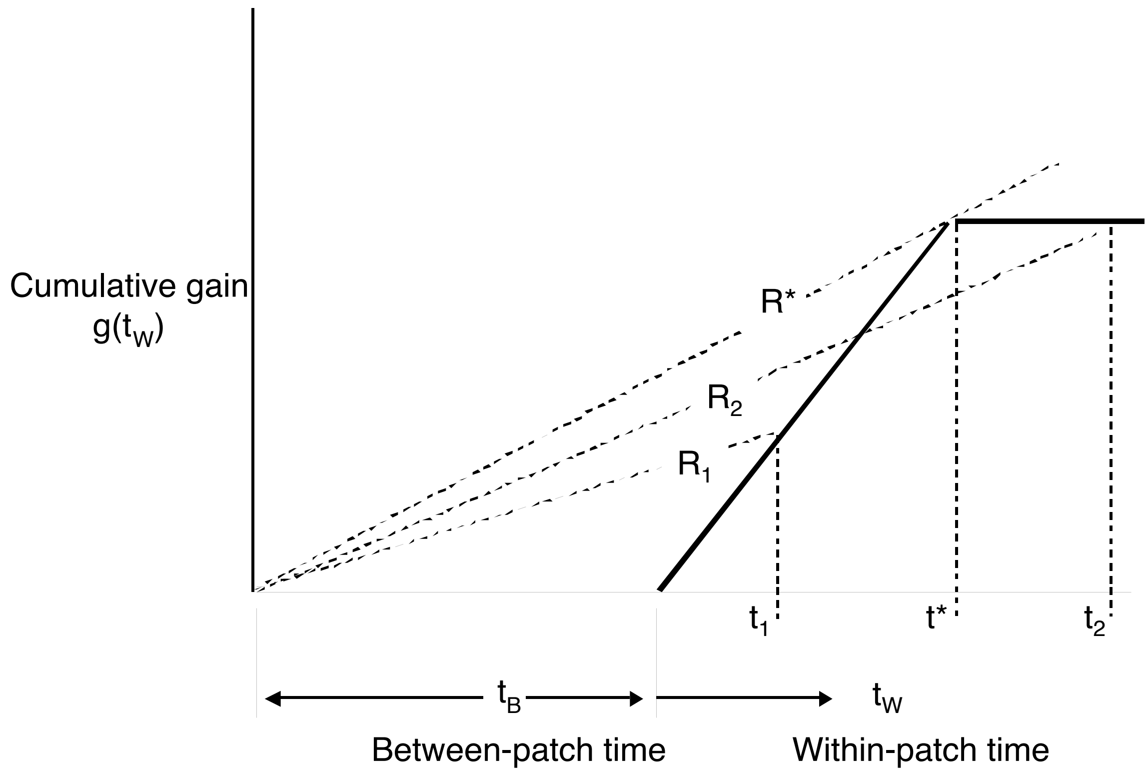

Figure 4. A linear, finite cumulative within-patch gain function (solid line). Dashed lines, Ri, have slopes equaling the average rates of gain produced by different within-patch time (tW) allocation policies.

Charnov's Marginal Value Theorem

Often, there will be diminishing returns as a function of within-patch foraging time, as in Figure 5. This might occur, for example, for an information forager who collects relevant citations from a list that has been automatically ranked with elements that are more likely to be relevant at the beginning of the list. It may also occur because of redundancy in such a list—because items encountered later on the list replicate information encountered earlier in the list. In such cases there will be a point at which the expected future within-patch gains diminish to the point that they are less than the expected gains that could be made by leaving the patch and moving to a new one. Later we discuss how such expectations are assessed in ACT-IF.

Charnov's (1976) *Marginal Value Theorem* was developed to deal with the analysis of time allocation for patches that yield diminishing returns. The theorem, presented in detail in the Appendix, deals with situations in which foraging within a patch has a decelerating cumulative gain function, such as those in Figure 5a. The theorem predicts that a forager should remain in a patch so long as the slope of g_i (i.e., the marginal value of g_i) is greater than the average rate of gain R for the environment.

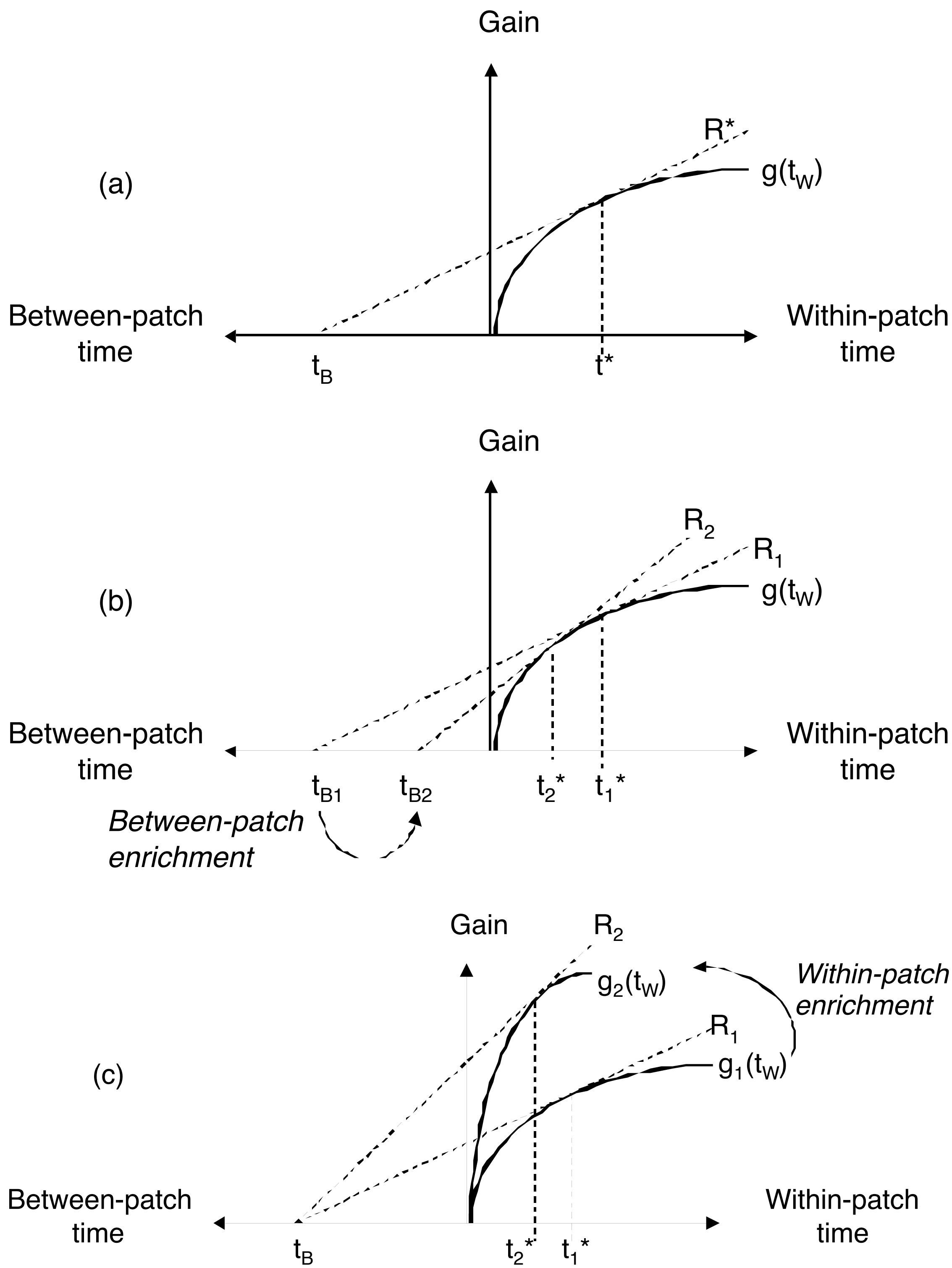

Figure 5. (a) Charnov's Marginal Value Theorem states that the rate-maximizing time to spend in patch, t*, occurs when the slope of the within-patch gain function g is equal to the average rate of gain, which is the slope of the tangent line R, (b) the average rate of gain increases with decreases in between-patch time costs, and (c) improvements in the gain function also increase the average rate of gain.

Figure 5 shows graphical representations of Charnov's Marginal Value Theorem that appear in many discussions of optimal foraging theory. Figure 5a captures the basic relations for the situation in which there is just one kind of patch-gain function. The prevalence of patches in the environment (assuming random distribution) can be captured by either (a) the mean between-patch search time, t_B, or (b) the rate at which patches are encountered is $\lambda = 1/t_B$. In Figure 5a, the average between-patch time t_B is plotted on the horizontal axis, starting at the origin and moving to the left. To determine the optimal rate of gain R^*, one draws a line tangent to the gain function $g_i(t_W)$ and passing through t_B to the left of the origin. The slope of the tangent will be the optimal rate of gain, R. The point of tangency also provides the optimal allocation to within-patch foraging time, t^*. The point of tangency is the point at which the slope (marginal value) of g_i is equal to the slope of tangent line, which is the average rate of gain R.

Effects of Between-Patch and Within-Patch Enrichment

Throughout our discussion of our field studies of information foraging, we noted activities that enriched future returns on foraging. Some activities seemed aimed at reducing between-patch foraging times. These included the arrangement of office layout so as to minimize costs of accessing piles of information. Other activities seemed aimed on improving the gains that would eventually be made from within-patch foraging. These included the filtering activities noted in both case studies. We may use Charnov's Marginal Value Theorem to reason qualitatively about these enrichment activities.

Figure 5b illustrates the effects of enrichment activities that reduce between-patch time costs. As between-patch time costs are reduced from t_{B1} to t_{B2} the overall rate of gain increases from the slope of R_1 to the slope of R_2, and optimal within-patch time decreases from $t_1{}^*$ to $t_2{}^*$. Reducing between-patch costs not only improves the overall average rate of gain, the optimal gain is achieved by spending less time within a patch (when the conditions satisfying Charnov's Marginal Value Theorem hold, see the Appendix).

Figure 5c illustrates the effects of enrichment activities that improve the returns from a patch. Figure 5c shows that as within-patch foraging gains are improved, from g_1 to g_2, the optimal average rates of gain improve from the slope of R_1 to R_2, and the optimal within-patch time decreases from $t_1{}^*$ to $t_2{}^*$. Again, not only does within-patch enrichment improve the overall rate of gain, it also reduces the optimal amount of time needed to spend within patches (when the conditions satisfying Charnov's Marginal Value Theorem hold, see the Appendix).

Information Diet Selection

There are many instances in which a person attempting to find relevant documents in response to a query has to decide to gather and consume the most profitable information types. Thus, the MBA students in Example 1 decided to gather company reports for the target company and short news articles. They declined to gather company reports for other companies, longer news articles, and other types of documents. Many such documents were relevant documents, but the students' judgments seemed to concern selecting or rejecting items based on their profitability.

A second variant on Equation 5 is to distinguish different types of information items, and to consider which items should be pursued. The conventional *diet model* of optimal foraging theory provides some insights concerning the selection of items during foraging. The details and derivation of the model are presented in the Appendix. The model assumes that information items (or patches) can be classified by the forager into i= 1, 2,... n types, and that the forager knows information concerning the profitability and prevalence of these items.[8] The average time between processing items of type i (or average between-patch time) is t_{Bi}, and the rate of encountering items of type i is $\lambda_i = 1/t_{Bi}$. Let g_i be the average gain of valuable information yielded by processing items of type i, and let t_{Wi} be the time required to process such items. Let D be a set representing the diet of a forager, e.g., $D = \{1, 2, 3\}$ represents a diet consisting of items of types 1, 2, and 3. The average rate of gain, R, yielded by such a diet would be given by another variation on Equation 5,

$$R = \frac{\sum_{i \in D} \lambda_i g_i}{1 + \sum_{i \in D} \lambda_i t_{Wi}}. \tag{10}$$

---

[8] We are purposely ambiguous in our interchangeable use of "item" and "patch." It may sometimes be more natural to think of things like documents as items and collections of documents as patches, however, one could conceivably develop diet models that treat collections as items , or patch models that treat documents as patches of content that require time-allocation decisions. Stephens and Kreb (1986) present a combined diet and patch model in which elements are simultaneously patches requiring time allocation decisions and item types requiring diet decisions.

Optimal Diet Selection Algorithm

If we assume that the time costs needed to recognized the item types are effectively zero, then an optimal diet can be constructed by choosing item types in an all-or-none manner according to their profitabilities (this is known as the *zero-one rule*, see the Appendix). The profitabilities of each item type, π_i, are defined as the value of the item divided by its time cost,

$$\pi_i = \frac{g_i}{t_{Wi}}. \tag{11}$$

In general (Stephens & Krebs, 1986), the following algorithm can be used to determine the rate-maximizing subset of the n types that should be selected:

- Rank the item types by their profitability, $\pi_i = g_i / t_{Wi}$ To simplify our presentation, we and let the index i be ordered such that

$$\pi_1 > \pi_2 > \ldots > \pi_n .$$

- Add item types to the diet in order of increasing rank (i.e., decreasing profitability) until the rate of gain for a diet of the top k item types is greater than profitability of the k + 1st item type,

$$R(k) = \frac{\sum_{i=1}^{k} \lambda_i g_i}{1 + \sum_{i=1}^{k} \lambda_i t_{Wi}} > \frac{g_{k+1}}{t_{Wk+1}} = \pi_{k+1} \tag{12}$$

The left side of the inequality in Equation 12 concerns the rate of gain obtained by the diet of the k highest profitability item types, computed according to Equation 10. The right side of the inequality concerns the profitability of the k+1st item type.

Conceptually, one may imagine an iterative process that considers successive diets of the item types. Initially, the diet, D, contains just the most profitable type, $D = \{1\}$, the next diet considered contains the two most profitable types, $D = \{1, 2\}$, and so on. At each stage, the process tests the rate of gain $R(k)$ for the current diet containing $D = \{1, 2, ..., k\}$ types against the profitability of the next type π_{k+1}. So long as the gain of the diet is less than the profitability of the next item type, $R(k) \leq \pi_{k+1}$, then the process should go on to consider the next diet $D = \{1, 2, ...k +1\}$. Otherwise, Equation 12 is true, the iterative process terminates, and one has obtained the optimal diet. Adding in the next item type would decrease the rate of gain for the diet.

To illustrate this graphically consider Figure 6. Figure 6 presents a set of hypothetical information items having an exponential distribution of profitabilities indicated by π_k. We assume that these items are all encountered at an equal rate of $\lambda_k = 1$. Figure 6 also presents *R(k)* calculated according to Equation 10, for diets including items up to and including each item *k*. One can see that *R(k)* increases at first as the diet is expanded up to an optimum diet containing the top two item types, and then decreases as additional items are included in the diet. The optimum, *R**, occurs just prior to the point where *R(k)* crosses under π_k . Exploration of Equation 12 shows that increasing the profitability of higher-ranked items tends to change the threshold, yielding fewer types of items in the diet. A similar diet-narrowing effect is obtained by increasing the prevalence (λ) of higher ranked items. Increases in profitability or prevalence of high-ranked items are enrichment activities that yield a narrowing of the information diet.

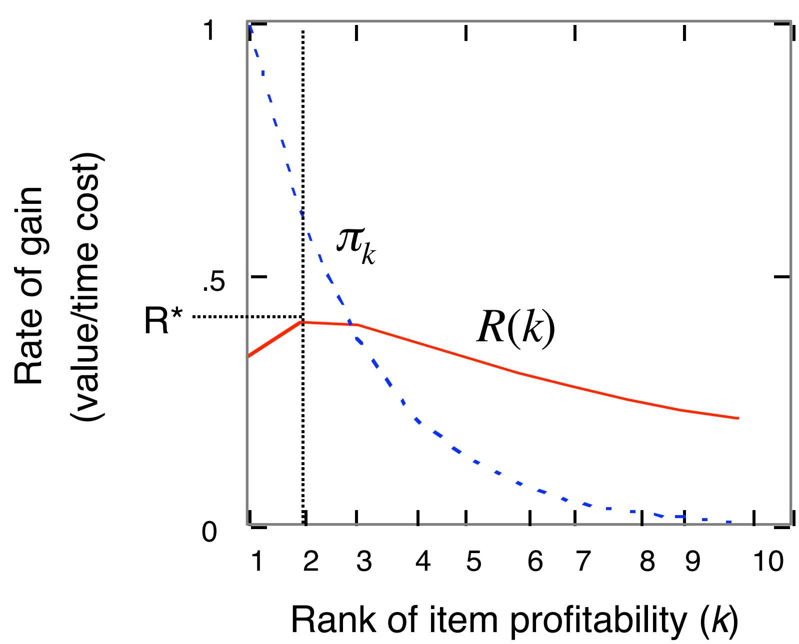

Figure 6. A hypothetical example of the relationship between profitability (π) and rate of gain (R) for diets including items 1, 2, ...k.

