Information foraging theory: A framework for intelligence analysis

David J. Bryant
DRDC - Toronto Research Centre

Defence Research and Development Canada

Scientific Report
DRDC-RDDC-2014-R115
November 2014

Information foraging theory: A framework for intelligence analysis

David J. Bryant
DRDC - Toronto Research Centre

Defence Research and Development Canada
Scientific Report
DRDC-RDDC-2014-R115
November 2014

IMPORTANT INFORMATIVE STATEMENTS

This report was completed under the Joint Intelligence Collection and Capability (JICAC) project (JFD 4.1), sponsored by the Chief of Defence Intelligence (CDI).

Abstract

Information Foraging Theory (IFT) is proposed as a framework in which model information search in the military intelligence analysis domain. Information Foraging Theory explains human information search and exploitation as adaptations to the informational structure of the environment and has been used to model peoples' preferences for information types, rules for exploiting discrete information sources, and the use of semantic cues to enhance the search process. A plan for the application of Information Foraging Theory to the military intelligence domain is described, beginning with the process of describing the task environment in which analysts work and moving to the issue of defining key Information Foraging Theory concepts in that environment. The report ends with a discussion of ways application of IFT may benefit military intelligence analysis, such as automated goal analysis and parameter tracking, enhancing information scent cues, and information visualisation techniques.

Significance to defence and security

One of the most important tasks of military intelligence analysis is information search, which currently consumes a great amount of analysts' time. This report reviews Information Foraging Theory to determine its suitability for modelling intelligence analysts' information search processes and assessing their adaptive efficiency. Support for analysts in the form of training and decision support systems may be developed within this theoretical perspective and address the important constraints facing intelligence analysts - information overload and severe time limitations.

Résumé

La théorie du butinage des renseignements (TBR) est proposée comme cadre et modèle de recherche des renseignements dans le domaine de l'analyse du renseignement militaire. La théorie du butinage des renseignements explique la recherche et l'exploitation humaine des renseignements comme des adaptations à la structure informative de l'environnement et a été utilisée pour modéliser les préférences des individus selon les types de renseignements, les règles pour l'exploitation de sources de renseignements discrètes, et l'utilisation d'indices sémantiques pour améliorer le processus de recherche. Un plan pour l'application de la théorie du butinage des renseignements dans le domaine du renseignement militaire est décrit, en commençant par le processus de description de l'environnement de travail dans lequel les analystes évoluent et abordant la question de la définition des concepts clés de la théorie du butinage des renseignements dans cet environnement. Le rapport se termine par une discussion sur les façons dont la TBR peut bénéficié l'analyse du renseignement militaire, comme l'analyse automatisée des objectifs et le suivi des paramètres, l'amélioration des renseignements des signaux olfactifs et les techniques de visualisation des renseignements.

Importance pour la défense et la sécurité

L'une des tâches les plus importantes de l'analyse du renseignement militaire est la recherche des renseignements, celle-ci consommant actuellement une grande partie du temps des analystes. Ce rapport passe en revue la théorie du butinage des renseignements pour déterminer la capacité de cette dernière à modéliser les processus de recherche des renseignements des analystes du renseignement et évaluer leur efficacité adaptative. Le soutien aux analystes sous la forme d'instruction et de systèmes de soutien aux décisions peut être développé dans cette perspective théorique et aborder les contraintes importantes auxquelles font face les analystes du renseignement, à savoir la surcharge de renseignements et des délais très serrés.

Table of contents

List of figures

Introduction

Background

Intelligence analysis is a process of gaining knowledge and understanding of an operational area to support informed decision making. In the military context, intelligence analysts strive to "provide commanders and staffs with timely, relevant, accurate, predictive, and tailored information about the enemy and other aspects of the area of operations" [1]. More than this, however, the goal of military intelligence is to gain information superiority by gaining an understanding of the operational environment that is more comprehensive, more accurate, and timelier than that of one's opponent [2]. This makes intelligence a "force multiplier" that allows greater precision in the application of force [2][3].

Although the ultimate aim of intelligence analysis is to make sense of an operational area, much of the effort of analysts is directed at simply gathering data that will be used to build situation awareness (SA) [4]. Analysts make use of a wide range of data sources that provide an equally wide range of information types. These include communications and electronic signals (SIGINT), geospatial intelligence (GEOINT), imagery, meteorological and oceanographic information, human intelligence (HUMINT), open-source intelligence (OSINT), and information provided by other governmental departments [1][5]. Despite the large volume of data that can potentially be surveyed, analysts must select only the relevant subset of data that are useful in building SA. Relevant data may comprise smaller parts of documents or other sources and the extraction of data can be one of the most time-consuming parts of the sensemaking process [4].

Intelligence analysis can be described in terms of two process loops, one comprising information search (searching, filtering, and extracting), and the other comprising sensemaking (iterative generation and testing of hypotheses, determining information needs, etc.) [6]. The two loops continually interact with the sensemaking loop framing information needs for the search loop and the search loop identifying relevant information for use in sensemaking (e.g., [7]). Thus, sensemaking and information search are mutually reinforcing parts of intelligence analysis. Information search is, in large part, a process of seeking evidence that bears on the hypotheses being developed as part of sensemaking. Indeed, evidence that can potentially disconfirm hypotheses is of most value as it speeds the winnowing of competing hypotheses.

Overall, the role of the analyst is to narrow the range of uncertainty and eliminate incorrect, irrelevant, and ambiguous understandings [5]. Individual analysts have different roles in the overall process but overall they play roles related to both information gathering and sensemaking, such as answering questions, providing warnings, monitoring and assessing developments, and providing guidance to data collectors [8]. A key objective of monitoring events is to identify patterns that can be related to underlying factors and intentions [9]. Identifying patterns requires the close interaction of sensemaking and information search functions.

The Canadian Army intelligence process mirrors this interactive search and sensemaking model with its four general steps: direction, collation, processing, and dissemination [10]. Direction involves the communication of information needs to the intelligence staff which subsequently communicates components of those information needs to the primary data collectors. The staff interprets information needs as communicated by higher levels of command to identify the specific kinds of information required (i.e., to operationalize instructions to collectors) and passes along only what is relevant to each collector. In the Collection stage, the collectors obtain information to meet the information requests given to them with little or no aggregation, integration, or interpretation. The Processing step encompasses many activities, such as collation, evaluation, and integration, aimed at selecting and transforming gathered data to produce useful intelligence. The final step of Dissemination sends confirmed intelligence to appropriate users in as timely a fashion as possible.