Principles of Diet Selection

The diet selection algorithm suggests the following:

- *Principle of Lost Opportunity* . Intuitively, the information diet model states that a class of items should be ignored if the profitability, π_i , for those items is less than the expected rate of gain, R, of continuing search for other types of items. This is because the gain obtained by processing items of that low-profitability item type is less than the lost opportunity to get higher-profitability types of items.

- *Independence of Inclusion from Encounter Rate*. A counterintuitive implication of the diet selection algorithm (Stephens & Krebs, 1986) is that the decision to pursue a class of items is independent of its prevalence. The decision to include lower-ranked items in a diet is solely dependent on their profitability, and not upon the rate at which they are encountered, λ_i. However, the inclusion of a class of items is sensitive to changes in the prevalence of more profitable classes of items. This can be seen by examination of Equation 12, where λ_i appears on the left side of the inequality but not the right side. Generally, increases in the prevalence of higher-profitability items (or equivalently increases in their encounter rates) make it optimal to be more selective. To use an everyday information foraging analogy: if reading junk mail is categorized as a too-low profitability item (because there are better things to pursue), then that decision should be made regardless of the barrage of junk mail received. Reading any junk mail would cost the opportunity of doing more profitable activities. In general, as the prevalence of profitable information increases, one should expect a narrowing of the information diet

Conventional models of optimal foraging theory—the patch model and the diet model—have generally proven to be productive and resilient in addressing food-foraging behaviors studied in the field and the lab (Stephens, 1990). They are, however, cast at a level that abstracts away from mechanisms that organisms actually use to achieve adaptive foraging strategies. The conventional models also make the strong assumption that the forager has perfect "global" information concerning the environment. Moreover, the models are static rather than dynamic (dependent on changing state or time). To make information foraging predictions at a level of behavioral analysis typically carried out in laboratory studies, we formulate a cognitive model that is dynamic, mechanistic, and having access only to information currently attended to or gained from past experience.

ACT-IF: A COMPUTATIONAL COGNITIVE MODEL OF INFORMATION FORAGING

So far, we have shown some basic predictions about information foraging. To test these predictions we specify a detailed process model. This moves the discussion from analyses that explain why actions are adaptive, to mechanistic specifications that explain how the actions are effected. This model, called ACT-IF, assumes a network representation of declarative information and a spreading activation mechanism that computes estimates of the relevance of external sources of information. ACT-IF also assumes a production rule representation of cognitive skill and a set of heuristics that select productions in ways that achieve adaptive information foraging behavior. Since ACT-IF is a behaving production system, we may compare traces of its behavior against those of human information foragers.

We present a production system model of data collected in Pirolli et al. (1996) in a study of a information system for very large collections of full-text documents, called Scatter/Gather. The production system model operates by heuristics that instantiate the information diet and information patch models. This model may be considered as an extensive revision of the ACT-R production system architecture (Anderson, 1993) that incorporates information foraging predictions. We have called our model ACT-IF in recognition of its dual heritage. We think this model is the most stringent test of the information foraging theory.

The Scatter/Gather Browser

The Scatter/Gather system (Cutting, Karger, Pedersen, & Tukey, 1992) uses the clustering of documents as the basis of a browser suitable for large numbers of documents. Figure 7 presents a conceptual overview of how a person interacts with Scatter/Gather. The system uses an automatic clustering algorithm, based on comparing the full text of documents in a collection. Scatter/Gather scatters documents into a set of automatically induced clusters. Level 1 in Figure 7 represents a set of clusters created by Scatter/Gather (only five clusters are presented in this hypothetical interaction but 10 clusters is typical in the real system). Scatter/Gather summarizes the contents of the clusters in a concise way that can be presented to users. Figure 7 uses single-word topic labels on the clusters to represent these summaries. The user may gather the documents of interesting clusters in into a new subcollection, as in Figure 7, and request that the system repartition the subcollection into a new set of clusters, as in Level 2 of Figure 7. This process may continue until a small interesting collection of documents is created, and the user decides to read them.

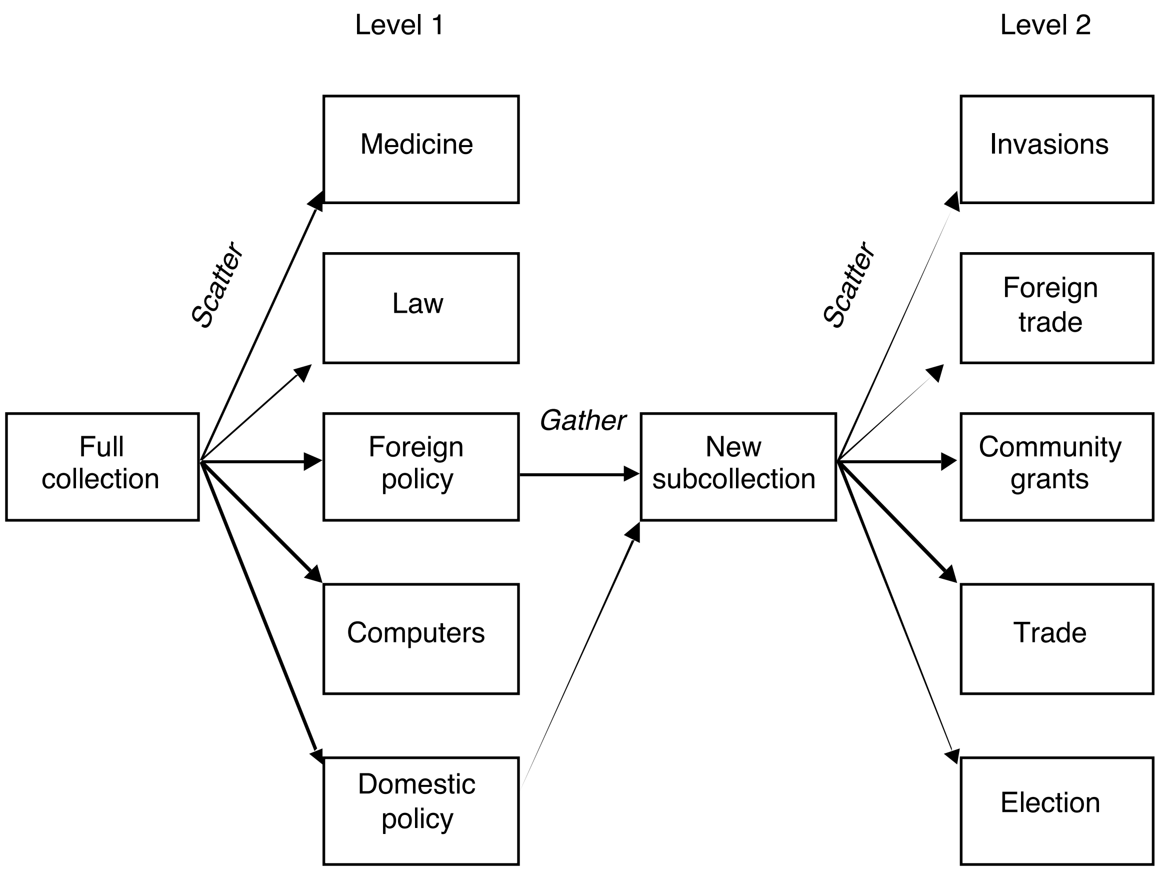

Figure 7. A conceptual overview of the Scatter/Gather interaction process.

Figure 8 presents a typical view of the Scatter/Gather interface.[9] The document clusters are separate areas on the screen. The user may gather those clusters of interest by pointing and selecting buttons above each cluster. On command, the system will pool the subset of documents in these clusters, then automatically scatters that pooled subcollection into another set of clusters. A new screen like Figure 8 is presented to the user containing the new set of clusters. With each successive iteration of scattering and gathering clusters, the total number of documents in the clusters becomes smaller, eventually bottoming out at the level of individual documents. At some point, the user may choose one or more clusters and request that the system display the titles in those

[9]This interface was developed by Marti Hearst at Xerox PARC.

clusters. That display window contains a list of document titles. By using the mouse to click on document titles, the user may bring up the full-text of a document for viewing.

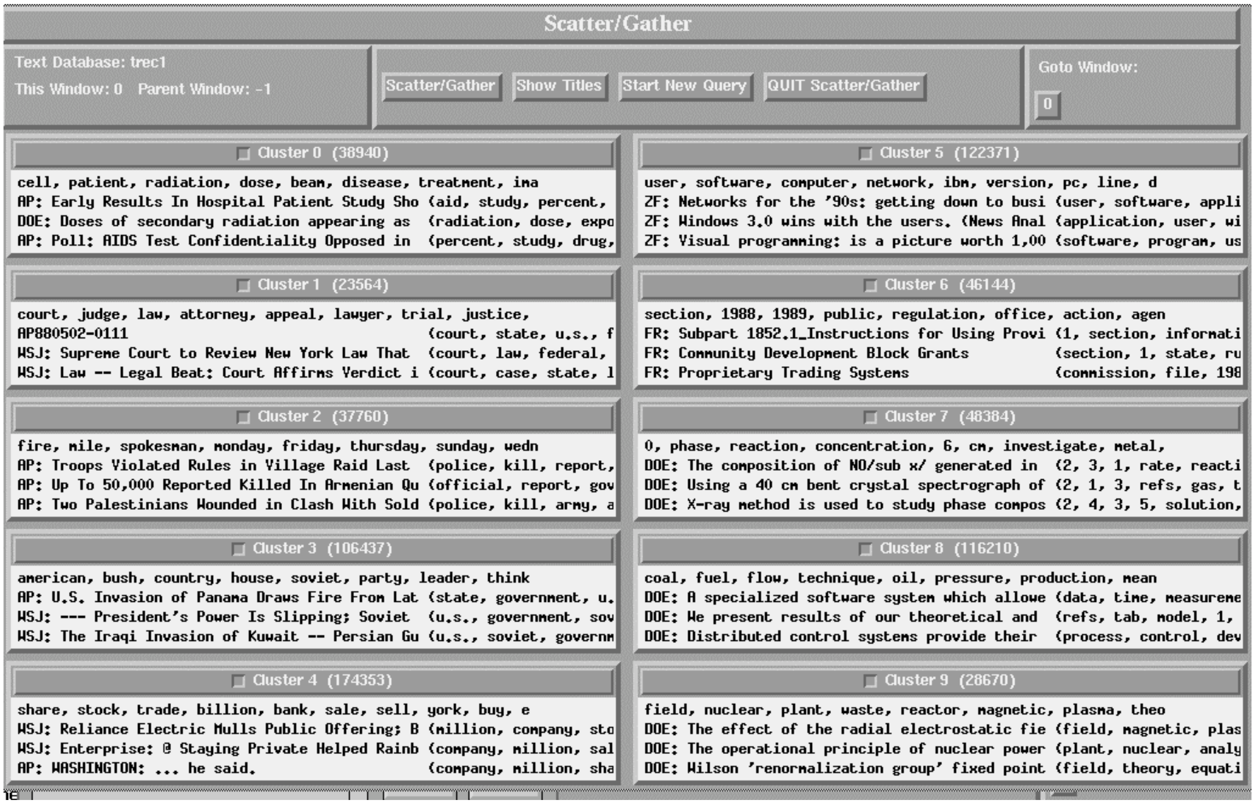

Figure 8. The Scatter/Gather document browsing interface.

Internally, the system works by precomputing a *cluster hierarchy,* recombining precomputed components as necessary. The clustering in Scatter/Gather depends on a measure of inter-document similarity. Conceptually, this approach (vanRijsbergen, 1979) represents documents as vectors of equal length, where each component of a vector is associated with one of the unique content words in the document collection. In some schemes, the component may contain a value indicating the presence of a word in the document (i.e., a binary coding). In other schemes, a vector component may indicate the frequency or some normalized frequency of a word in the document. The similarity of two documents may then be computed by a *cosine measure*, which is the cosine of the angle between two vectors, sometimes also known as a *normalized correlation.*

Scatter/Gather browsing and clustering employs methods that take the same amount of time on each iteration, independent of the number of documents clustered (Cutting, Karger, & Pedersen, 1993). The method summarizes document clusters by *meta-documents* containing profiles of topical words central to the cluster and the most typical titles. Topical words are those that occur most frequently in a cluster, and typical titles are those from documents with the highest similarity to a centroid of the cluster. These topical words and typical titles are presented to users to provide them with a summary of the documents in a cluster. This summary is called a *cluster digest*, and it is the cluster digests that appear in boxed subareas of Figure 8 to represent each cluster.

We have developed an ACT-IF model of foraging tasks with a specific version of Scatter/Gather. In that version, Scatter/Gather was applied to the TIPSTER text collection, which was created for the TREC text retrieval conference (Harman, 1993). This is test corpus used to evaluate information retrieval systems. The version we used contained 742,833 full-text documents collected from the Wall Street Journal, the Associated Press newswire, Department of Energy technical abstracts, the Federal Register, and computer articles published by Ziff-Davis. The corpus has been extensively used by the information retrieval community. Standard information retrieval tasks (queries) have been defined on it together with lists of known relevant Tipster documents, as judged by experts. The test corpus provides us with a common standard against which to compare performance. The retrieval tasks involve finding as many relevant documents as possible within a time limit.

Overview of ACT-IF

ACT-IF consists of a *production memory* and a *declarative memory*. The declarative memory models the information being attended to, goal information, and information that has been recalled from long-term declarative memory. The production memory contains production rules which are patterns of the general form *Condition* → *Action*. ACT-IF operates on a basic *match-execute cycle*. During the match phase, the condition part of the production rule patterns are matched against information in working memory. Those that match are then ranked by evaluation functions, the best match is selected, and its action pattern is executed during the execution phase. Actions specify updates to declarative memory, setting of goals, and actions to be performed in the world.

The evaluation functions that select production rules in ACT-IF were developed by refining the optimal foraging models discussed above. We will elaborate the diet model to address how Scatter/Gather clusters are selected, and we will elaborate the patch model to address how time is allocated to the (a) process of collecting clusters and reclustering them vs (b) displaying the document titles in clusters and scanning them for relevant ones.

In information foraging tasks, we assume that people must assess the relevance or utility of information based on available cues, such as bibliographic citations, abstracts, key words, titles, etc. This is what we call the detection of information scent. Our ACT-IF model of Scatter/Gather uses spreading activation mechanisms (Anderson, 1993) as an integral part of the assessment of information scent. We assume that text summaries on the Scatter/Gather interface spread activation through the declarative memory of the user, and activation simultaneously spreads from the task query. Activation levels are used by ACT-IF evaluation functions to determine which production rules are best to execute. These evaluation functions implement rate-optimizing information foraging heuristics.

Model-Tracing

For the Scatter/Gather study discussed below, we modeled the task structure with a common set of production rules. For each simulation of each individual Scatter/Gather user, we also generated a *spreading activation network* (Anderson & Pirolli, 1984) to represent words and inter-word memory associations in participants' declarative memory. These networks are used to model spreading activation effects in the evaluation of information foraging productions.

A *model-tracing methodology* (Anderson, Boyle, Corbett, & Lewis, 1990) was used to parse the logged interactions of Scatter/Gather participants and to match these logs against ACT-IF simulations (Pirolli, 1997). To model-trace (match) a participant's log file, the ACT-IF production system is initialized with (a) production rules for the task and (b) the spreading activation network for the individual. The model-tracer then parses the participants' actions from the log file, and it uses this information to maintain a model of the Scatter/Gather screen state. Changes in screen state are "perceived" by the ACT-IF production system, which means that declarative memory elements are created when corresponding objects "appear" on the screen in the screen state model. The two main types of windows of interest are (1) the Scatter/Gather windows, which present the cluster digests, and (2) titles display windows which present the titles of all the documents in a cluster. Each cluster summary consists of topical words and typical titles.