Issues of intelligence analysis

Military intelligence analysis differs from similar research tasks in other fields (e.g., business) in several important respects [11]. First and foremost, military intelligence analysts consult an unusually wide number of information sources that can contain very large volumes of data [12]. This places a tremendous burden on analysts who must devote extensive time to the search and filtering processes. Search is further complicated by the dynamic and uncertain natures of many sources, which reflects operational environments in which events evolve and change rapidly. Military intelligence analysts must also confront the potential for denial and deception by an opponent which increases the uncertainty associated with data [11].

Military operations generally have a fast tempo and demand rapid analysis, putting analysts under extreme time pressure [12]. Time pressure is especially demanding because military intelligence analysts often must try to find patterns among scattered, seemingly unrelated events. Further complicating factors include the high-risk nature of military operations, the potential pressure on analysts to deliver desired results at the expense of accuracy, and the fragmented or “stove-piped” organization of intelligence staffs [11]. On top of all of this, analysts rarely get any feedback on their performance as the outcomes of events addressed in analysis become known only long after the analysis.

The problem of intelligence analysis has become more complex as surveillance and communication technologies have advanced. Although these technologies have expanded the capabilities of analysts, they have also expanded the capabilities of the militaries of other nations and other non-state actors [2]. Rapid technological advances have further contributed to the prevalence of asymmetric threats, including cyber warfare, and a greater diversity of human contexts in which operations must be performed [1].

Intelligence analysis is a collaborative activity performed by sizeable staffs. This allows analysts to divide labour and share information but only to the extent that there is active collaboration among analysts. Such collaboration has not always been present in the Canadian Armed Forces’ (CAF) intelligence analysis process, with Barber [3] remarking that the “CF’s strategic intelligence capability is characterized by a degree of ‘stove piping’, in which different elements of the intelligence puzzle are collected, processed and disseminated to decision makers by intelligence exploitation centres operating in various degrees of isolation from each other.” When analysts do not interact with one another, there are redundancies and inefficiencies in both the data collection and analysis processes.

In light of all this, perhaps the most significant concerns for intelligence analysts are the vast amount of data that must be searched and the extreme time pressure under which they operate [5][13][14]. Although there is no firm estimate of how much time analysts spend on information search, the prevailing belief is that analysts spend a significant majority of their time on data collection at the expense of processing and sensemaking activities (e.g., [15]). To be of value, collected information must be analysed and integrated. The imbalance of collection and analysis activities indicates that either a great deal of collected information goes unanalysed or that most collected data is analysed only incompletely or inadequately. Information overload, combined with time pressure, dramatically increases the cognitive burden placed on analysts who find it difficult to process large volumes of information rapidly and are thus more susceptible to cognitive biases and error [13][16].

Objective

The aim of this report is to present Information Foraging Theory (IFT) as a potential remedy to the problems of information overload and extreme time pressure faced by intelligence analysts.

Military intelligence analysis is a challenging activity in large part because of the problems of information overload and extreme time pressure. As a result, major improvement to intelligence analysis can potentially be made by improving the efficiency of the information search process. If analysts are better able to locate relevant information more quickly and with less effort, they would not need to devote as much time to information search and collection and could devote more effort to higher-level analysis.

The following section introduces IFT and provides a summary of its major concepts. This introductory section is intended to provide the reader with a general understanding of IFT as it has been formulated and employed in domains that bear some similarity to intelligence analysis. In particular, IFT has been developed in efforts to improve the use of the internet and complex, online information resources. After introducing IFT, the report examines how IFT might be applied to the military intelligence analysis domain and ways IFT could inspire practical ways to mitigate the problems of information overload and extreme time pressure.

Information foraging theory

IFT is a framework developed to explain human information search and exploitation activities in relation to the structure of the environment in which those activities take place. This theory has been developed primarily by Peter Pirolli and his colleagues (e.g., [17][18]) over the past 20 years. They have sought to understand information gathering from an ecological point of view as adaptations of human cognitive mechanisms to the informational structure of the environment. To develop their theory, Pirolli and his colleagues employed rational analysis techniques (Anderson [19][20][21]) to explore and formally describe the nature of the various environments in which people gather information. Rational analysis defines the problems posed by the environment, allows one to evaluate strategies for solving those problems, and suggests how those solutions could be implemented by cognitive mechanisms [22].

IFT is based on an explicit analogy between information gathering by humans and the search for food engaged in by all organisms. Information search can be viewed as "information foraging" in which the searcher is cast as an "informavore" [17] (p. 13) that hungers for information in the way an organism hungers for food. The information foraging problem is analogous to foraging for food in a number of respects and Pirolli [17] (p. 15) argues that the cognitive mechanisms employed to search for information are *exaptations* of the cognitive mechanisms evolved to solve food foraging problems. Exaptation is the principle by which an adaptation that evolved to serve one purpose comes to serve another related purpose [23]. In other words, the similarity of the food and information foraging problems allowed cognitive mechanisms that supported adaptive food foraging to be applied to information foraging. As a result, we can understand how humans look for information in terms of the degree to which information foraging behaviour is adaptive to the particular information environment in which someone forages.

The precise nature of an information environment – the physical media in which information is represented, the locations in which media can be accessed, and the meaning of information with respect to the objectives of the forager – depends on the specific tasks the forager is performing. Nevertheless, all cognitive tasks that require the use of information share a number of critical points of correspondence with the task of foraging for food see [18][22]. First, in both food and information foraging, the desired resources are distributed unevenly throughout the environment. For any species seeking any kind of food (plant, animal, or other), food items occur in irregular patterns throughout the physical landscape. Likewise, information exists as representations in various physical media that are themselves stored in a variety of ways and distributed throughout both a physical and informational space [18][22].

Second, both food and informational resources vary in value (e.g., [15][24]). Food items have values to an organism that can be measured in caloric benefit to the organism and they can be sorted according to those values. Likewise, information items vary in terms of their relevance to an information forager depending on the impact that information will have on the forager's knowledge and goals [18]. The value of information may be more ephemeral than that of food items, changing in value from one forager to another and from one task to another, but from the perspective of someone performing a given task, information items can be ranked in terms of value to accomplishing one's goals just as food items can be ranked in terms of the calories of energy obtained by consuming them.

Third, seeking and processing information requires the expenditure of energy and time in the same fashion as seeking and processing (consuming) food [18][22]. Thus, there is always some cost associated with locating and exploiting a resource, whether food or information. Costs in food foraging are typically physical energy expenditure required to move from one food source to another and to consume and digest food items as well as time. Information foraging requires some physical energy expenditure although that cost has dramatically declined in the age of electronic media and online databases. Information foraging, however, does require cognitive effort and time, just as does food foraging.