The model-tracer runs the ACT-IF production system for one cycle at a time, catching it just as it has evaluated and ranked the productions that match to the current goal and state of working memory. At this point, the ranked list of executable productions (known as the *conflict set*) serves as a prediction of the relative likelihood of potential user actions. The model tracer then examines the parsed log to determine the actual user action. The corresponding production in the conflict set is then chosen for execution in ACT-IF, and the model tracer duly updates statistics regarding the match of its predictions against the observed user actions. Following production execution, the model-tracer reads the next action from the log file, updates the Scatter/Gather screen state model, and ACT-IF updates its declarative memory in accordance with any "perceived" screen changes.

Scatter/Gather Task Structure

The model of the Scatter/Gather task in the evaluation study consists of 15 production rules. The production rules implement the task structure presented in Figure 9 and they are glossed in Figure 10. For each Scatter/Gather window (e.g., Figure 8), goals were set to process each of the clusters on the screen. Processing a clusters entailed looking at the clusters and processing the elements (the text summary) for the cluster. After a cluster had been looked at, the cluster could be selected, ignored, or deselected, or the gathered clusters could be re-clustered (scattered) or displayed.

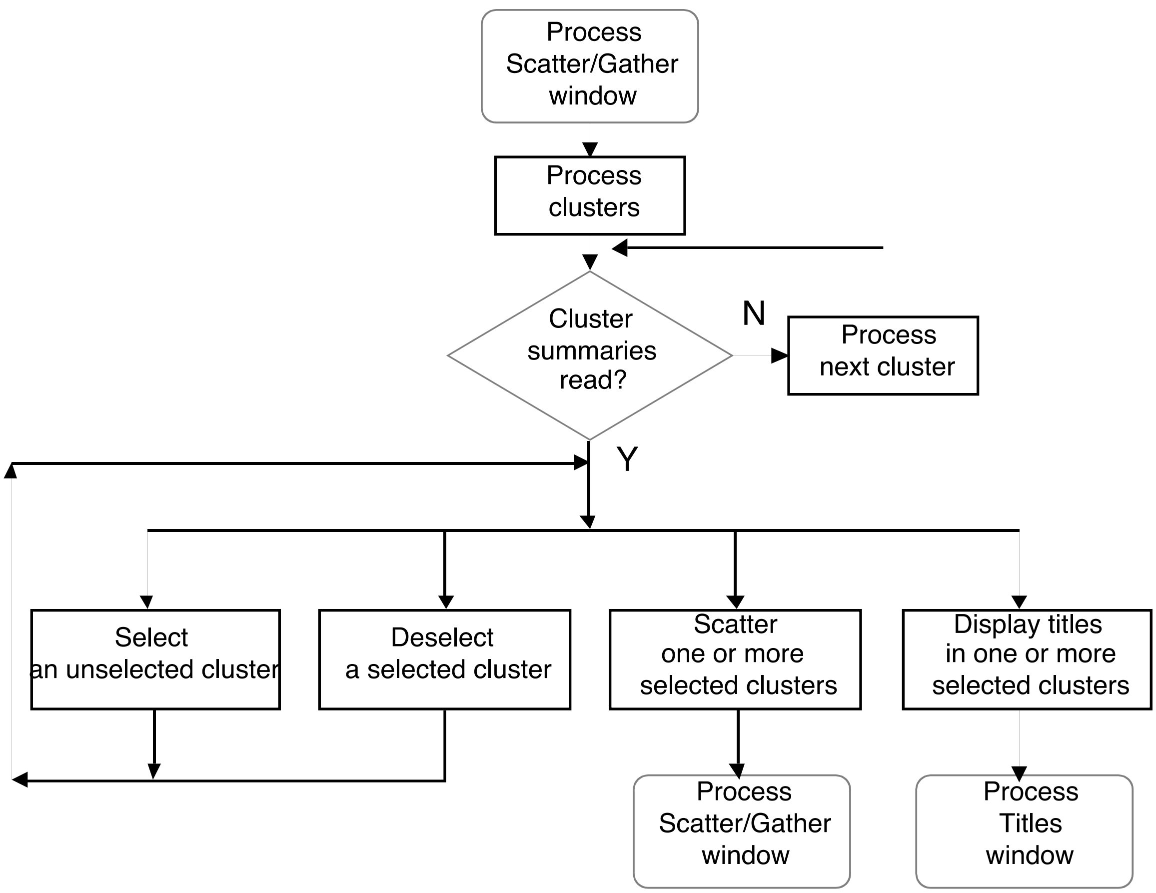

Figure 9. Task structure for processing cluster windows (Figure 8) as implemented by the ACT-IF production system model

On the left of the arrows in Figure 10 are mnemonic names for the productions and the conditions for matching declarative memory. The right side of the arrows are the actions of the production rules. Some productions are annotated with a "(2)" to indicate that there are actually two copies of the productions for the two different types of window.

NOTICE-NEW-WINDOW (2) New window on screen	→	Attend to it & set goal to process it
ATTEND-TO-WINDOW (2) Attend to window	→	Look at window
UNATTEND-TO-SCREEN (2) Goal is to process a window & different window has appeared	→	Pop the goal
SHIFT-ATTENTION Another window is present	→	Attend to that window
PROCESS-CLUSTERS Goal is to process S/G window	→	Set goal to process clusters
PROCESS-NEXT-CLUSTER Goal is to process S/G window clusters & one is unprocessed	→	Set goal to process next cluster
LOOK-AT-NEXT-CLUSTER Goal is to process next cluster	→	Look at cluster & pop the goal & set goal to process cluster elements
LOOK-AT-CLUSTER-ELEMENTS Goal is to look at cluster elements	→	Look at topics and typical titles & pop the goal
SELECT-RELEVANT-CLUSTER Goal is to process SG window & there is a query & there is an unselected cluster	→	Select the cluster
DESELECT-RELEVANT-CLUSTER Goal is to process SG window & there is a query & there is a selected cluster	→	Deselect the cluster
DO-SCATTER/GATHER Goal is to process SG window & some clusters have been selected	→	Scatter/Gather the window
DO-DISPLAY-TITLES Goal is to process SG window & some clusters have been selected	→	Display the titles in the window

Figure 10. Production rules used in the ACT-IF model of the Scatter/Gather protocols obtained in Pirolli et al. (1996).

Assessment of Information Scent by Spreading Activation

Spreading activation provides the mechanism modeling people's assessment of information scent. In this section, we describe the spreading activation model of information assessment and describe how spreading activation networks can be generated from texts in the world. These assessments are used in our ACT-IF model to predict which clusters are selected and how many.

Spreading Activation as a Bayesian Model of Relevance

Spreading activation theories of human memory generally predict how a resource called *activation* is spread from cognitive structures that reside in a focus of attention. The spread of activation from one cognitive structure to another is determined by some network representation where interstructure links weight the rate of activation flow. Spreading activation theories are usually interpreted as predicting that more activated structures will receive more favorable processing.

The ACT-IF simulations used an evaluation function that rated cluster matching productions based on the activation of the task query when a cluster summary and its words were in the focus of attention. The activation computation was based on that of ACT-R (though not actually computed by the ACT-R architecture). The activation of a query word i is

$$A_i = B_i + \sum_j W_j S_{ji} \tag{13}$$

where B_i is the base-level activation of i, S_{ij} is the association strength between cluster word j and query word i, and W_j is the base level activation of cluster word j. Following the adaptationist rationale of Equation 13 in ACT-R, we interpret Equation 13 as a Bayesian prediction of the relevance of one word in the context of other words. A_i in Equation 13 is interpreted as reflecting the log posterior odds that i is relevant, B_i is the log prior odds of i being relevant, and S_{ji} reflects the log likelihood ratios that i is relevant given that it occurs in the context of word j.

The spreading activation networks are based on the following equations used to derive the values in Equation 13. B_i reflects the log prior odds so

$$B_i = \ln\left(\frac{\Pr(i)}{\Pr(\bar{i})}\psi\right) \tag{14}$$

where $\Pr(i)$ is the probability of word i occurring in the world, $\Pr(\bar{i})$ is the probability 1 - $\Pr(i)$ that the word will not occur in the world, and ψ is a normalizing constant used to yield positive values (this was set to e^{17} based on inspection of the raw statistics). W_j is the analogous value for each word j.

s_{ij} reflects the log likelihood ratio

$$s_{ji} = \ln\left(\frac{\Pr(j \mid i)}{\Pr(j \mid \bar{i})}\psi\right) \tag{15}$$

where $\Pr(j|i)$ is the conditional probability of word j occurring in the context of word i and $\Pr(j \mid \bar{i})$ is the conditional probability of word j occurring in a context that does not contain word i.

Anderson (1993) provides arguments for the adaptiveness of Equations 13 to 15. It should be noted that the version of spreading activation embodied in Equations 13 to 15 is a mechanism that updates Bayesian *a posteriori* logarithmic odds based on *a priori* estimates and current contextual evidence.

Specification of Spreading Activation Networks

The ACT-IF simulations of Scatter/Gather require actual spreading activation networks to compute information scent assessments. These networks must specify the base-level activation and interassociation strengths for all the screen text encountered by an individual participant as they work with the Scatter/Gather interface. This requires statistics concerning the base rates and cooccurance statistics for the text. These statistics were computed from the raw text as described in Pirolli (1997). They were calculated by making use of some intermediate results produced as a side effect of building an experimental system that automatically creates a thesaurus. To create this system, Schuetze (1992) computed an index containing the base rate frequencies for all words in the Tipster corpus used in our Scatter/Gather study, as well as the pairwise cooccurance frequencies of words that occur within 40 words of one another. For each simulation of a participant interacting with Scatter/Gather we preprocessed their on-line log files, extracted all the text they encountered, looked up the relevant statistics in Schuetze's index, and generated a spreading activation network. Pirolli (1997) performed comparisons of an earlier version of ACT-IF spreading activation against two other information retrieval methods of determining the relevance of cluster digests to queries. In addition to being sensitive to raw word overlap between a query and a cluster summary, spreading activation is also sensitive to thesaurus relations (words with different surface form but related word sense). Moreover, as we discuss below, spreading activation from the Scatter/Gather interface and task queries seems to be especially adept at providing proximal assessments of the whereabouts of relevant information.

One nagging question is how well the spreading activation network computed for each person actually reflect the person's memory structure for words. Such an assumption would seem more justified if the text corpus were known to be representative of each participants' past experience with text. We do not really know this for sure, although the sheer size of the corpus might be expected to mean that its word statistics were more reflective of the world of text than a smaller sample. The assumption might also seem justified if we knew that the amount of experience with the corpus that was observed in our studies was enough to enable the users to learn the underlying corpus statistics. Again we do not know this for sure, although analyses reported in Pirolli et al (1996) suggest that people are learning quite a bit about the text corpus. Future work will be required to test the validity of the spreading activation networks obtained from corpora such as Tipster.

Combining Activation from Cues

When a person reads a citation, a summary of a document, or other pointer to an information source, it is a proximal stimulus (information scent) that provides information about some distal source of content, and the stimulus also (usually) suggests a path of access to that source. ACT-IF assumes this proximal stimulus is composed of a set of cues and spreading activation provides a model of how activation spreads from cues in the environment and features of the information goal. In the case of Scatter/Gather, the information goals are given as the task queries, and the cues from the environment are the cluster summaries and document titles presented on the Scatter/Gather interface. We now develop a model of how patterns of activation across these goal features and cues are integrated into assessments of the potential value or relevance of the distal information source.

ACT-IF assumes that cues forming a proximal stimulus, such as a cluster summary or document title, provide context for one another in determining their value or relevance to an internal representation of an information goal. The relevance or value of each cue element (e.g., a word) is not independent of the relevance of other cues. Rather, relevance or value assessment for the individual cues combines in an interactive (e.g., multiplicative) manner. ACT-IF assumes an assessment that is like the assessment of the match of stimulus cues to stored representations in exemplar-based categorization models (Kruschke, 1992; Medin & Schaffer, 1978).

For the ACT-IF model of Scatter/Gather, the assessment, $g(c, s)$ of the cluster summary text for the cluster c presented on screen s, is

$$g(c,s) = \exp\left(\frac{\sum_{i \in Q} A_i}{T}\right) \tag{16}$$

where the summation is over the activations of all the words, i, in the query, Q. T is a scaling factor that we estimate in the next section. Equation 16 is a variant of Kruschke's (1992) model of stimulus cue assessment. In the case of ACT-IF, activation levels reflect the log likelihoods that proximal cues are relevant to an information need (the query), and Equation 16 describes how these activations are integrated into a global assessment of all the proximal cues.

To estimate T, we fit activation-based assessments obtained from Scatter/Gather screen states to estimates of the actual number of relevant documents in clusters. It should be noted that, besides the spreading activation networks, that (a) T is the only parameter estimated for our simulations and (b) this estimation is done based on a priori characterizations of the information environment, not post-hoc from user data. The information scent model in ACT-IF is completely specified a priori from the structure of the information environment.

Characterization of Scent in the Scatter/Gather Environment

Match of Proximal Information Scent to Distal Structure of Scatter/Gather

Equation 16 provides us with a characterization of information scent: a prediction of the information forager's assessment of the prospects of finding relevant information from proximal cues. We wanted to understand if the proximal assessment of prospects tracked the actual prospects of finding relevant information in Scatter/Gather. Information scent may fail to track the underlying distribution of relevant documents for either of two reasons: (1) the model is wrong or (2) the Scatter/Gather interface provides a poor reflection of the underlying clustering structure. Good fits, however, should corroborate the validity of the model and the effectiveness of the interface.

For an average Scatter/Gather state s with clusters c = 1, 2, … 10, we may consider two aspects: (1) how the Scatter/Gather clustering algorithm has distributed relevant documents across the clusters and (2) how a person perceives the distribution of relevant documents across clusters, given the proximal cues available. We may then ask how well the two distributions match. We call the distribution computed by the clustering algorithm the *distal* distribution of relevant information $d_D(c)$. We call the perceived distribution the *proximal* distribution of relevant information $d_P(c)$. We now proceed to fit these distributions for the average Scatter/Gather state.

We made use of data from computational experiments on the Scatter/Gather clustering algorithm.[10] These data were calculated by an autonomous computer program that determined how many relevant documents could be found in each cluster as it traversed around the cluster hierarchy. The average proportion $d_D(c)$ of relevant documents in Scatter/Gather clusters for an average query is well-fit (R^2 = .92) by the exponential,

$$d_D(c) = \alpha \exp(-\beta(c-1)) \tag{17}$$

when the c = 1, 2, ... 10 clusters are ranked in decreasing order of proportion of relevant documents. The free parameter estimates from fitting Equation 17 to the data are $\alpha = .47$ and $\beta = .63$. Distribution $d_D(c)$ is presented in Figure 11.

Next, we computed $A(c, s)$, the summed activation received by query words from cluster summary texts (which is the numerator of Equation 16). We computed $A(c, s)$, for every cluster on every screen s available across all participant log files in the Scatter/Gather study. $A(c$, s) depends on each participant's particular queries. For every screen s we ranked the clusters c = 1, 2, ..., 10 in decreasing order of activation value. In other words, $A(1, s) > A(2, s) > \ldots > A(10, s)$, for a particular screen state s. For a particular cluster rank c we computed the average activation $\overline{A}(c)$. Using $\overline{A}(c)$ as the numerator in Equation 16, we computed the average information scent of relevant documents as $\overline{g}(c)$, for a cluster of rank c. The distribution of relevant documents across clusters on a Scatter/Gather screen, can be calculated by dividing the activation-assessment of a cluster by the total activation-assessments of all clusters on the same screen,

$$\begin{aligned} d_P(c) &= \frac{\overline{g}(c)}{\sum_{i=1}^{10} \overline{g}(i)} \\ &= \frac{\exp(\overline{A}(c)/T)}{\sum_{i=1}^{10} \exp(\overline{A}(i)/T)}. \end{aligned} \tag{18}$$

Equation 18 was fit by numerical methods[11] to Equation 17, to obtain the curve in Figure 11 and to estimate the scaling factor T.[12] Note that achieving the fit in Figure 11 with only one free parameter (T) is not always possible.

---

[10] These unpublished experiments were conducted by Marti Hearst at Xerox PARC on 29 TREC queries.

[11] Specifically, we used the Levenberg-Marquardt method as developed in the public domain MINPACK alrgorithms (More, Garbow, & Hillstrom, 1980).

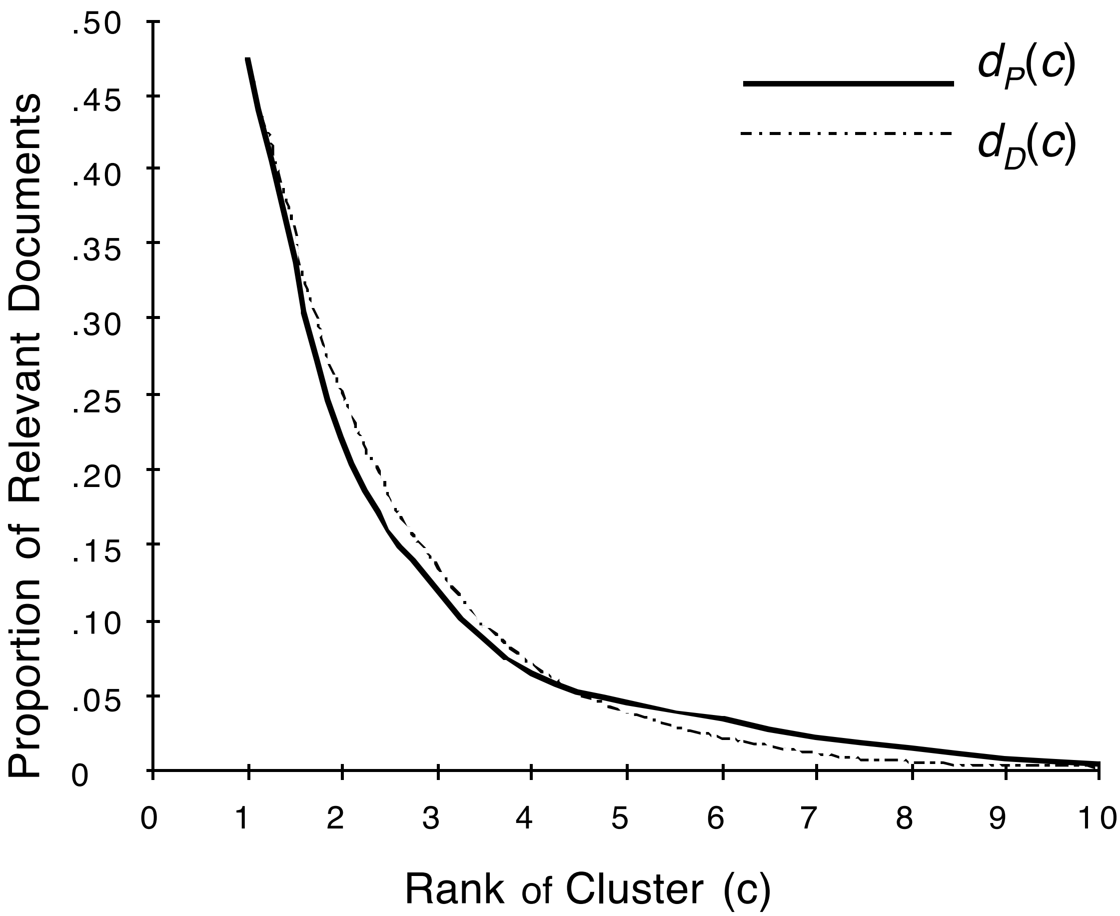

Figure 11. The underlying, distal, distribution of relevant documents, dD, characterizing the clustering algorithm, and the proximal distribution, dP, characterizing information scent from the Scatter/Gather screen. The distributions characterize the proportion of all the relevant documents (in an average system state) that fall in each cluster. Clusters are ranked in decreasing order of proportion relevant documents.

The match of $d_D(c)$ to $d_P(c)$ in Figure 11 illustrates how well the assessment of prospects from proximal cues fits the actual distribution of relevant documents across clusters in an average Scatter/Gather state. One may also examine what happens these prospects as a person moves from one Scatter/Gather state to another—as they iteratively gather clusters and then scatter them into a new set of clusters. Optimally, this iterative

[12] Equation 18 is a Boltzman equation, sometimes found in categorization research (Kruschke, 1992). The scaling factor T is known as "temperature" in the Boltzman formula.

process should reduce the total number of documents under consideration while increasing the proportion of relevant documents.

First, consider the changes in the underlying clusters as a user works with Scatter/Gather. Assume that there are no backups in the process and that people iteratively gather and scatter clusters until they finally decide to display the cluster contents. Any task will involve a sequence of Scatter/Gather cluster states, 1, 2, ..., s, ..., S produced by the iterative gathering and scattering of clusters. The basic observation is that the proportion of relevant documents across all of the clusters in state $s + 1$ should equal the proportion of relevant documents (relevant documents divided by total number of documents) in the clusters that were gathered in state previous state s,

$$\frac{relevant}{total}\ documents\ in\ all\ 10\ clusters\ in\ s+1$$
$$= \frac{relevant}{total}\ documents\ in\ all\ k\ gathered\ clusters\ in\ s.$$

This is how the proportion of relevant documents changes from one state to the next as one interacts with Scatter/Gather. It is another characterization of the distal structure of relevant information in the environment. It is a characterization of how the distal structure changes over states.

We can ask if the proximal assessment of relevant information tracks the change in distal structure from state to state. Letting $i = 1, 2, \ldots k$ index the k gathered clusters at any state, we can write this relationship as:

$$\frac{\sum_{c=1}^{10} g(c,s+1)}{\sum_{c=1}^{10} N(c,s+1)} = \frac{\sum_{i=1}^{k} g(i,s)}{\sum_{i=1}^{k} N(i,s)}, \tag{19}$$

where $N(c, s)$, is the total number of documents in a cluster in state s. Each side of Equation 19 is a proportion where the numerator is the number of relevant documents assessed by activation and the denominator is the total number of documents (this value is presented on the Scatter/Gather screen). Equation 19 says that the total proportion of relevant documents in state $s + 1$ is equal to the proportion of relevant documents in the k clusters gathered from state s.

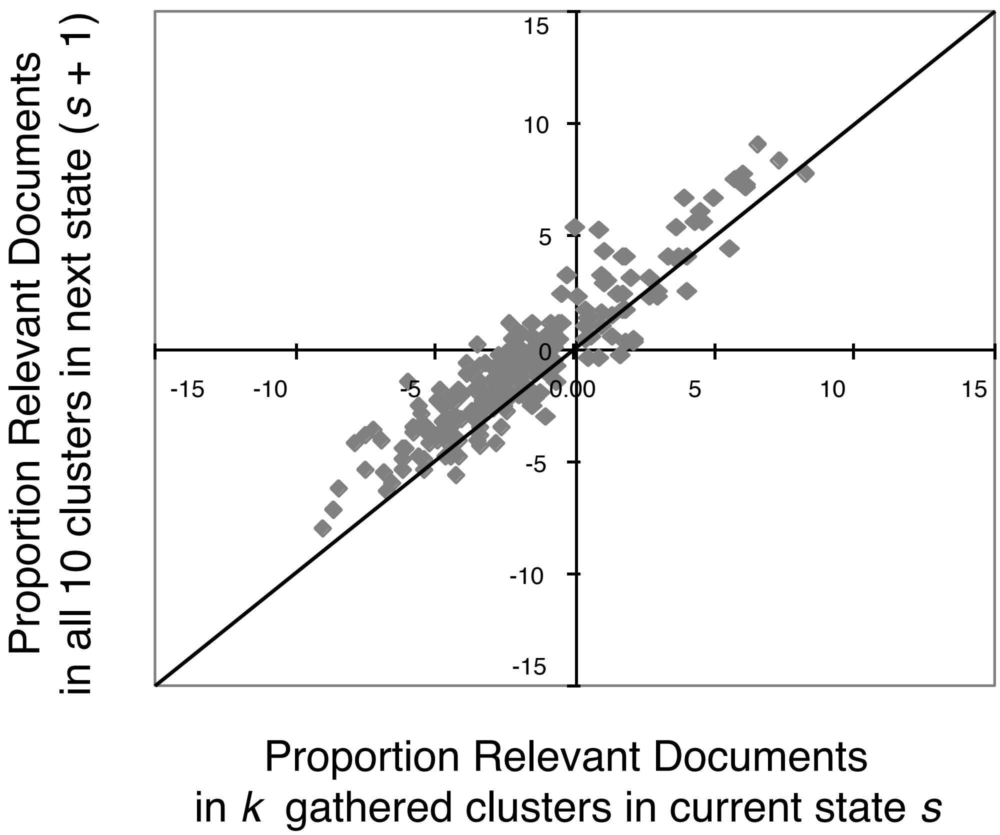

Figure 12. Expected proportion of relevant documents in a new state (s + 1) vs the expected proportion of relevant documents in the gathered clusters from previous state s. The expected values are computed by information scent assessments by model-traces of ACT-IF. Logarithmic transformations have been performed on both dimensions. See text for details.

Figure 12 plots the Equation 19 from data obtained from our log files. We found all screen states, *s*, in which a person gathered clusters and then scattered them into a new screen state, *s* + 1. We used ACT-IF to compute the information scent provided by each cluster summary in each state using the scaling parameter *T* as estimated above. Figure 12 plots each one of these *s* to (*s* + 1) transitions (*N* = 302) as points, where the abscissa plots the model values for the right side of Equation 19 and the ordinate plots the model values for the left side of Equation 19 (both scales are logarithmic). If Equation 19 matched every transition in the log files, then all the points in Figure 12 would fall on a diagonal line though the origin. There is a good correlation (R^2= .76)[13], at the predicted slope = 1, without any new parameters estimated from the data.

---

[13] Estimated using Equation 7 of Kvålseth (1985).

Information scent is the proximal means by which a forager judges the value or relevance of distal information sources. ACT-IF computes information scent based on spreading activation from proximal cues on the Scatter/Gather screen. Figures 11 and 12 show that the ACT-IF model of information scent tracks quite well the underlying (distal) structure of relevant information in the Scatter/Gather clustering system. Information scent matches the underlying distribution of relevant documents across clusters in the average Scatter/Gather state (Figure 11). Information scent tracks the proportion of relevant information available as one progresses from state to state in Scatter/Gather (Figure 12). The accuracy of the information scent judgements by ACT-IF corroborate the validity of the scent model and the effectiveness of the Scatter/Gather interface at communicating the underlying clusters of documents.

Optimization Analysis of the Scatter/Gather Task

ACT-IF uses a set of heuristics to evaluate production rules that match current conditions. These heuristics posit proximal mechanisms that generate behavior that approximate the adaptationist analysis of the conventional foraging models discussed above. These evaluation heuristics build upon the ACT-IF information scent mechanisms. To motivate the specific ACT-IF heuristics presented in the next section, we first present a more refined patch model and diet model analysis of Scatter/Gather.

This refined analysis is a state space model, which is presented in the Appendix. It is based on an engineering analysis of Scatter/Gather in Pirolli (1998). The time cost parameters used in this state-space model were estimated from data collected from Scatter/Gather experts (Pirolli & Card, 1995). The analysis assumed an optimizing user that interacted as fast as possible, and who chose actions so as to maximize the overall average rate of gain, R.

Characterization of Patch and Diet Problems

We assume that the information patches are the Scatter/Gather display windows that contain lists of document titles. The goal of the forager is to select relevant titles from these displays. The patch (time-allocation) problem facing the information forager is the choice between (a) continuing to cluster and re-cluster documents (between-patch enrichment), or (b) beginning to display titles and forage (within-patch exploitation). The diet problem facing the information foraging is one of selecting the optimal set of clusters on each Scatter/Gather window.

Expected Rate of Gain for Displaying Clusters

During the within-patch display phase, the forager scans through a scrollable list of document titles, and must spend time processing each document citation, plus an additional amount of time processing each relevant document (i.e. cutting and pasting them into their answer file). Scanning a list of document titles should produce a within-patch gain function such as Figure 4. As discussed above, the optimal strategy in this situation is to forage until the end of the list.

The rate of return for displaying the k best clusters in a Scatter/Gather state may be characterized as

$$R_D = \frac{\text{No. relevant documents in gathered clusters}}{\text{time between patches so far} + \text{future time within gathered clusters}},$$

The expected within-patch (i.e., within gathered clusters) time would be the time it takes to process all documents in a cluster plus the estimated additional time it would take to handle the relevant ones in the cluster (the task involves cutting and pasting relevant titles into another window),

future time within gathered clusters
= time on all documents in gathered clusters
+ time on relevant documents in gathered clusters.

The Scatter/Gather interface presents the total number of documents in each cluster. For a particular cluster i, on a particular Scatter/Gather screen s, we designate the total number of documents as $N(i, s)$. The expected time to process all documents in a set of k gathered clusters in a particular state can be characterized as the sum of the all the documents in all the gathered clusters times the estimated time it would take to process each document title, t_N ,

$$t_N \sum_{i=1}^{k} N(i,s).$$

Using the ACT-IF formulation of information scent in Equation 16, we can characterize the expected time it would take to process all the relevant documents in all the k gathered clusters as,

$$t_g \sum_{i=1}^{k} g(i,s),$$

where t_g is the additional time it takes to process a relevant document title.

The rate of return for displaying k gathered clusters in state s can be estimated by dividing the information scent estimate of the relevant documents in the gathered clusters by the expected time cost, which will be sum of the time spent so far t_B plus the expected time within gathered clusters t_W ,

$$\begin{aligned} R_D(k,s,t_B) &= \frac{\sum_{i=1}^{k} g(i,s)}{t_B + t_W} \\ &= \frac{\sum_{i=1}^{k} g(i,s)}{t_B + \left(t_N \sum_{i=1}^{k} N(i,s) + t_g \sum_{i=1}^{k} g(i,s) \right)}. \end{aligned} \tag{20}$$

In our simulations we fixed our time cost parameters to be

$$t_N = t_g = 1 \text{ second/information item.}$$

Positive deviations from these values have very little effect on the results reported below.

Optimal Diet of Clusters to Display

To calculate the optimal collection of k clusters to display, we can evaluate R_D for different collections of k clusters and choose the collection that has the maximum value of R_D. According to the diet selection algorithm presented above, we may rank the clusters by their decreasing profitabilities and by considering the rates of gain produced by collections of the topmost k = 1, 2, …9 clusters. The rate-maximizing collection of clusters to display, will be,

$$R_D^*(s,t_B) = \max_{k=1,2,\ldots 9} R_D(k,s,t_B) \tag{21}$$

Expected Rate of Gain for Gathering and Re-Scattering Clusters

To evaluate gathering and re-clustering clusters we assume that the information forager projects the rate of gain that would be achieved by performing one more step of clustering and then displaying the titles and foraging. The look-ahead assumes that the gathering and reclustering will take t_Δ additional time, that the number of total documents will change by factor Δg and the number of relevant documents will change by factor ΔN,

$$R_{SG}(k,s,t_B) = \frac{\Delta g\left(\sum_{i=1}^{k} g(i,s)\right)}{(t_B + t_\Delta) + \Delta N\left(t_N \sum_{i=1}^{k} N(i,s)\right) + \Delta g\left(t_g \sum_{i=1}^{k} g(i,s)\right)}. \tag{22}$$

Similar to Equation 21, the rate-maximizing collection of k clusters to gather and re-cluster is found by ranking clusters in decreasing order of profitability, and determining the rates of gain for collections of the topmost k = 1, 2, …9 clusters. The rate-maximizing collection of clusters to gather and re-scatter will be,

$$R^*_{SG}(s,t) = \max_{i=1,2,\ldots 9} R_{SG}(i,s,t). \tag{23}$$

Enrichment

The optimizing forager invests in between-patch Scatter/Gather time in order to enrich within-patch foraging. Each cycle of gathering and re-scattering clusters is aimed at improving the proportion of relevant documents under consideration, and it also reduces the total number of documents under consideration. Figure 13 shows the effect of this enrichment process. Figure 13 plots the proportion of relevant documents in Scatter/Gather states that occur along a sequence of interactions with Scatter/Gather that assumed that the forager always gathered the rate-maximizing set of clusters and re-clustered them. The data for Figure 13 were produced by running the state space model and varying the amount of time invested in cycles of scattering and gathering clusters. At each cycle of scattering and gathering clusters, the model selects the rate-optimizing set of clusters according to Equation 23.