Finally, in both food and information foraging, the objective is to the greatest cumulative resource value while expending the least energy/time possible [18][22]. A key determinant of whether an organism will pass on its genes to offspring is the survival of that organism, and a key determinant of survival is the capability to locate and obtain sufficient food energy. Thus, evolution has favoured individuals able to gather more food more efficiently than other members of its species, leading to adaptations that maximize energy gain per unit of cost. IFT proposes that cognitive mechanisms recruited for information foraging also maximize the value of information obtained in relation to the cost of foraging.

IFT offers great promise as a framework for understanding the cognitive aspects of intelligence analysis. Mantovani [25] notes that, rather than conceiving of knowledge as the processing of “given” information, IFT views knowledge as a set of complex activities that include seeking relevant info, gathering it, and making sense of it. Thus, IFT captures the purposeful, directed nature of information gathering. This helps us understand the strategic nature of information foraging and the way information seeking behaviour must be adaptive to the task environment. IFT is also promising for practical reasons, as humanity has experienced tremendous growth of recorded information that allows more people access to more information than at any previous time in history [18]. Because the central problem of information gathering and sensemaking is the allocation of attention, IFT can be used to generate new techniques and technologies to aid people in finding and exploiting information.

Optimal foraging

For many organisms, finding food is a problem-solving exercise. It is likely that human intelligence has, to some degree, been driven by the need to develop sophisticated strategies for finding and obtaining food [26]. Certainly, organisms have developed a vast array of sophisticated cognitive mechanisms to solve that problem in their particular environmental niche. IFT assumes that these cognitive mechanisms can serve as, or form the basis of, information foraging processes.

To understand IFT, one must learn basic principles of Optimal Foraging Theory (OFT), from which it is derived [18]. OFT was developed to explain how organisms are adapted to their environment or, more precisely, how physical and behavioural traits of the organism evolved to address food foraging problems in the environment of the organism. OFT has been used to explain findings of ethological studies of food seeking and prey selection for many species (e.g., [27][28][29]). OFT relates these traits to affordances determined by environmental factors, such as available food types, the physical distribution of food, and energy costs of foraging and processing food items.

Physical and behavioural traits show their adaptation to the environment through their conformity to the demands of the environment. In other words, a trait is adaptive to the extent it offers the best means by which to solve some environmental problem. In the case of food foraging, the problem is to maximize the amount of energy gained from food while minimizing the amount of effort expended. Drawing an analogy, Pirolli [22] argues that the central problem of information foraging is to maximize the value of information gained from information sources while minimizing the amount of cognitive effort and time expended.

Pirolli [17](p. 23) describes optimization models as having three main components. First they must contain *decision assumptions* that specify the problem to be analysed in terms of strategies for deciding what food items to pursue, how much time to spend on processing a particular food item, and so on. This requires representations of the organism and the environment in which it lives, specifying the kinds of food available, their distribution, and so on. Second, optimization models contain *currency assumptions* that identify how choices are to be evaluated. These specify the measurement of resource value and costs. Finally, optimization models contain *constraint assumptions* that limit the relationships among decision and currency variables. These assumptions take into account the physical and biological processes that govern how the organism can operate within its environment.

OFT is based on the concepts of *gain* and *cost*. Gain refers to the accumulation of resource value in the currency of analysis. In optimal foraging models, the currency is typically caloric value as this is a suitable proxy for overall nutritional value [30]. Cost refers to expenditures of resources that bear on the organism`s survival. This can take the form of direct resource costs that are expenditures of energy incurred by foraging activities. Typically, energy in calories is considered a key cost as an organism must expend energy in moving to food sources and in processing food to extract its value (chewing, digesting). However, time is also a crucial cost. Time spent in foraging a particular food source incurs opportunity costs, which are benefits that could be gained by engaging in other activities but are forfeited by engaging in the chosen activity.

IFT makes use of the concepts of gain and cost but defines them with respect to cognitive tasks that require information. The resource currency of IFT can be generally considered to be “relevance” in terms of furthering the forager’s goals [17] (pp. 49-53). Relevance can only be determined contextually and, thus, defining gain is more complicated in IFT than OFT. Likewise, the costs of information foraging are different than those in food foraging, with less cost in terms of physical energy expended and more in cognitive effort. The opportunity costs in time associated with information foraging are more directly related to those of food foraging.

Conventional foraging models make use of a few simplifying assumptions to make the mathematical expressions of key functions more tractable [17] (pp. 31-33) [18]. First, these models assume that food is distributed in roughly co-localized groups in the environment. This is termed the “patch” structure of foraging environments, with resources occurring in aggregations that are themselves distributed unpredictably throughout the environment. Second, these models assume that a forager’s activities can be divided into two mutually exclusive types of activity: between-patch activities involving search for, and movement to, the next place to forage, and within-patch activities involving the exploitation of food resources [18]. A third assumption is that resource items vary in value, some being more valuable than others. This makes it important to try to find high-value items in order to maximize gain. Fourthly, foraging activities aimed at finding and processing resource items expend energy and time (i.e., costs). Finally, optimization models assume that an organism's goal is to obtain the greatest cumulative value in food resources while expending the least amount of energy/time as possible.

Working with these assumptions, it is possible to formalize the relationship between gain and costs. The following discussion summarizes Pirolli and Card's [18] description of this relationship.

The rate of gain of resource value per unit of cost can be calculated as the ratio of the total net amount of resource gained, divided by the total amount of time spent between patches and exploiting within patches.[1] If we let the rate of gain be R, the total net amount of resource gained be G, the time spent between patches be T_B, and the time spent within patches be T_W, then this relationship can be expressed by the formula [18]:

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

The rate of gain, R, is an important concept in optimal foraging theories as this is typically what organisms will attempt to maximize [18]. However, it is generally not possible to directly measure an organism's actual resource gain or between- and within-patch time expenditures. And even when this can be done, the calculation of the rate of gain, R, only provides an historical value for a particular foraging session. To use this concept as a guide to future foraging decisions, calculations must be based on observable averages of resource value and time.

To calculate a *predicted* rate of gain, it is assumed that the number of patches that can be processed is linearly related to the amount of time spent on between-patch foraging behaviour (searching, travelling) [18]. Further, it is assumed that the organism has estimates of the average time between patches, t_B, the average gain per item, g, and the average time taken to process items within patches, t_W.

From these estimates, it is possible to calculate the average amount of time taken to locate and move to a new patch; i.e., the average rate of encountering patches [18]:

$$\lambda = {}^{1}\!/_{t_B}. \tag{2}$$

The total amount of resource gained can then be represented as a linear function of between-patch foraging time [18]:

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

Equation 3 gives the total gain estimated to be obtained from foraging for a specified total amount of between-patch time, T_B, given the average rate of encountering patches, λ, and the average value gained from a resource item, g.

[1] Physical energy expenditure is typically left out of considerations in IFT as the actual physical effort of foraging for information is very small in relation to food foraging. Cognitive effort is highly correlated with time, meaning that it is possible to capture foraging costs solely with respect to time.