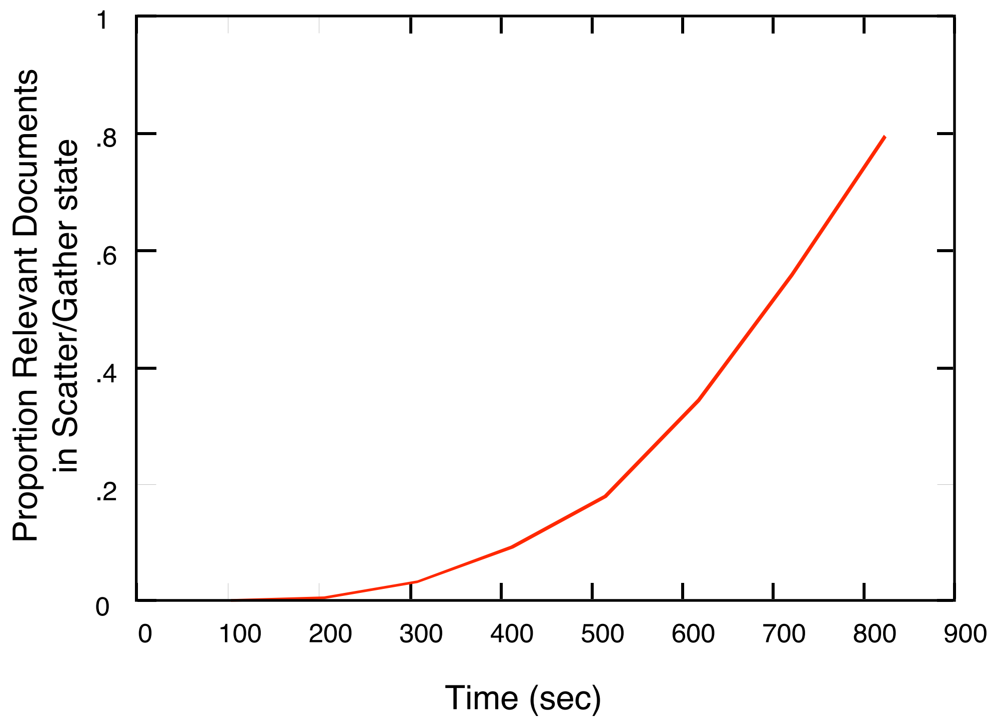

Figure 13. Proportion of all documents that are relevant on Scatter/Gather windows, plotted as a function of time invested in scattering and gathering, assuming hill-climbing heuristics.

Patch Decision Problem

The patch problem for Scatter/Gather involves the decision between the options of (a) continuing to gather and re-scatter clusters vs (b) displaying clusters and foraging through the display. To show the decision rule for this problem we present Figure 14, which displays the changing evaluations of these options over time. Figure 14 was produced by running the state space model, varying the amount of time invested in cycles of scattering and gathering clusters, and always selecting the rate-maximizing collection of clusters. At each cycle of scattering and gathering, the simulation would determine (a) R_D*, the maximum rate of gain that could be achieved by foraging through a display of titles, and (b) R_{SG}*, the maximum rate of gain, that could by achieved by one more round of scattering and gathering clusters.

Figure 14 shows that both these rates rise at first then drop. During the rising portion of the graph, this means that displaying titles for the current state will not be as productive as displaying titles in the next state, after another round of gathering and re-scattering clusters. During this phase, the forager should continue scattering and gathering clusters. During the declining portion of the graph, displaying titles for the current state will be more productive than displaying titles in the next state. As soon as this phase is detected, the forager should stop scattering and gathering clusters and should display titles. Figure 14 suggests a greedy hill-climbing regime: If $R_D^* \geq R_{SG}^*$ then the forager should display titles and if $R_{SG}^* > R_D^*$ the forager should continue scattering and gathering clusters.

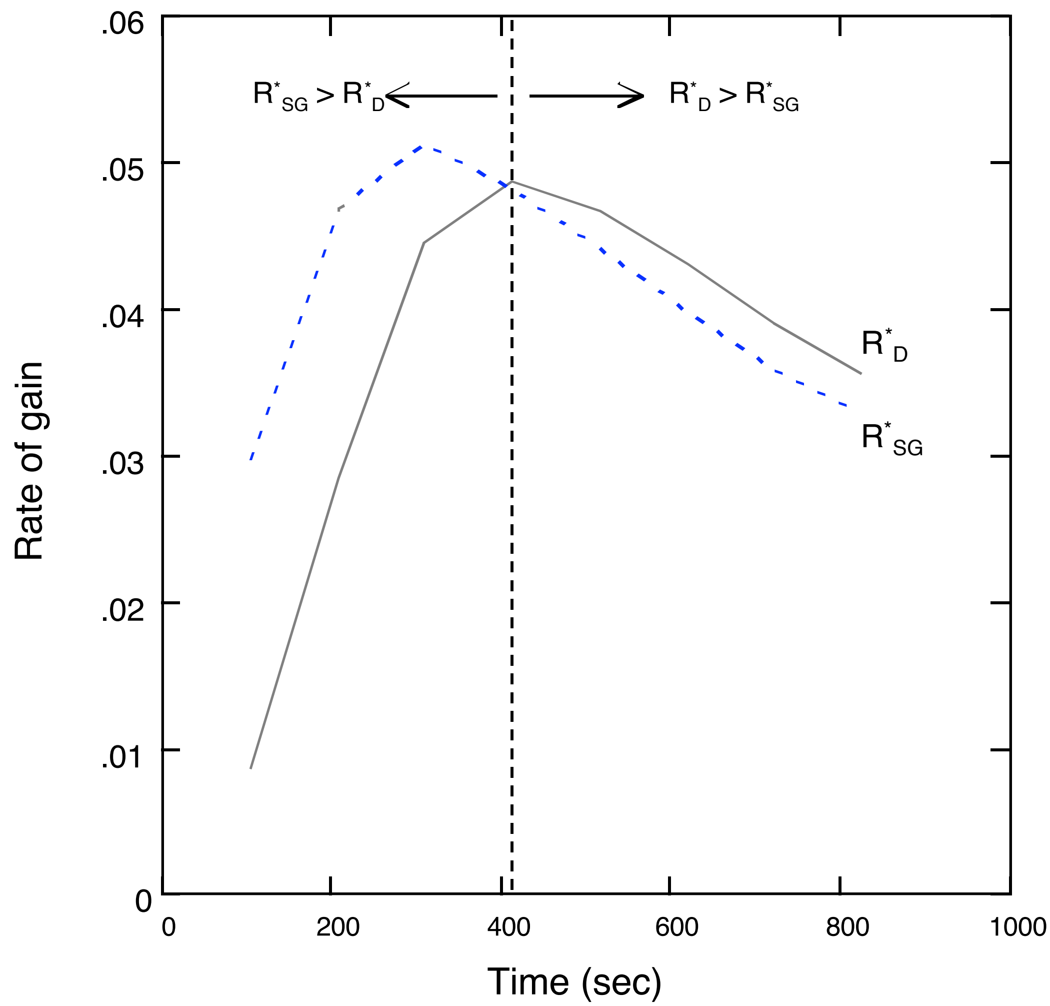

Figure 14. The average rate of gain yielded by different investments in time spent scattering and gathering clusters, assuming hill-climbing heuristics.

Next we discuss how the ACT-IF model evaluates productions in ways that achieve effects similar to those discussed in this analysis of the state-space model.

ACT-IF Evaluation Functions

ACT-IF is a production system that selects a single production rule to execute on each cycle of operation. On each cycle, ACT-IF matches production rule conditions to information in declarative memory, evaluates instantiations of matching rules, and executes the action of the highest-evaluated instantiation. The ACT-IF model for Scatter/Gather has evaluation functions of production rules in Figure 10 that are based the adaptation analysis given in the previous section:

- The DO-DISPLAY-TITLES production is evaluated on the basis of the rate of gain that would be achieved by displaying and foraging through k clusters that have already been gathered. The evaluation function is R_D in Equation 20.
- The DO-SCATTER-GATHER production is evaluated on the projected rate of gain that would be produced by one more round of having the system scatter k clusters that have already been gathered. The evaluation function is R_{SG} in Equation 22.
- The SELECT-RELEVANT-CLUSTER production is evaluated on the basis of an assessment of the profitability $\pi(c, s)$ of the cluster, c, that it matched on the current Scatter/Gather cluster display window (state s). The ACT-IF model for the Scatter/Gather task assumes that the profitability is evaluated by:

$$\pi(c,s) = \frac{g(c,s)}{t_g g(c,s) + t_N N(c,s)} \tag{24}$$

  where $g(c, s)$ is the activation-based scent assessment in Equation 16. $N(c, s)$ is the total number of documents in a cluster. The time cost to process all the documents in a cluster is therefore the term $t_N N(c, s)$, and the additional costs of processing relevant documents is the term $t_g \; g(c, s)$.
- The DESELECT-IRRELEVANT CLUSTER, matches an already selected cluster, but is evaluated on the basis of the maximum of the current rate of gains estimated by R_{SG} and R_D.

These evaluations are computed locally, based only on declarative information matched by a production rule plus a time parameter. In concert, however, the evaluations instantiate important aspects of the adaptation analysis provided by the state-space model:

- The rule evaluations select clusters when their profitability, $\pi(c, s)$, is greater than the current rate of gain for k clusters that have already been gathered so far. Clusters will continue to be gathered so long as the evaluation of SELECT-RELEVANT-CLUSTER is greater than the evaluation of DO-SCATTER-GATHER or DO-DISPLAY-TITLES. This solves the diet selection problem and implements the rate-maximization summarized in Equations 21 and 23.

- Clusters are deselected when their profitability is less than the expected rate of gain for already gathered clusters.
- The evaluations work to continue gathering clusters and scattering them until the overall rate of gain shows a projected decrease, i.e., until $R^*_{SG} < R^*_D$. So long as DO-SCATTER-GATHER has a higher evaluation than DO-DISPLAY-TITLES, the ACT-IF model will continue to gather and scatter clusters. When the evaluation of these two rules reverses, the model will display titles.

EXPERIMENT: FORAGING WITH SCATTER/GATHER

An evaluation study of Scatter/Gather performed by Pirolli, Schank, Hearst, and Diehl (1996) provides data (mostly unreported) to test our information foraging predictions. In this study, two groups of participants used Scatter/Gather under slightly different task instructions. Here we presented additional analyses that focus on information foraging analyses and fits of the ACT-IF model.

Method

Participants

Eight adults solicited through Xerox PARC or the Stanford University graduate program participated in the Scatter/Gather portion of the study as volunteers or were paid $10/hour.

Materials and Procedure

Participants were asked to read the instructions for the experiment and then use the Scatter/Gather system to find articles relevant to given topics in a large and complex collection of text documents. The experiment used the 2.2 gigabyte TIPSTER text collection created for the TREC conference (Harman, 1993). Twelve topics were drawn from the first 100 query topics used in the TREC conference. A typical topic description is:

> A relevant document will include a prediction about the prime lending rate (national-level or major banks'), or will report a prime rate move by major banks, in response to or in anticipation of a federal/national-level action, such as a cut in the discount rate.

As mentioned above, information retrieval experts associated with TREC have identified relevant documents in the Tipster collections for each of the topics (Harman, 1993). The twelve topics chosen for this experiment were selected at three levels of difficulty, where difficulty was measured by the mean number of expert-identified relevant documents in the Tipster collection. The four topics with the fewest (expert-identified) relevant documents (M = 46) were placed in the Hard group, the four topics with the most relevant documents (M = 865) were placed in the Easy group, and the four topics about the median number of relevant documents were placed in the Medium group (M = 303 relevant documents).

Four blocks of topics were constructed. Each topic-block contained one easy topic, one medium topic, and one hard topic, in that order. Each participant completed two blocks of topics using Scatter/Gather, and the other two blocks were used for other activities not reported here. The presentation order of blocks was counterbalanced over participants, within groups, according to a randomized Latin square.

Scatter/Gather users read a topic query and then proceeded to find as many documents relevant to the query as possible. This required that the participants repeatedly scatter and gather clusters using windows such as Figure 8, then choose clusters for display. The display window would present a list of document titles from the chosen clusters and the participants would select relevant titles from the list. The titles selected by the participant would then be saved to a file, as the participants' answer to the topic query.

Scatter/Gather participants were randomly assigned to one of two study conditions: Scatter/Gather Speeded (N = 4) or Scatter/Gather with Relevance Ratings (N = 4). In the Scatter/Gather Speeded condition, participants were given one hour per block to find articles. In the Scatter/Gather Relevance Rating condition, participants were not given a time limit, and were asked to complete additional classification and relevance activities: Given worksheets, they were asked to indicate how they would classify each presented cluster (i.e., using words or short phrases), and to estimate what percentage of texts in a cluster seemed relevant to the topic. For most of the analyses below we combine the data for the Scatter/Gather groups.

The activities of participants interacting with Scatter/Gather were automatically logged. These log files provide the test data for our ACT-IF simulation. The log files contained time-stamped records of every display presented to the participants and every Scatter/Gather action performed by participants.

Results

A General Analysis of Diet Selection for Scatter/Gather

As a general test of the information diet model, we examined data concerning the selection of clusters for the first step of the interactive Scatter/Gather process. In particular we will test the diet model prediction concerning which clusters are selected and how many. This corresponds to the selection of clusters at Level 1 of document clustering in Figure 7. Later, we present the more detailed ACT-IF model for selecting clusters throughout the Scatter/Gather task. Participants worked on queries at three levels of difficulty, where difficulty corresponded to the number of expert-identified relevant documents in the collection. The number of expert-identified relevant documents across the Easy, Medium, and Hard query tasks was such that Easy No. relevant > Medium No. relevant > Hard No. relevant. Participants selected more clusters for easier tasks: for Hard queries they selected $M = 1.38$ clusters, for Medium queries they selected $M = 1.63$ clusters, and for Easy queries they selected $M = 2.25$ clusters. This is generally consistent with the diet model, which predicts that a forager should select more profitable clusters over less profitable ones: The clusters in the Easy conditions contained more relevant documents than those in the Medium or Hard conditions. As we will illustrate, a coarse application of the diet model predicts the qualitative ordering of the query conditions with respect to the number of clusters selected (Easy > Medium > Hard), and reasonably approximate quantitative predictions.

Table 1 presents estimates relevant to the application of an information diet model to Scatter/Gather. To apply the diet selection model we estimated the subjective assessments of the cluster profitabilities, π_i, for each cluster i at the first Scatter/Gather iteration in all query conditions. The first Scatter/Gather interaction corresponds to the selection of clusters at the very top level of clustering in Figure 7. To estimate these subjective profitabilities, we first derived g_i, the subjective estimates of the number of relevant documents in each cluster. These were obtained from the $N = 4$ participants (out of the eight participants total) who provided ratings of the percentage of relevant documents in each cluster. Then we estimated the time to process each cluster selected at the first Scatter/Gather iteration, t_{Wi} for the three levels of query difficulty. This was calculated by dividing the total time to complete the query tasks by the number of clusters selected at the first iteration.[14] For clusters selected at the first iteration of

[14] We assume that participants had gained sufficient knowledge of the Scatter/Gather system for performing these estimates based on warm-up tasks done prior to experimental conditions.

Scatter/Gather (at Level 1 of clustering): The Easy condition required 957 sec per selected cluster, Medium queries required 994 sec per selected cluster, and Hard queries, 1261 sec per selected cluster. Table 1 lists the profitability, $\pi_i = g_i/t_{W,i}$ for the top 10 profitability clusters from all three query conditions.

Table 1.
Optimal information diet analysis for Scatter/Gather (data from Pirolli et al., 1996)**.**
The optimal diet includes the four highest profitability clusters.

Task Condition (Rank of cluster within query condition)	Participants' estimate of net relevant documents (g_i)	Handling time in sec (t_{Wi})	Estimated profitability (π_I) in relevant documents per sec.
Easy (1)	13,670	957	14.28
Medium (1)	10,400	994	10.46
Hard (1)	11,890	1,261	9.43
Easy (2)	5,824	957	6.08
Easy (3)	2,526	957	2.64
Medium (2)	2,607	994	2.62
Hard (2)	1,991	1,261	1.58
Easy (4)	1,040	957	1.09
Easy (5)	891	357	.93
Hard (3)	379	1,261	.30

Inspection of Table 1 reveals that the general qualitative prediction is that the number of clusters selected in query conditions should be such that, Easy ≥ Medium ≥ Hard, which is consistent with observation. To see this, recall that Equation 12 tells us that the optimal diet can be constructed by selecting the k most profitable clusters. Since the rows of Table 1 are in descending rank profitability, drawing an imaginary line across Table 1 is like choosing the clusters whose profitabilities lie above the imaginary line. One can verify that drawing imaginary lines for all diets of greater than three clusters will select more Easy query clusters than Medium.