The total amount of within-patch time can be calculated from the specified total amount of between-patch time, T_B, the average rate of encounter, λ, and the average time taken to exploit a patch, t_W [18]:

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

Given this, the estimated rate of gain can be calculated by Holling's Disc Equation [18][31]:

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

Holling's Disc Equation indicates that, given known averages for between- and within-patch times and the average gain derived per resource item, it is possible to estimate the total gain an organism can expect to derive by foraging that item type over any given amount of time. This equation is important in OFT because it serves as the basis for deriving other foraging models. For example, the profitability, π, of patches is the ratio of net value gained per patch to the cost of within-patch processing [18]:

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

Equation 6 indicates that increasing the profitability of within-patch activities will increase overall rate of gain, R. Decreasing the between-patch costs, t_B, increases the overall rate of return R towards an asymptote equal to the profitability of patches, $R = \pi$ [18]. Equivalently, increasing the prevalence of patches, λ, also increases the overall rate of return R towards an asymptote equal to the profitability of patches, $R = \pi$. Originally developed to explain food foraging behaviour, Equations 1-6 offer a means of quantifying aspects of information seeking behaviour [18].

Extending the analogy between food and information foraging, it is possible to create quantitative relationships that capture fundamental aspects of the cognitive strategies underlying foraging decision making. Specifically, diet models and patch models are two important types of models that have been promulgated within the framework of OFT. Diet models consider an organism's decisions to accept or reject food items and predict choices of what resource items the organism will pursue [32]. Patch models predict choices of where to forage and, in particular, how much time to spend exploiting a grouping of items before moving to another location [18].

Basic information foraging concepts

The analogy between OFT and IFT is illustrated by the detailed correspondence between basic concepts, including resource value, cost, etc. Notably, IFT posits that human cognitive systems, analogous to food foraging mechanisms, have evolved in ways that allow them to maximize the value of external knowledge gained while minimizing the cost of foraging in terms of effort and time [22]. Pirolli [17](p. 13) refers to the concept of an extended phenotype, in which an organism's genotype has extended effects on the world that go beyond the actual body and behaviour of individual, to argue that cognitive mechanisms adapted to foraging for food have been extended to other tasks, such as foraging for information.

The challenge of IFT is in determining how to operationalize the analogy between information foraging and food foraging. This means determining exactly how to define concepts such as information value and information foraging costs, and to describe the information environment. Information value might be defined in terms of its task relevance; i.e., the extent to which a particular piece of information provides some useful contribution to the information forager's knowledge [17] (pp. 49-53). Relevance, however, is a difficult concept to operationalize as relevance depends on the forager and his/her task [15].

Pirolli [17] (p. 21) defines information value by its practical effect on advancing the forager's task goals. In a simple, well-structured problem, the value of knowledge gained from foraging can be expressed as a difference between the expected result of using that information and the expected result of not having that information. Pirolli [17] (p. 21) offers the example of purchasing a product on the Web. At one website the product costs \$X, but after visiting a price comparison site one can find an equivalent but cheaper product that costs \$Y. In this case, the net value of the information gained by foraging the price comparison website is \$X-\$Y-\$C, where \$C is a measure of cost of gaining the information. Thus the value of information is determined by the change in outcome that results from information foraging and by the cost of doing that foraging. If the change in outcome is small or the cost of foraging is high, the information has less value.

Pirolli's example illustrates some difficulties in measuring information value and foraging cost. Not all information-using tasks deal with monetary exchanges, so it is not always the case that value can be measured against a convenient external quantitative standard. Intelligence analysis focuses on developing situation awareness (SA) and predicting future events. To apply IFT to that domain, it will be necessary to determine some metric, whether objective or subjective, that can be applied to information value. In addition, we must be able to measure foraging cost in that system. Even in Pirolli's example, it is unclear that foraging cost can be easily expressed as a monetary cost.

An information forager can be seen as operating in two environments [22]. One is a task environment defined by the physical, social, cognitive structures that determine the task performer's activities and information requirements. The other is an information environment that consists of all the external knowledge that exists and permits person to perform his or her task. An information environment has its own inherent structure, with physical media separated according to type and location. But the organization of the information environment is also defined in part by the task environment, which determines the relevance or importance of various types of information in terms of the task performer's information needs.

Intelligence analysis can be characterized as an ill-defined problem in which the analyst's goals are dynamic, the scope of the problem is vast, and there are multiple valid solutions [1]. Such problems require that a great deal of information be collected to define goals, constraints, and courses of action [17]. Much of this information is needed to define the problem itself before potential solutions can be considered. This makes the definition of information value difficult as the forager's task goals may be hard to clearly express and those goals may change frequently as the forager develops a better understanding of the task.

The interface between the forager and the environment (i.e., information sources) determine the time costs associated with foraging and the opportunity costs incurred by choosing one information source to exploit and consequently foregoing all others [18]. Foraging costs include work done to access information sources, handling costs in manipulating the medium in which information is stored, and cognitive effort required to formulate and execute an information search strategy.

Diet models

Organisms face the problem of determining which food items to seek and consume. Because there is a cost associated with food foraging in addition to the value of food items, the organism must consider the value to be had from each potential food item in relation to the cost of obtaining it [32]. As a result, optimization of the overall energy gain from foraging generally means that some food items are not worth the effort of seeking them because those items do not return a sufficient gain of energy.

An organism, however, usually cannot know the exact energy value of an individual food item until it is eaten. For this reason, an organism can only make inferences about the expected value of food items based on their type. A given type of food (apple, tuber, chipmunk, or what have you) will be characterized by an average value for members of this type. Any given apple will be somewhat more or less valuable, depending on its size and composition, but there is a predictable range of energy that can be obtained from consuming an apple. An organism that knows the average value associated with each type of food in its environment can predict the relative rate of gain to be obtained from foraging one type of food rather than another.

This is the "diet problem," in which the organism decides what kinds of food items are likely to be worthwhile and what kinds will likely not to be worthwhile. An organism must determine a diet (i.e., an exclusive set of food types that will be foraged) that optimizes its energy gain per unit of cost. This problem is complicated by the facts that food types will differ in their prevalence and/or ease of finding, their energy value, and in the effort needed to process to extract their energy.

In information foraging, the diet problem has more to do with how an individual filters out irrelevant information than actively searches for relevant information [18]. The individual's task goals will usually define what kinds of information are needed but the environment will present the individual with a wide range of information types. It is essential that an information forager sets up some sort of attentional filter to discriminate information items that do not meet some criterion level of relevance. The diet problem can be thought of as determining a strategy for setting the filter criterion that will exclude information items. This criterion should be set according to some computation of benefit-cost ratio for the types of information in the environment.

For a complex task, there are often multiple types of information that can be employed, where types of information might be, for example, textual descriptions versus pictures, illustrations versus physical models, or verbal reports from a person versus transcriptions of conversations [18]. The types of information are defined by their physical medium as well as conceptual characteristics such as topic. Information types, like food types, vary in their prevalence, average information value, and cost in terms of search and processing activities [18]. Thus, an information forager faces the problem of deciding on which types of information should be sought in order to maximize his or her information gain per unit of cost.

Pirolli and Card [18] described an algorithm for creating an optimal information foraging diet based on conventional diet models developed for food foraging. Their model assumes that information types can be classified into $i = 1, 2, \ldots n$ types and that the forager has some knowledge of the average value and prevalence of items. The model contains the following variables:

- The *encounter rate* (frequency) for information items of type i: λ_i;
- The average *gain* of relevant information yielded by processing items of type i: g_i; and
- The *time* to process (within-patch time) individual items of type i: t_{Wi}.

Given these variables, one can define a diet, D, as a finite set of information item types, i, that a forager will seek. In other words, $D = \{1,2,3\}$ represents a diet consisting of items of type 1, 2, and 3 of n types of information such that the forager will only look for and exploit types 1, 2, and 3 and ignore all other types. The average rate of gain, R, yielded by diet D is given by the equation [18]:

$$R = \frac{\sum_{i \epsilon D} \lambda_i g_i}{1 + \sum_{i \in D} \lambda_i t_{Wi}}. \tag{7}$$

Equation 7 indicates that the overall gain consists of the sum of the average gains of each item type, weighted by each type's encounter rate, divided by the sum of the costs (in processing time) of each type, weighted by each type's encounter rate, plus a constant 1. Thus, the predicted gain can be calculated for any given diet if the individual average gains for all the item types are known in addition to the average within-patch processing times and encounter rates of the item types. Using this equation it is thus possible to evaluate all possible diets and identify the set of item types that, if exclusively foraged, will yield the optimal overall gain.

Although it is possible to evaluate all possible diets in principle, there may be a large number of information types, meaning a very large number of potential diets. It would be computationally complex and inefficient to calculate R for every possible diet and then select the one with the highest R. Instead, there is an algorithm that simplifies the determination of the optimal diet.

Pirolli and Card [18] employed the Optimal Diet Selection Algorithm [33] as a means to identify a forager's optimal diet. According to this algorithm, an optimal diet can be constructed by choosing item types in an all-or-none fashion according to their profitabilities. The profitability of an item type, π_i, is defined as the value of the item type divided by its cost in time to process [18]:

$$\pi_i = \frac{g_i}{t_{Wi}}. \tag{8}$$

According to the Optimal Diet Selection Algorithm, item types are rank-ordered according to their profitabilities. An initial diet is defined as the most profitable item and item types are added sequentially in order of decreasing profitability (i.e., the second most profitable, followed by the third most profitable, etc.). As an item type is added, the rates of gain of the initial and next diet are compared according to the formula [18]:

$$(k) = \frac{\sum_{i=1}^{k} \lambda_i g_i}{1+\sum_{i=1}^{k} \lambda_i t_{Wi}} > \frac{g_{k+1}}{t_{Wk+1}} = \pi_{k+1}. \tag{9}$$

The left side of Equation 9 concerns rate of gain obtained by the diet of the *k* highest profitability items (computed by Equation 7), whereas the right side of Equation 9 concerns the profitability of the *k+1* item types.

Conceptually, the algorithm can be understood by imagining an iterative process that considers successive diets of all item types. Initially, the diet includes only most profitable item, $D = \{1\}$. The next diet considered contains the two most profitable items, $D = \{1,2\}$, and so on. At each stage, the process tests the rate of gain, $R(k)$, for the current diet containing $D = \{1,2,\ldots,k\}$ types against the profitability of the next item type, π_{k+1}. As long as the gain of the diet is less than the profitability of the next type, $R(k) \leq \pi_{k+1}$, then the process should go on to consider the next diet, $D = \{1,2,\ldots,k+1\}$. Otherwise, Equation 9 is true and process terminates as adding the next item type would decrease the rate of gain for the diet.

Principles of diet selection

One implication of the optimal diet selection algorithm (Equation 9) is that the decision to include or not include a given information item type will be dependent on the profitabilities of information types ranked higher than that item [18]. Pirolli and Card [18] demonstrated that the steeper the decline in profitability from one item type to the next, generally the fewer items will be included in the optimal information foraging diet. In this case, when the profitabilities of higher ranked items are much greater than that of a given item type, the costs associated with including that item type will reduce the expected value to be gained from the diet below the threshold of the expected gain of including just the higher ranked item types. Similarly, when the prevalence of higher ranked item types is very large, this has the effect of making it worthwhile to exclude lower ranked items from the diet. Increases in the profitability or prevalence of high-ranked items leads to a narrowing of the optimal foraging diet.

This leads to some general principles of optimal diet selection. The first is the *Principle of Lost Opportunity*, which states that a class of items should be ignored if the profitability for those items is less than the expected rate of gain of continuing to search for other types of items [18]. That is, the gain derived by processing items of that low-profitability type is less than the cost of the lost opportunity to obtain higher-profitability types of items. By definition, foraging a given information item means that other items cannot be foraged at that moment. An opportunity cost is incurred when the value of the item being foraged is less than the expected value of other items that could be foraged in its stead.

The second principle is that of *Independence of Inclusion from Encounter Rate*, which states that the decision to pursue a class of items is independent of its prevalence [18]. Instead, the decision to include low-ranked items is solely dependent on their profitability, not upon the rate at which they are encountered. The decision to include a class of items *is* sensitive to changes in the prevalence of more profitable types of items (i.e., higher ranked item types). Generally, increases in the prevalence of higher-profitability items make it optimal to be more selective.

Another general principle is that the length of time it takes to process individual items will affect the decision to include that type of item in the diet. Generally, longer processing times result in greater opportunity costs because the time spent processing an item is necessarily time that cannot be used to search for other items [32]. This means that difficult-to-process items are less attractive than items that require less effort and time. But this factor must be considered in relation to the overall availability of resources in the environment. Diet models generally predict that when the environment is rich (i.e., a forager can quickly find desirable items), a forager will accept a narrow range of the most profitable item types. In contrast, when the environment is sparse (i.e., a forager requires a long time to find items), the opportunity costs are lower and the forager will accept a diet containing more low-profitability items [32].