We calculated the overall average rate of finding relevant document citations, $R(k)$, in Equation 12, for diets that included the k = 1, 2,...10 highest profitability clusters in Table 1. To do this we used t_{Bi} = 91.9 min, which is the average time between query tasks in the same query condition. Figure 15 presents the profitability estimates from Table 1. The leftmost points are labeled with the query conditions in which the clusters occurred.

Figure 15 also presents the estimates of the overall average rate of finding relevant items $R(k)$ for diets of increasing numbers of clusters, k. The predicted optimal diet of $k^* = 4$ clusters, as calculated by Equation 12, is indicated by the hashed line.

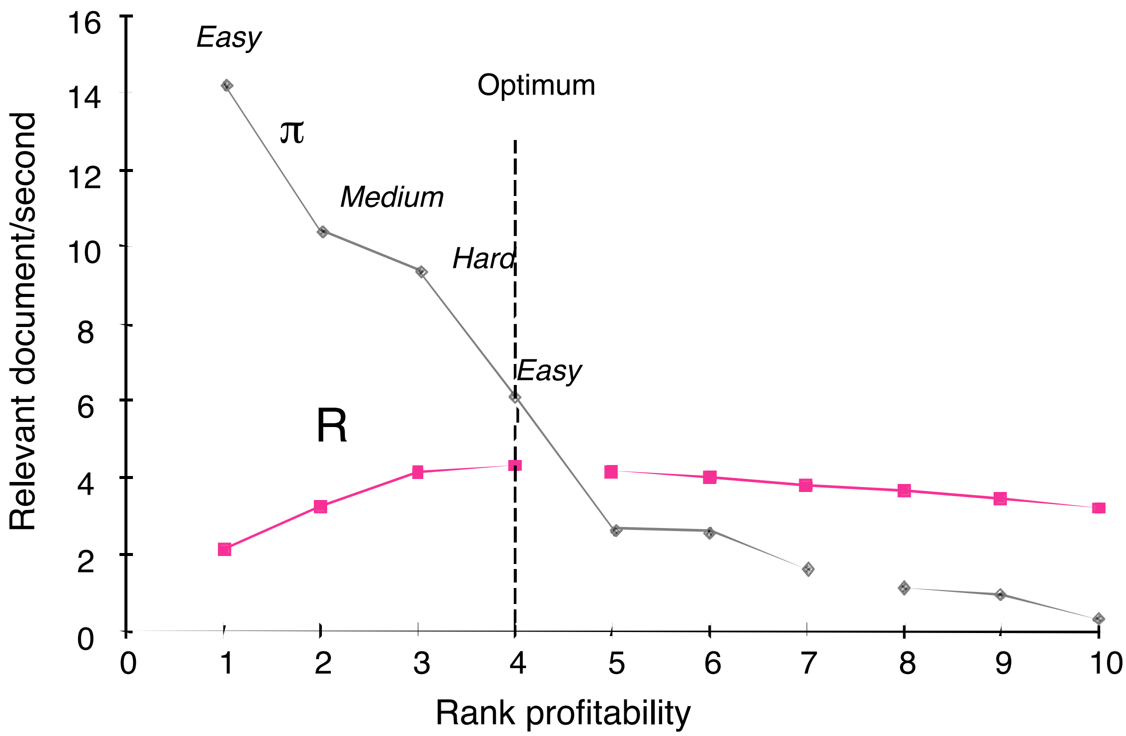

Figure 15. Analysis of the optimal information diet. The profitability (π) of clusters is ranked and added to the diet in order of decreasing profitability until the rate of gain, R, so long as the profitability of the item is greater than R.

According to this application of the diet model, for the Easy tasks, the topmost 2 clusters should be chosen, and for the Medium and Hard tasks the topmost 1 cluster should be chosen. These compare favorably to the observed values: Easy M = 2.25 clusters, Medium M = 1.63 clusters, and Hard M = 1.38 clusters.

Match to Subjective Ratings

Half (N = 4) of our Scatter/Gather participants provided subjective ratings of the percentage of relevant documents they expected to find in each cluster, i.e., ratings of

(No. relevant documents/No. total documents) x 100%,

in each cluster.

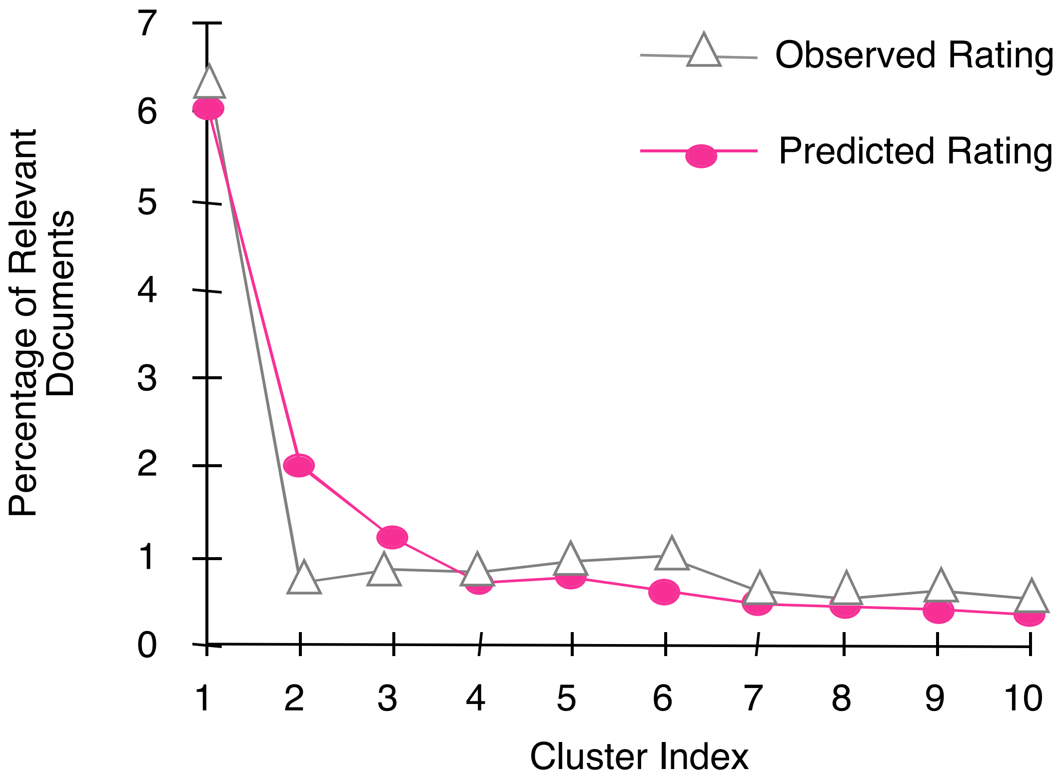

Figure 16. Observed ratings of the percentage documents in each cluster that are relevant and the ratings predicted by activation-based assessment of information scent.

Information scent in ACT-IF, $g(c, s)$, is assumed to provide an assessment of the quantity of relevant information in a cluster. A straightforward hypothesis is that estimates of information scent weighted by the total number of documents in a cluster, N (which is displayed on the Scatter/Gather screen) should map onto these subjective ratings,

$$g(c, s)/N.$$

Our ACT-IF simulation obtains $g(c, s)$ and N for productions that match cluster digests on the Scatter/Gather screen (Figure 8). We extracted these values for each cluster rated by participants. We assumed a simple linear mapping from the simulation estimates onto observed ratings. We fit a linear regression to the geometric mean of the participants' ratings:

$$Rating = a + b\left(\frac{g(c,s)}{N}\right), \tag{25}$$

where a and b are free parameters. Figure 16 presents the observed and predicted ratings (R^2 = .92; a = .32 and b = 232). Figure 16 illustrates that an information scent analysis, based on an analysis of the proximal cues available in the environment, can provide good predictions of the assessments that users will make of the prospects of finding relevant information

Match of ACT-IF to Traces

The ACT-IF model was matched to the log files from all of the Scatter/Gather participants. On each cycle of each simulation of each participant, the ACT-IF production system ranked productions whose conditions were matched (the conflict set of productions). We investigated how the observed cluster-selecting actions of users compared to the predicted rankings of those actions in the ACT-IF simulation. Ideally, the observed action of the user should have been the highest ranked production in the conflict set of productions on the corresponding simulation cycle.

The histogram in Figure 17 shows the frequency with which productions at different ranks matched the observed actions. The histogram can be interpreted as reflecting the probability that the actions at a particular rank match the observed actions. Higher ranked productions show better chances of matching user actions. There are a total of N = 858 observations in Figure 17. The χ^2 statistic for comparing the distribution of predicted action selection against random selection was $\chi^2(10) = 400.77$, $p < .0001$.

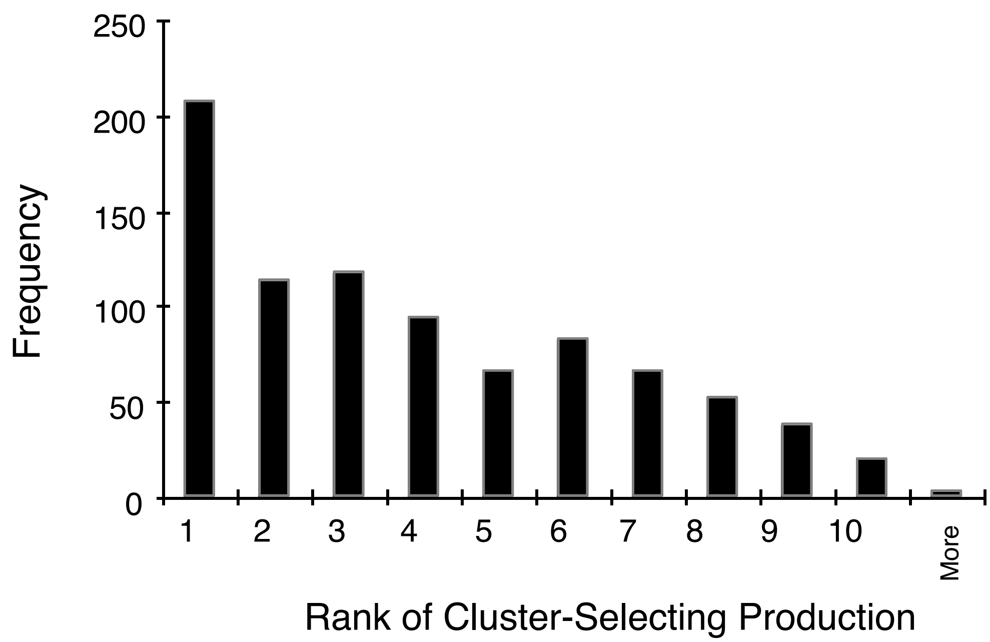

Figure 17. Frequency that ACT-IF productions match observed cluster-selection actions. The ACT-IF production rankings are computed at each simulation cycle over all productions that match.

A strong test of Information Foraging Theory concerns the selection of Scatter/Gather clusters. Clusters at state s should be selected so long as their profitability, $\pi(c, s)$, is greater than the overall rate of gain for the clusters gathered at that state $R_D(k, s, t)$. If we let,

$$\begin{aligned} x &= \textit{Cluster Profitability} - \textit{Expected Rate of Gain} \\ &= \pi(c,s) - R_D(k,s,t), \end{aligned} \tag{26}$$

then decisions should be (a) select a cluster when $x > 0$ and (b) do not select a cluster when $x < 0$. The threshold, $x = 0$, separating the decision to select vs not select clusters occurs when profitability equals rate of gain.

We used the ACT-IF model-tracing simulation to collect the statistics relevant to these predictions regarding cluster selection. For all clusters seen by Scatter/Gather users, we determined (a) if the clusters was selected or not by the user, and (b) the value x = (*Cluster Profitability* - *Expected Rate of Gain*) as predicted by the ACT-IF simulation. From these observations, We estimated the probability (density) of selecting a cluster, *select*(x), and the probability (density) of not selecting a cluster, *unselect*(x). We should expect that *select*(x) > *unselect*(x) for positive values of x (cluster profitabilities greater than expected rate of gain), and *unselect*(x) > *select*(x) for negative values of x (cluster profitabilities less than expected rate of gain).

Figure 18 presents these probability density functions. These are based on $N = 2929$ observations. These densities are plotted against $x = \pi(c, s) - R_D(k, s, t)$. Figure 18 shows that most of the profitabilities for all clusters are close to the value of $R_D(k, s, t)$ computed at the time the clusters are presented to users . Despite this, it is clear that there are two distributions whose modes occur on opposite sides of $x = 0$. As predicted, it appears that the threshold, $x = 0$, separates the decision to select vs not select clusters. It seems to occur precisely when the cluster profitability equals the expected rate of gain. Figure 19 gives a clearer indication of the placement of the threshold. In this figure, we have plotted *select*(x) - *unselect*(x), and again: $x = \pi(c, s) - R_D(k, s, t)$. The shift in probability of selecting vs not selecting clusters across the threshold, $x = 0$, is significant, $\chi^2(1) = 50.65$, $p < .0001$.

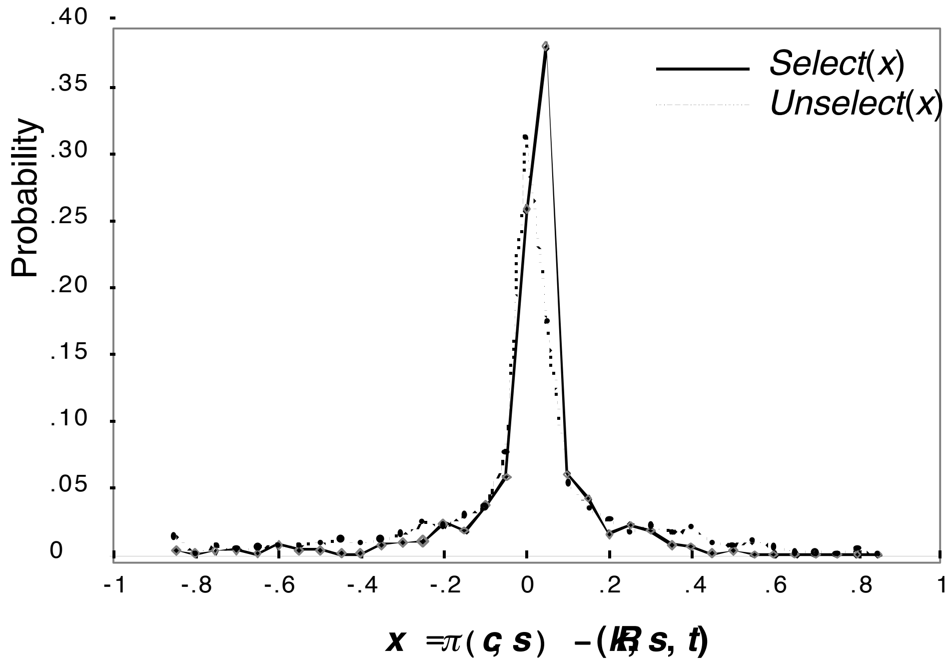

Figure 18. The probability density distributions for selecting clusters, select(x), and not selecting clusters, unselect (x), as a function of the difference between cluster profitability and current estimate of rate of gain: x = _(c, s) - R(k, s, t).

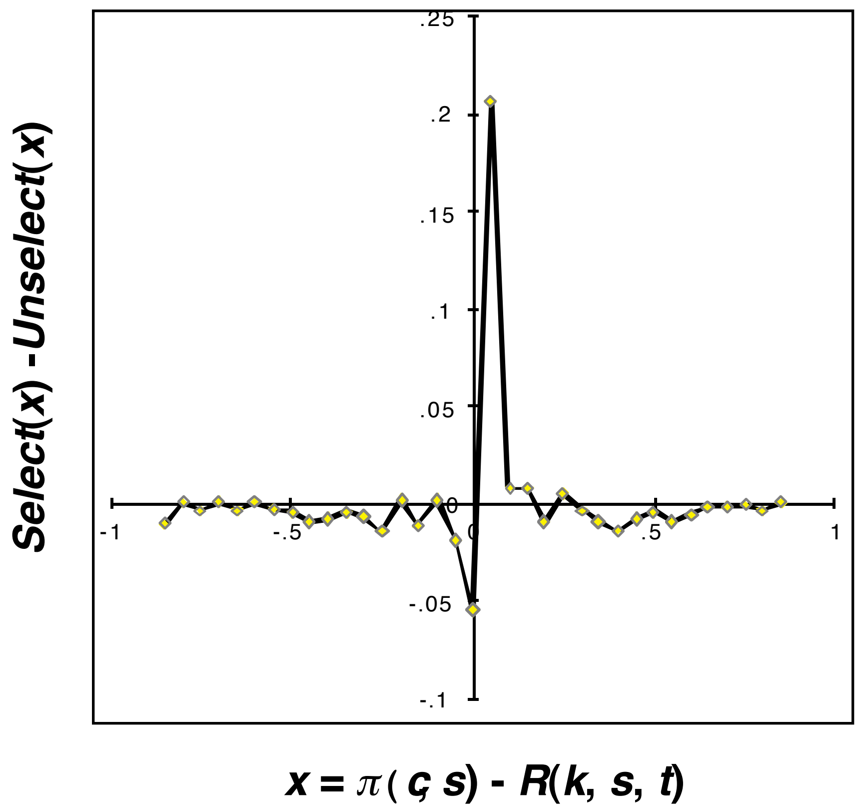

Figure 19. The difference in density distributions from Figure 18.