Example of diet optimization

Pirolli [17] (pp. 42-43) provides the following example to illustrate the principles of diet selection in an information foraging context. A woman runs a small business that she conducts using email. Each email from a customer is an order for a product and she makes g_o = \$10 profit on each order. The woman also receives unsolicited junk email (spam) that occasionally offers a relevant service or product. It is assumed that 1 in 100 spam emails offers something that saves woman \$10, so the average gain from a spam email is g_s = \$10/100 = \$0.10. Initially, the woman received two orders per 8-hour day, for an encounter rate of $\lambda_o = 1/240$ orders per minute but this has improved to one order per hour for $\lambda_o = 1/60$ orders per minute. It takes 1 minute on average to read and process an email, both orders and spam ($h_o = h_s = 1$). By computing rates of gain, we can see that when the order rate is low ($\lambda_o = 1/240$), the woman should read both orders and spam (i.e. rate of gain of a diet of orders + spam is greater than that of a diet of orders only). But when the order rate is high ($\lambda_o = 1/60$), she should read only orders (i.e., rate of gain of diet of orders only is larger than a diet of orders + spam).

This example illustrates the Principle of Lost Opportunity, as the gain obtained from low-profitability items (spam) is less than the lost opportunity to gain higher profitability types of items (orders). Thus, even though there is some gain from spam, the woman can get greater value from searching for orders. The example also illustrates the Independence of Inclusion from Encounter Rate in that the decision is based solely on the profitability and prevalence of the more valuable type of item. Even if spam is encountered at a high rate, the woman should not include it in her diet. But the inclusion of spam *is* sensitive to changes in the prevalence of more profitable orders. Only when orders are encountered at a relatively low rate is it worthwhile to consider spam. Generally, increases in the prevalence of higher profitability items (or increases in encounter rate) make it optimal to be more selective.

Patch models

Patch models are extremely important in the study of food foraging behaviour due to the patchy structure of most environments. Often, food items occur in closely associated groupings (i.e., patches) that are themselves more sparsely distributed throughout the environment. When an organism locates a patch it is able to process a number of items sequentially with little between-item time but moving between patches takes effort and time [32]. Thus, decisions pertaining to where to look for patches and when to stop exploiting a given patch and move on are crucial to an organism's success.

Patchy information environments

Any given information environment can likewise be described as "patchy" to the extent information items occur in an organized fashion that causes subsets of items to be closely related or grouped and the resulting groups to be distributed in a way that renders them effortful to locate and access. In an information environment, patches are the result of different physical media as well as the organization of information by the creators of those media. Unlike food in a terrestrial environment, information is purposely organized, generally in ways that the creator believes will facilitate its use.

The patchy structure of information environments is illustrated by the World Wide Web (WWW). The WWW is not centrally controlled and thus unstructured to some degree but it generally exhibits a node-network structure [34][35]. Web sites are typically organized in a hierarchical fashion that reflects the site creator's intent [22]. But the WWW itself is largely uncontrolled and web sites are not organized in the same way. Thus, the overall structure is patchy, with web pages grouped in hierarchical trees of web sites but web sites are distributed in an unpredictable fashion [22]. This can also be termed a hub-structure, with web pages tending to be linked to a few "hubs" that serve to collect pages with related content [36] Hubs are sometimes purposely created but can arise because they are perceived as valuable or authoritative and attract links to pages [36]. It generally requires more effort to locate a web site than to navigate to a page within a web site.

Defining information patches can be difficult because the definition depends on the level of analysis one chooses [37]. Just as an information item must be defined through the interaction of the medium in which the information is represented and the goals and existing knowledge of the forager, an information patch can be an aggregation of items that corresponds to a web page, a web site, or a directory, etc. [38]. According to Pirolli [17] (pp. 49-53), the web page is a basic information patch that collects content and a variety of other interactive hypermedia elements such as links, pull-down menus, etc. Web sites are larger entities that collect multiple web pages as well as other info technologies such as databases into related groups. Even higher, Web portals act as central hubs, some of which (like Google) contain hundreds of millions of links hierarchically arranged by hundreds of thousands of semantic categories. Web portals also use search engine technologies to index the Web and dynamically generate links to Web content relevant to user's query.

Pirolli [38] argues that the WWW is "probabilistically textured," meaning that foragers are uncertain about the location, quality, relevance, veracity, etc. of the information being sought and the effects that foraging actions will have. This is the case because the WWW is not constrained (like a text editor or spreadsheet in which actions have fairly determinate outcomes) and information is not added and distributed by an overall controlling authority or single set of rules.

The patchy structure of the WWW has been demonstrated by several empirical studies. In a survey of over 600 million web pages from 12.5 million web sites, Eiron and McCurley [39] found that 41.1% of links in web pages connected to other pages within same directory (intra-directory links), 11.2 % linked to pages higher in the web site's directory (up-directory links), 3.9% linked to pages lower in the directory (down-directory links), 18.7% linked to other pages within the web site, and 25.0% linked to pages external to the web site. This means 75% of the links from web pages go to other pages in the same web site and 58% of the within-site links go to web pages in the same directory. Also there is a strong correlation between the tree distance measured in the directory structure of a web site and the probability of occurrence of a hyperlink at that distance. The probability of a hyperlink existing between two web pages decreases exponentially with their distance in the directory structure of the web site host [17] (pp. 49-53).

The content of web pages also indicates a patchy organization. Content is generally most similar among pages that have a direct link, followed by pages that are linked from the same parent web page, followed by pages linked to different parents, and finally random pages [40]. Similarly, increasing the link distance between web pages within the same domain is associated with decreases in similarity of content among those pages [41].

Optimizing time spent foraging in a patch

How well someone adapts to foraging in a patchy environment depends on that person's decisions concerning how long to spend exploiting resources in a patch and when to move on. These decisions are complicated by the fact that real-world environments are highly variable. A forager cannot count on all patches being equivalent or equally spaced, so there is not a fixed gain to each patch or a fixed amount of time required to move from one patch to another [18].

According to Pirolli and Card [18], foragers determine the optimal length of time to spend in a patch by comparing the rate of information gain derived within a patch to the cumulative gain function that describes the forager's overall expected rate of gain in the environment [18]. The forager controls the within-patch foraging time, t_{wi}, as he or she can leave that patch whenever desired. The amount of relevant information derived through foraging in a patch is expressed by the function $g_i(t_{wi})$. To calculate the cumulative gain, it is assumed that the environment contains multiple different patches, indexed as $i = 1, 2, \ldots P$. It is further assumed that the forager encounters patches of type i at rate λ_i as a linear function of total between-patch foraging time T_B. Given these assumptions, the total gain from foraging is computed by the formula (Pirolli & Card, 1999):

$$G = T_B \sum_{i=1}^{P} \lambda_i g_i[t_{wi}]. \qquad (10)$$

Thus, the total gain is the sum of the gains derived from each patch, weighted by the encounter rate for the patch, weighted overall by the total time spent between patches. Based on Equation 10, the overall average rate of gain is computed by the formula [18]:

$$R = \frac{\sum_{i=1}^{P} \lambda_i g_i[t_{wi}]}{1+\sum_{i=1}^{P} \lambda_i t_{wi}}. \tag{11}$$

Equation 11 provides the patch model for information foraging [18]. This model can be illustrated by plotting gain, G, as a function of time, as in Figure 1. A line, R*, is shown extending from the beginning of the average between-patch time, t_B. The slope of this line corresponds to the amount of value gained from patches, $g_i(t_{wi})$, divided by the time spent on between- and within-patch activities. The tangent of this line to the function gives the optimal within-patch foraging time, t^*.

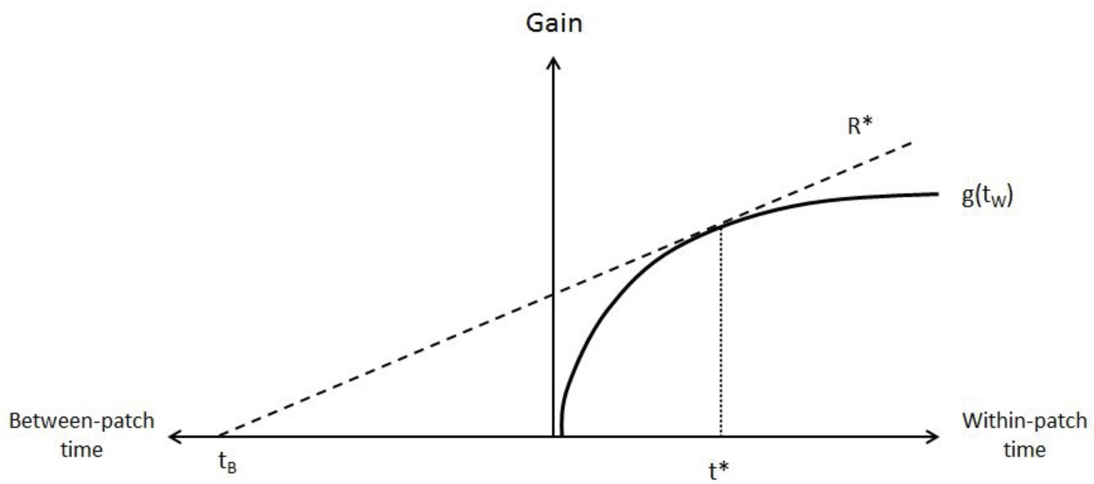

Figure 1: *Cumulative gain function showing diminishing returns as a function of within-patch foraging time. (Reproduced from Figure 5a, Pirolli & Card [18]).*

The problem of diminishing returns (i.e., a decelerating cumulative gain function as illustrated in Figure 1) is addressed by Charnov's [42] Marginal Value Theorem (MVT). Figure 1 illustrates the basic relations in Charnov's MVT for the case in which there is only one kind of patch-gain function. As described by Pirolli and Card (1999), 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, $\lambda = 1/t_B$. In Figure 1, the average between-patch time, t_B, is plotted on x-axis, starting at the origin and moving to the left. The optimal rate of gain, R^*, is determined by drawing a line tangent to the gain function, $g_i(t_W)$, and passing through t_B to left of origin. The slope of the tangent is the optimal rate of gain R. The point of tangency also provides the point at which the slope (marginal value) of g_i is equal to the average rate of gain R, the slope of tangent line.

Diminishing returns in information foraging is to be expected not just as a result of patches containing information items of varying value but also because there is likely to be overlap in the content of items [17] (pp. 37-39). As a forager exploits information items, he or she will increasingly encounter previously encountered ideas as redundancies are revealed. Charnov's MVT predicts that people should stay in a patch only as long as the slope of the within-patch gain function, $g(t_w)$, is greater than the average rate of gain, R, for the environment.

Enrichment of information patches

The optimal time to forage within a patch will vary according to the precise form of the gain function. This relationship can be exploited by a forager who is able to alter the environment in ways that change the gain function. Enrichment is the process of modifying the environment in ways that benefit the forager. Two basic approaches to enrichment are to reduce the between-patch costs of foraging and to enhance the results of within-patch foraging activities [18]. Between-patch foraging costs can be reduced by improving access to information sources and linking different sources via a common interface. An example of enhancement of within-patch foraging activities is the use of additional keyword queries to return a higher proportion of relevant results in online search engine.

Information foragers can enhance their environment in ways that reduce the average time taken to move between patches. For example, Pirolli and Card [18] note that a person foraging on the WWW can use saved links (i.e., “favorites”) to quickly move between well-known info sources. Also, physical media can be more efficiently organized and labelled to make it easier to select a desired source. Figure 2 illustrates the general effect of reducing between-patch foraging time. By decreasing the between-patch cost, the opportunity cost associated with staying in a patch is increased. Or, in other words, decreasing the between-patch cost increases the overall average rate of gain. This means that a forager should be more selective and leave patches earlier when between-patch time is lower.

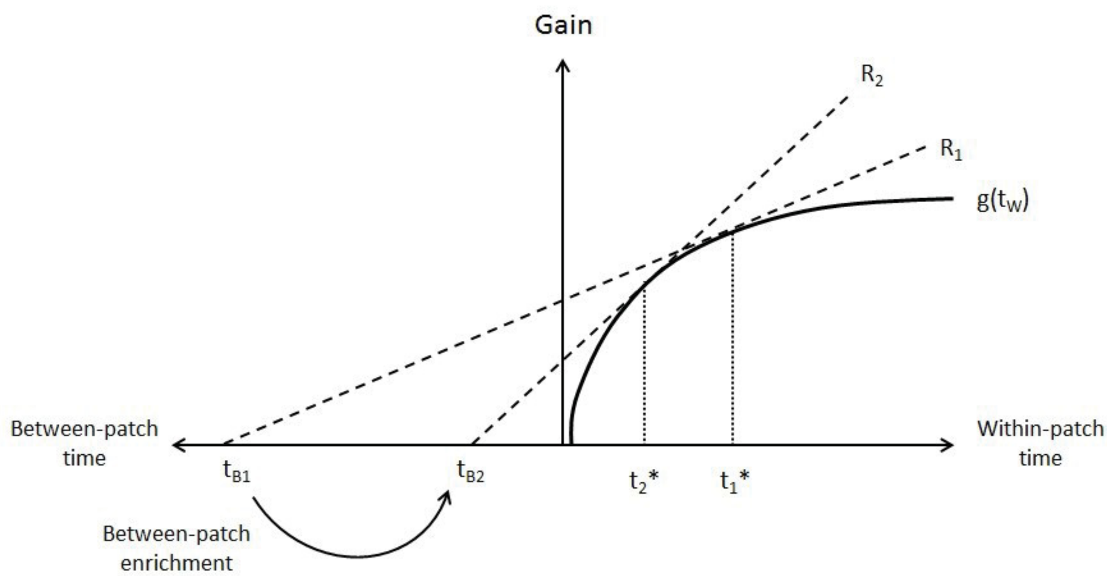

Figure 2: *Decreased between-patch cost results in a shorter optimal within-patch time. (Reproduced from Figure 5b, Pirolli & Card [18]).*