Discussion

Scatter/Gather is a complex information foraging environment. A cognitive model was developed in ACT-IF by using production rules to implement a task analysis of Scatter/Gather interaction and using spreading activation to compute judgements of information scent. The the spreading activation model was determined by the statistics of word frequency and word cooccurance in the document corpus. The assessment of information scent from spreading activation between external cues and a goal was modeled by a form of interactive cue combination found in exemplar-based models of categorization. This required a single scaling parameter estimated from an analysis of the concordance of proximal cues on the Scatter/Gather screen to the underlying distribution of relevant documents. Heuristics for selecting productions in ACT-IF were developed from an adaptation (rational) analysis of the Scatter/Gather task, by instantiating the information diet and information patch models. In other, words, the ACT-IF model was determined by *a priori* analysis of the information foraging task and the information environment.

This ACT-IF model yielded good fits to users' ratings of the prevalence of relevant documents in given clusters. The likelihood of cluster selection by users correlated with the ACT-IF rankings of clusters. These correspondences support the basic spreading activation model of information scent.

A general analysis showed that the information diet model could explain the differences in the number of clusters selected for queries of different difficulties. ACT-IF also contains heuristics that implement the information diet model and information patch model. These heuristics determine which clusters will be selected and which will not be selected. The threshold determining the choice of clusters varies with Scatter/Gather state and task time. Fits of ACT-IF to traces of Scatter/Gather users suggest that this varying threshold has a good correspondence to the varying threshold of Scatter/Gather users.

GENERAL DISCUSSION

In this paper, we have cast the problems of finding information in terms of coadaptation of people and their information environments. We have proposed that adaptive pressures work on users of information that are analogous to ecological pressures on animal food foraging or mate selection. In all cases there is an advantage to adopting methods that confer competitive advantages measured in benefits per unit cost. We have therefore explored the use of quantitative models that have been developed to explain food-foraging strategies for analyzing adaptive pressures in human information-gathering activities.

Information Foraging under Alternate Assumptions

Up to this point, the information foraging models have assumed a certain set of constraints. Relaxing these constraints or imposing new stronger constraints gives rise to a family of related models. We now proceed to explore some of these models.

Search Motility and Time Allocation

One difference between the two field cases was that the MBA students seemed to be much more active in their search than the business intelligence professional. The difference is analogous to a well-known distinction in behavioral ecology between widely-foraging predators, such as sharks, and sit-and-wait foragers, such as web-building spiders (Pianka, 1966). The environment moves past a sit-and-wait forager, much as it does for our business intelligence analyst, but widely-foraging organisms move through the environment, much like the MBA students. In actuality, sit-and-wait and widely-foraging foragers are two ends of a continuum, and many individuals cycle between the two extremes (Bell, 1991). More generally, different foraging situations and strategies will involve different allocations of time to subsidiary tasks. For instance, Bell (1991), presents the proportion of time allocated to search, pursuit, and handling for a number of species.

Rate Maximization: Time Minimization v. Resource Maximization

The maximization of rate of gains may take the form of *time minimization* or *resource maximization*. That is, an individual may maximize the resource accumulated or may minimize the time spent in resource accumulation (Hames, 1992). The difference can become apparent when the average rate of return of the activity is made more efficient. For instance, human fishers may improve their foraging efficiency as they change technologies from spears, to baited lines on small craft, to large-scale fine-sieved drift nets, and one may imagine similar technological effects on information foraging. As discussed by Hames (1992): (a) a time-minimizer does not forage for more resources if efficiency is improved, and alternative activities that are being forfeited have a large impact but (b) a resource maximizer either spends the same amount of time, or more, if foraging efficiency is improved. In either case, the average rate of return is improved.

Information Encounters That Overlap in Time

Equation 5 assumes that between-patch and within-patch foraging activities are mutually exclusive. There are many systems, however, that allow people to perform information foraging tasks in parallel. For instance, some systems will retrieve items and place them on some sort of queue for processing at the users discretion. The analogous situation in optimal foraging theory is that of web-building spiders who can process (eat) one insect while others are "queued up" by the spiders' web. Information filtering systems (Belkin & Croft, 1992), typically adopt such an approach. A model developed by McNair (1983), addresses foragers such as web-building spiders and a variant of it has been applied to information foraging systems that have multi-threaded processing capabilities (Pirolli & Card, 1998).

The Appendix shows how we can elaborate Equation 5 with simple results from queuing theory. Assume that while a user forages in an information patch, that other information patches are retrieved and placed in a queue. Assume that while the user processes one information patch, other overlapping items arrive on the queue as a Poisson process at rate $\hat{\lambda}$. Assuming that $\hat{\lambda} < 1$, then the rate of gain, R, is:

$$R = \frac{\lambda g(t_W)}{1 + (\lambda - \hat{\lambda})t}. \tag{27}$$

Note that when $\hat{\lambda} = 0$, Equation 27 (multi-threaded with queued items) becomes Equation 5 (single-threaded with zero queued items).

Information Foraging with Deadlines and Uncertainty

Another strong assumption made in the conventional models was that choices made by an information forager yielded certain results. If the MBA students chose to read newspaper articles or company reports or whole-industry reports, then the reading of these sources of information would yield information as predicted by the average yield of these types. Now we take into account two more complications of this simple account. First, the outcome of a particular choice is uncertain. Selecting a particular information source entails the risk that it will produce little or no information and that the time will be wasted. Second, if there is a deadline involved, then there is a chance that sufficient information will not be foraged in time. That is, we move from a pure consideration of maximizing the rate of information gain per unit time to the related problem of maximizing the probability that the information gatherer will have obtained sufficient information by the time of a deadline. This aspect of information-intensive work is familiar to all involved in it and it is a part of both of our examined field examples.

The analysis of strategy choice under risk and uncertainty is a large topic unto itself in disciplines such as economics (e.g., Katz & Rosen, 1994, Ch. 6) as well as psychology (Slovic, Lichtenstein, & Fischhoff, 1988). Some analyses have been developed in optimal foraging theory. One model for risk and deadline-pressure foraging is the *Extreme Variance Rule* developed by Stephens and Charnov (1982). Dynamic optimization models of foraging (Mangel & Clark, 1988) have also addressed foraging with risk and deadlines. A dominant idea in the these analyses of strategy choice, risk, and deadlines is that organisms are generally risk-aversive when they can be, but will become more risk-prone if facing a projected shortfall in some resource. For example, impending starvation often leads to higher risk strategies. The pervasiveness of similar trade-offs in human endeavors (e.g., economic behavior; sports) suggests the we should find them also in information foraging.

Dynamical Models of Information Foraging

Dynamic optimization techniques have been used in optimal foraging theory (Mangel & Clark, 1988), and they have been applied to the analysis of Scatter/Gather to explore the effects of hypothetical design variations on human-computer interaction (Pirolli, 1998). The essential components of such models (Mangel & Clark, 1988) are: (1) a state space, (2) a set of constraints, (3) a strategy set, (4) an optimization criterion, and (5) a specification of state dynamics. Optimization techniques like dynamic programming can be used to understand the optimal foraging strategies in such systems.

Adaptation of Information to People

Not only do people adapt to core complex information environments, but the environments of information to which people adapt are themselves complex and dynamic. Dennett has aptly characterized the situation, using the notion of cultural knowledge units called *memes* (Dawkins, 1976):

> Human language, first spoken and then, very recently, written, is surely the principal medium of cultural transmission, creating the infosphere in which cultural evolution occurs. Speaking and hearing, writing and reading–these are the underlying technologies of transmission and replication most analogous to the technologies of DNA and RNA in the biosphere...We are all well aware that today we live awash in a sea of paper-borne

> memes, breathing in an atmosphere of electronically-borne memes. Memes now spread around the world at the speed of light, and replicate at rates that make even fruit flies and yeast cells look glacial in comparison. They leap promiscuously from vehicle to vehicle, and from medium to medium, and are proving to be virtually unquarantinable. (Dennett, 1995), p. 347.

This analysis suggests that, in addition to humans consuming information for their survival, there is a complementary notion of information consuming humans for *its* survival. Analyses presented in Pitkow and Pirolli (1997) shows by survival analysis techniques that WWW page survival rates (until deletion) are correlated with the amount of visits they receive and the kind of visitors.

Related Theories

There are several other approaches to cognitive psychology that are strongly motivated by ecological and evolutionary concerns. Information Foraging Theory is different from and, for the most part, complementary to these other approaches. In contrast to other approaches, Information Foraging Theory emphasizes universals governing adaptation at the level of the modern-day task environments.

Evolutionary psychology (Barkow, Cosmides, & Tooby, 1992) argues for cognitive universals in mental architecture that have evolved as adaptations over evolutionary time, with an emphasis on specific complexes for such activities as social exchange (Cosmides & Tooby, 1992) and certainly language (Pinker & Bloom, 1992). The motivating analogy is the eye, an exquisite, function-specific, complex arrangement of tissue that has evolved due to strong selection pressures (Tooby & Cosmides, 1992). Evolutionary psychology seeks to explain complex cognitive modules, rather than unprogrammed universal computational architecture, typically by looking back at the physical and social environments that shaped human evolution during the Pleistocene (Tooby & Cosmides, 1992).

Rational analysis (Anderson, 1990) does not necessarily deny the existence of function-specific complexes in the cognitive architecture, but instead focuses on general mechanisms that solve general information-processing problems posed by the environment. For instance, the mechanisms of memory (Anderson & Milson, 1989) are characterized as adaptive solutions to the recurrent structure of events in the world (Anderson & Schooler, 1991), and mechanisms of categorization (Anderson, 1991) are characterized as adaptive solutions to recognizing naturally occurring phenotypes in the biological world. It is argued that these information-processing problems are universal (at least at the space-time scale of terrestrial evolution) and would operate as strong selection forces to drive the evolution of general information-processing solutions at the level of cognitive architecture.

To understand information foraging behavior, to analyze and design information access and visualization technologies, and to understand the dynamics of the environment of information sources, Information Foraging Theory adopts an evolutionary ecology approach. That is, it pursues explanations that take environmental structure and variation as an essential element in the explanation of the observed behavioral structure and variation. Unlike evolutionary psychology, however, Information Foraging Theory focuses on understanding adaptation to current environments. As we stated in the introduction, modern-day information foraging mechanisms may be exaptations of food foraging mechanisms that evolved in our ancestors, but it would be difficult to trace this connection.

The basic working heuristic of Information Foraging Theory is to assume that people adapt to the constraints and problems they face in complex, dynamic, often technology-based environments in which they perform tasks that require processing external information-bearing resources. Unlike rational analysis at the level of the cognitive architecture, Information Foraging Theory frames the rational analysis of adaptations at the level of tasks rather than at the level of cognitive architecture. Information Foraging Theory frames the analysis of complex ensembles of cognitive mechanisms and knowledge that are shaped by information foraging environments. The problems and constraints of such environments can be thought of as forming abstract landscapes of information value and costs, such as the costs of accessing, rendering, and interpreting information-bearing documents. The task of Information Foraging Theory, then, is to explain and predict adaptive solutions: how people will best shape themselves to the environment and how the environment can best be shaped to people.

REFERENCES

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

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409-429.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Boyle, C. F., Corbett, A., & Lewis, M. W. (1990). Cognitive modelling and intelligent tutoring. *Artificial Intelligence, 42*, 7-49.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703-719.

Anderson, J. R., & Pirolli, P. L. (1984). Spread of activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 791-798.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396-408.

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind*. Oxford: Oxford University Press.

Belkin, N. J., & Croft, W. B. (1992). Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM, 35*, 29-38.

Bell, W. J. (1991). *Searching behavior: The behavioral ecology of finding resources.* London: Chapman and Hall.

Bharat, K., & Broder, A. (1998). Estimating the size and overlap of public Web search engines. *Proceedings of the Seventh International World Wide Web Conference* .

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Case, D. O. (1991). The collection and use of information by some American historians: A study of motives and methods. *Library Quarterly, 61*, 61-82.

Cashdan, E. (1992). Spatial organization and habitat use. In E. A. Smith & B. Winterhalder (Eds.), *Evolutionary ecology and human behavior* (pp. 237-266). New York: de Gruyer.

Charnov, E. L. (1976). Optimal foraging: The marginal value theorem. *Theoretical Population Biology, 9*, 129-136.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163-228). Oxford: Oxford University Press.

Cox, D. R., & Smith, W. L. (1961). *Queues*. London: Wiley.

Cutting, D. R., Karger, D. R., & Pedersen, J. O. (1993). Constant interaction-time Scatter/Gather browsing of very large document collections. *Proceedings of the SIGIR '93* New York: .

Cutting, D. R., Karger, D. R., Pedersen, J. O., & Tukey, J. W. (1992). Scatter/gather: A cluster-based approach to browsing large document collections. *Proceedings of the SIGIR '92* (pp. 318-329). New York: ACM.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little, Brown and Co.

Dennett, D. C. (1995). *Darwin's dangerous idea*. New York: Simon and Schuster.

Denning, P. J. (1980). Working sets past and present. *IEEE Transactions on Software Engineering, 6*, 66-84.

Furnas, G. W. (1997). Effective view navigation. *Proceedings of the Human Factors in Computing Systems, CHI '97* (pp. 367-374). Atlanta, GA: Association for Computing Machinery.

Hames, R. (1992). Time allocation. In E. A. Smith & B. Winterhalder (Eds.), *Evolutionary ecology and human behavior* (pp. 203-235). New York: de Gruyter.

Harman, D. (1993). Overview of the first text retrieval conference. *Proceedings of the 16th Annuam International ACM/SIGIR Conference* (pp. 36-38). Pittsburgh, PA: ACM.

Holling, C. S. (1959). Some characteristics of simple types of predation and parasitism. *Canadian Entomology, 91*, 385-3998.

Kaplan, H., & Hill, K. (1992). The evolutionary ecology of food acquisition. In E. A. Smith & B. Winterhalder (Eds.), *Evolutionary ecology and human behavior* (pp. 167-201). New York: de Gruyter.

Katz, M. L., & Rosen, H. S. (1994). *Microeconomics (2nd Ed.)*. Boston, MA: Irwin.

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

Malone, T. (1983). How do people organize their desks? Implications for the design of office systems. *ACM Transactions on Office Systems, 1*, 25-32.

Mangel, M., & Clark, C. W. (1988). *Dynamic modeling in behavioral ecology*. Princeton, NJ: Princeton University Press.

McNair, J. N. (1983). A class of patch-use strategies. *American Zoologist, 23*, 303-313.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review, 85*, 207-238.

Pianka, E. R. (1966). Convexity, desert lizards, and spatial heterogeneity. *Ecology, 47*, 1055-1059.

Pinker, S., & Bloom, P. (1992). Natural language and natural selection. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 451-493). Oxford: Oxford University Press.

Pirolli, P. (1997). Computational models of information scent-following in a very large browsable text collection. *Proceedings of the Conference on Human Factors in Computing Systems, CHI '97* (pp. 3-10). Atlanta, GA: Association for Computing Machinery.

Pirolli, P. (1998). Exploring browser design trade-offs using a dynamical model of optimal information foraging. *Proceedings of the Conference on Human Factors in Computing Systems, CHI '98* (pp. 33-40). Los Angeles: ACM.

Pirolli, P., & Card, S. (1997). *The evolutionary ecology of information foraging* (Technical Report UIR-R97-01). Palo Alto, CA: Xerox PARC.

Pirolli, P., & Card, S. K. (1995). Information foraging in information access environments. *Proceedings of the CHI '95, ACM Conference on Human Factors in Software* (pp. 51–58). New York: ACM.

Pirolli, P., & Card, S. K. (1998). Information foraging models of browsers for very large document spaces. *Proceedings of the Advanced Visual Interfaces Workshop, AVI '98* (pp. 83-93). Aquila, Italy: ACM.

Pirolli, P., Schank, P., Hearst, M., & Diehl, C. (1996). Scatter/Gather browsing communicates the topic structure of a very large text collection. *Proceedings of the Conference on Human Factors in Computing Systems, CHI '96* Vancouver, BC: Association for Computing Machinery.

Pirolli, P., & van Gigch, J. P. (1995). *Information foraging and sensemaking in strategic management: An analysis of an in-situ study* (Technical Report). Palo Alto, CA: Xerox PARC.

Pitkow, J., & Pirolli, P. (1997). Life, death, and lawfulness on the electronic frontier. *Proceedings of the Conference on Human Factors in Computing Systems, CHI '97* (pp. 383-390). Atlanta, GA: Association for Computing MAchinery.

Price, D. J. d. S. (1963). *Little science; big science*. New York: Columbia University Press.

Reitman, W. R. (1964). Heuristic decision procedures, open constraints, and the structure of ill-defined problems. In M. W. Shelly & G. L. Bryan (Eds.), *Human judgements and optimality.* . New York: Wiley & Sons.

Reitman, W. R. (1965). *Cognition and thought: An information-processing approach*. New York: Wiley.

Repo, A. J. (1986). The value of information: Approaches in economics, accounting, and management science. *Journal of the American Society for Information Science, 40*, 68-85.

Resnikoff, H. L. (1989). *The illusion of reality*. New York: Springer-Verlag.

Sandstrom, P. E. (1994). An optimal foraging approach to information seeking and use. *Library Quarterly, 64*, 414-449.

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science, 26*, 321-343.

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Towards a dynamic situational definition. *Information Processing and Management, 26*, 755-776.

Schuetze, H. (1992). Dimensions of meaning. *Proceedings of the Supercomputing '92* (pp. 787-796). Minneapolis, MN: .

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99-118.

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence, 4*, 181-204.

Slovic, P., Lichtenstein, S., & Fischhoff, B. (1988). Decision theory. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Steven's* *handbook of experimental psychology (Second Edition; Vol. 2)* (pp. 673-738). New York: Wiley.

Smith, E. A. (1981). The application of optimal foraging theory to the analysis of hunter-gatherer group size. In B. Winterhalder & E. A. Smith (Eds.), *Hunter-gatherer foraging strategies* (pp. 36-65). Chicago: University of Chicago.

Smith, E. A. (1987). Optimization theory in anthropology: Applications and critiques. In J. Dupré (Ed.), *The latest on the best* (pp. 201-249). Cambridge, MA: MIT Press.

Smith, E. A., & Winterhalder, B. (Eds.). (1992). *Evolutionary ecology and human behavior*. New York: de Gruyter.

Soper, M. E. (1976). Characteristics and use of personal collections. *Library Quarterly, 46*, 397-415.

Stephens, D. W. (1990). Foraging theory: Up, down, and sideways. *Studies in avian biology, 13*, 444-454.

Stephens, D. W., & Charnov, E. L. (1982). Optimal foraging: Some simple stochastic models. *Behavioral Ecology and Sociolobiology, 10*, 251-263.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Stigler, G. J. (1961). The economics of information. *The Journal of Political Economy, 69*, 213-225.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 19-136). Oxford: Oxford University Press.

vanRijsbergen, C. J. (1979). *Information retrieval*. (2nd ed.). Boston, MA: Butterworth & Co.

Williams, G. C. (1992). *Natural selection: Domain, levels, and challenges*. New York: Oxford University Press.

Winterhalder, B., & Smith, E. A. (1992). Evolutionary ecology and the social sciences. In E. A. Smith & B. Winterhalder (Eds.), *Evolutionary ecology and human behavior* (pp. 3-23). New York: de Gruyer.

AUTHOR NOTES

This research has been supported by an Office of Naval Research Grant, No. N00014-96-0097. Correspondence should be directed to Peter Pirolli (pirolli@parc.xerox.com), Xerox PARC, 3333 Coyote Hill Road, Palo Alto, CA 94304. We would like to thank John R. Anderson and Judith Olson for their extensive constructive comments on earlier drafts.

APPENDIX

Information Patch Residence Time Model

For the patch model, Holling's disc equation (Equation 5) is instantiated (Stephens & Krebs, 1986) as Equation 7. The task is to determine the optimal vector of collection residence times ($t_{W1}, t_{W2}, ..., t_{WP}$) for a set of patches, $\rho = \{1, 2, ..., i, ...P\}$ that maximizes the rate of gain R. To differentiate R in Equation 7 with respect to an arbitrary t_{Wi}, we first get

$$R = \frac{\lambda_i g_i(t_{Wi}) + k_i}{c_i + \lambda_i t_{Wi}}, \tag{A.1}$$

where k_i is the sum of all terms in the numerator of Equation 7 not involving t_{Wi},

$$k_i = \sum_{j \in \rho - \{i\}} g_j(t_{Wj}),$$

and c_i is the sum of all terms in the denominator of Equation 7 not involving t_{Wi},

$$c_i = 1 + \sum_{j \in \rho - \{i\}} \lambda_j t_{Wj}.$$

So, for a given t_{Wi} we get

$$\partial R / \partial t_{Wi} = \frac{\lambda_i g'_i(t_{Wi})\left[\lambda_i t_{Wi} + c_i\right] - \lambda_i\left[\lambda_i g_i(t_{Wi}) + k_i\right]}{(\lambda_i t_{Wi} + c_i)^2}. \tag{A.2}$$

R is maximized when $\partial R / \partial t_{Wi} = 0$ (Charnov, 1976), and so

$$g'_i(t_{Wi})\left[\lambda_i t_{Wi} + c_i\right] - \left[\lambda_i g_i(t_{Wi}) + k_i\right] = 0, \tag{A.3}$$

which becomes

$$g'_i(t_{Wi}) = \frac{\lambda_i g_i(t_{Wi}) + k_i}{\lambda_i t_{Wi} + c_i}, \tag{A.4}$$

so the right-hand side of Equation A.1 (average rate of gain) is the same as the right-hand side of Equation A.4 (instantaneous rate of gain when the average rate of gain is maximized),

$$g_i'(t_{Wi}) = R. \tag{A.5}$$

If we replace R with a function $R(t_{W1}, t_{W2},...t_{Wi}, ..., t_P)$, the full vector of rate maximizing t_{Wi}'s, $(\hat{t}_{W1}, \hat{t}_{W2}, \ldots, \hat{t}_{Wp})$, must fulfill the condition specified by

$$\begin{aligned} g'_1(\hat{t}_{W1}) &= R(\hat{t}_{W1}, \hat{t}_{W2}, \ldots, \hat{t}_{Wp}) \\ g'_2(\hat{t}_{W2}) &= R(\hat{t}_{W1}, \hat{t}_{W2}, \ldots, \hat{t}_{Wp}) \\ &\vdots \\ g'_p(\hat{t}_{WP}) &= R(\hat{t}_{W1}, \hat{t}_{W2}, \ldots, \hat{t}_{Wp}) \end{aligned} \tag{A.6}$$

This is the formal condition (Charnov, 1976) of Charnov's Marginal Value Theorem: long-term rate of gain is maximized by choosing patch residence times so that the marginal value (instantaneous rate) of the gain at the time of leaving each patch equals the long-term average rate across all patches.

Information Diet Model

Following Stephens and Krebs (1986), we assume that the items encountered can be classified into n types. The average rate of gain R can be represented as

$$R = \frac{\sum_{i=1}^{n} p_i \lambda_i g_i}{1 + \sum_{i=1}^{n} p_i \lambda_i t_{Wi}}, \tag{A.7}$$

where for each item type i, λ_i is the encounter rate while searching, t_{Wi} is the expected processing time for each item type, g_i is the expected net currency gain, and p_i is the probability that items of type i should be pursued (the decision variable to be set by the optimization analysis). In the case of food foraging, Equation A.7 might be applied under the assumption that the modeled organism partitions the space of the observed feature combinations exhibited by its potential prey into discrete categories, $i = 1, 2, \ldots n$. One may also think of Equation A.7 as being applicable when an organism can predict (recognize) the net gain, processing time, and encounter rate for an encountered prey. To maximize with respect to any given p_i we differentiate

$$R = \frac{p_i \lambda_i g_i + k_i}{c_i + p_i \lambda_i t_{Wi}}, \tag{A.8}$$

where k_i is the sum of all terms not involving p_i in the numerator of Equation A.7, c_i is the sum of all terms in the denominator not involving p_i, and we assume that the gain, processing time, and encounter rate variables are not dependent on p_i. Differentiating Equation A.8 obtains

$$\frac{\partial R}{\partial p_i} = \frac{\lambda_i g_i c_i - \lambda_i t_{Wi} k_i}{(c_i + p_i \lambda_i t_{Wi})^2}. \tag{A.9}$$

Zero-One Rule. Inspection of Equation A.9 shows that R is maximized by either p_i = 1 or $p_i = 0$ (Stephens & Krebs, 1986). Note that this occurs under the constraint that the time it takes to recognize an item is assumed to be zero. This is known as the *Zero-One Rule* which simply states that the optimal diet will be one in which items of a given profitability level are chosen all-or-none, where profitability, π_i, is defined as

$$\pi_i = \frac{g_i}{t_{Wi}}. \tag{A.10}$$

The decision to set $p_i = 1$ or $p_i = 0$ is reduced to the following rules which determine the numerator of Equation A.9:

set $p_i = 0$ if $g_i/t_{Wi} < k_i/c_i$ (the profitability for i is less than for everything else)
set $p_i = 1$ if $g_i/t_{Wi} > k_i/c_i$ (the profitability for i is greater than for everything else)

For the n item types, there are n such inequalities. This provides the basis for the diet optimization algorithm presented in the main text.

State Space Model of Scatter/Gather

The interaction of users with the Scatter/Gather system can be represented in a state-time space, where a state variable $X^{<k>}$ takes on values $X^{<k>} = x$. The state $X^{<k>}$ is the state at step k of the Scatter/Gather process, and $X^{<0>}$ is the initial state. The state has the following components: $t_B(X^{<k>})$ is time taken so far on a query task, $G(X^{<k>})$ is the number of relevant documents in all clusters in the subcollection at state $X^{<k>}$, and $T(X^{<k>})$ is the total number of documents in all clusters in the subcollection at state $X^{<k>}$. To produce the simulation summarized in Figures 13 and 14 we set the initial state such that,

$$\mathrm{G}(X^{<0>}) = 303 \text{ relevant documents},$$

$$N_T(X^{<0>}) = 742{,}833 \text{ total documents},$$

$$t_B(X^{<0>}) = 0 \text{ seconds}.$$

We let $g(c, X^{<k>})$ be the number of relevant documents in cluster c in state $X^{<k>}$ and $N(c, X^{<k>})$ is the number of total documents in cluster c in state $X^{<k>}$. We use the convention that c is indexed in order of decreasing profitability,

$$\frac{g(c, X^{<k>})}{N(c, X^{<k>})} > \frac{g(c+1, X^{<k>})}{N(c+1, X^{<k>})}.$$

The optimal set of clusters, C^*, is chosen such that it maximizes $R_{SG}{}^*$ as specified in Equations 22 and 23. We assume that the distribution of relevant documents across clusters is specified by Equation 17.

The state space is evolved for some number of steps $s = 1, 2, \ldots k, \ldots K$, such that,

$$t_B(X^{<k>}) = t_B(X^{<k-1>}) + t_\Delta, \tag{A.11}$$

$$N_T(X^{<k>}) = \sum_{c=1}^{C^*} N(c, X^{<k-1>}), \tag{A.12}$$

$$G(X^{<k>}) = \sum_{c=1}^{C^*} g(c, X^{<k-1>}). \tag{A.13}$$

Patch Model with Queuing [15]

A foraging model appropriate for overlapping encounters can be developed by elaborating the patch model with simple results from queuing theory (Cox & Smith, 1961). Assuming the strong but simple case in which items arrive locally as Poisson events, then the average length of a queue of such encounter events is determined by traffic intensity, λt_W, where λ is the rate of encounter with patches and t_W, is the average time to process a patch. From queuing theory, the average length of a queue for our simple case would be

$$\frac{\lambda t_W}{(1-\lambda t_W)}. \tag{A.14}$$

Including the item currently handled, a forager may expect to encounter

$$1+\frac{\hat{\lambda} t_W}{(1-\hat{\lambda} t_W)}=\frac{1}{(1-\hat{\lambda} t_W)}, \tag{A.15}$$

items, where is the rate at which overlapping items arrive during the handling of the current item. Assuming that $\hat{\lambda} < 1$, the revised rate of long-term gain becomes

$$R=\frac{\dfrac{\lambda g(t_W)}{1-\hat{\lambda} t_W}}{1+\dfrac{\lambda t_W}{1-\hat{\lambda} t_W}}=\frac{\lambda g(t_W)}{1+(\lambda-\hat{\lambda} t_W)}, \tag{A.16}$$

which is maximized under the first-order condition

$$\frac{dR}{dt}=0,$$

where

$$g'(t_W)=\frac{g(t_W)}{\dfrac{1}{\lambda-\hat{\lambda}}+t_W}. \tag{A.17}$$

---

15 This section elaborates the overlapping encounters patch model of Stephens and Krebs (1986).

Table A.1
Notation Using in the Foraging Models

Conventional Information Foraging Models	
R	Rate of gain of information value per unit time cost.
G	Total information value gained.
T_B	Total time spent in between-patch foraging.
T_W	Total time spent in within-patch foraging.
g	Average information value gained per item
g_i	Average information value gained per item of type i.
$g(t_W)$	Cumulative value gained in information patches as a function of time t_W
$g_i(t_{Wi})$	Cumulative value gained in information patches of type i as a function of time t_{Wi}
t_B	Average time cost for between-patch foraging
t_W	Average time cost for within-patch foraging
λ	Average rate of encountering information patches
t_{Bi}	Time spent between patches of type i
t_{Wi}	Time spent foraging within patches of type i
λ_i	Average rate of encountering information patches of type i
π_i	Profitability of item type i
$\hat{\lambda}$	Overlapping encounter rate (arrival of new items while an item is being processed)
p_i	Probability of pursuing items of type i (diet decision model)
Information Scent and Spreading Activation Models	
A_i	Total activation of cognitive element i
B_i	Base-level activation of cognitive element i
S_{ji}	Association strength from element j to element i
$g(c, s)$	Activation-based estimate of expected No. relevant documents in cluster c in Scatter/Gather state s
T	Scaling factor (Boltzman temperature)
Scatter/Gather Analyses	
$d_D(c)$	Distribution obtained from computational experiments on the clustering algorithm: the proportion of all relevant documents in a Scatter/Gather state that are allocated to cluster of rank c, when ranked by those proportions.
$d_P(c)$	Distribution obtained from information scent analysis of Scatter/Gather displays: the proportion of all relevant documents in a Scatter/Gather state that are allocated to cluster of rank c, when ranked by those proportions.
$\overline{A}(c)$	Average activation received by clusters of rank c.

$\bar{g}(c)$	Average information scent received by clusters of rank c.
$N(c, s)$	No. of total documents in cluster c in Scatter/Gather state s
$R_D(k, s, t_B)$	Expected rate of gain if k gathered Scatter/Gather clusters in state s at task time t_B are displayed and judged for relevance.
$R_{SG}(k, s, t_B)$	Expected rate of gain if k gathered Scatter/Gather clusters in state s at task time t_B are re-clustered
$\pi(c, s)$	Profitability of cluster c in Scatter/Gather state s
Δt	Time it will take to perform a re-clustering
$\Delta g, \Delta N$	Local rate of change in number of relevant documents and total number of documents if a reclustering is performed.

Scatter/Gather State Space Model

$X^{<k>}$	Scatter/Gather state k.
$G(X^{<k>})$	Number of relevant documents in a Scatter/Gather state
$N_T(X^{<k>})$	Total number of relevant documents in a Scatter/Gather state
$t_B(X^{<k>})$	Time spent getting to a Scatter/Gather state.
$N(c, X^{<k>})$	Total number of documents in cluster c in state $X^{<k>}$.
$g(c, X^{<k>})$	Number of relevant documents in cluster c in state $X^{<k>}$.