Other enrichment processes result in a reduction in the average time needed to process information items and, hence, a reduction in average within-patch time [18]. This type of enrichment is illustrated in Figure 3. Increasing the rate at which items can be processed results in a steeper gain function, $g_2(t_W)$, which in turn results in a steeper overall rate of gain, R_2. Thus, increasing the gain obtained from each item in a patch not only improves the overall gain rate but also reduces the optimal time needed to spend within the patch.

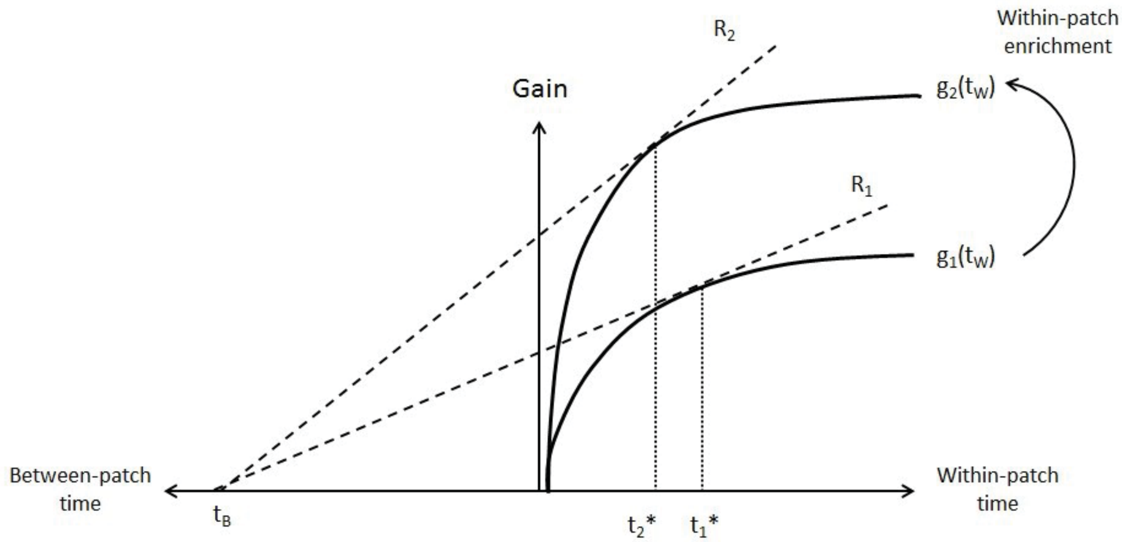

Figure 3: *Increasing the rate of gain within a patch results in a shorter average optimal within-patch time. (Reproduced from Figure 5c, Pirolli & Card [18]).*

The situation depicted in Figure 3 need not refer to an increase in the rate at which items can be processed but could also reflect enrichment that allows a forager to extract greater resource value from each item. If the value of each item is increased, holding processing time constant, the gain function for a patch still increases more steeply.

Optimizing patch foraging

The conventional patch model addresses the problem of determining how much time to spend in patches before leaving. Based on this, Pirolli [22] performed a rational analysis of foraging in information patches and developed a stochastic model that allows the forager to determine a stopping rule that guides the forager in this decision. This model is summarized in this section.

Pirolli [22] focused on search for information on the WWW and defined an information patch as a single web page within which text and pictures could be information items. In Pirolli's stochastic model, the experiential state of the forager at time i is represented as a state variable X_i, such that $X_i = x$ is a particular state value. This state variable includes some representation of the web page that has just been revealed and perceived by forager (i.e., the patch). The utility of continued link browsing in this information patch is represented as a function of the forager state, $U(x)$, as an expectation [22]:

$$U(x) = E[U|X_i = x]. \tag{12}$$

The expected time cost of future link browsing is also expressed as an expectation [22]:

$$t = E[T|X_i = x]. \tag{13}$$

In Equation 11, *T* is a random variable representing future time costs. The value, *U(x)*, of foraging for a period of time, *t*, in this patch must be balanced against the opportunity cost, *C(t)*, of foraging in the patch for that amount of time. This defines a function *h(x)* that describes the potential (i.e., expected gain) for continued foraging in this patch [22]:

$$h(x) = U(x) - C(t). \tag{14}$$

The optimal forager is one who maximizes this potential function. The opportunity cost, *C(t)*, is defined in terms of the overall long-term average rate of gain for foraging, $R*$ [22]:

$$C(t) = R^*t. \tag{15}$$

The overall rate of gain, $R*$, must be defined with respect to the same, or similar, task being performed by the forager. This determines the historical expected rate of gain when foraging the WWW for information relevant to the task.

The rationale behind Equations 12 and 13 is that the utility of foraging in the current patch must be greater than or equal to the average rate of returns for foraging in general among other potential web pages. If it is not, then continued foraging in the patch is incurring an excessive opportunity cost. This means an information forager should continue in the current patch as long as [22]:

$$U(x) - R^*t > 0. \tag{16}$$

The overall average rate of gain, $R*$, can be characterized in terms of: a) the mean utility of going to a relevant web site $\overline{U}$, b) the mean time spent on going to the next relevant site $\overline{t_s}$, and c) the mean time spent foraging at the new site $\bar{t}$. Thus, the Inequality 14 can be expressed as [22]:

$$\frac{U(x)}{t} > R^* = \frac{\overline{U}}{\overline{t_s} + \bar{t}}. \tag{17}$$

Equation 15 indicates that, in general, a forager should forage in an information patch as long as the value he or she derives per unit of time is greater than the expected value of foraging another patch per unit of the sum of within-patch and between-patch foraging time.

One feature of the WWW is that different web sites can generally be accessed quickly, through URLs (Uniform Resource Locators), links, or saved favorites [22]. This means that the time to access a web page within the current web site is often approximately the same as the time to go to a Web page at another web site. In this case, a forager's decision to continue foraging can be reduced to [22]:

$$U(x) > \bar{U}. \tag{18}$$

In this case, a forager should forage in an information patch until the expected potential of that patch is less than the mean expected value of going to a new patch.

Research has shown that people do behave adaptively when performing tasks requiring information foraging. Payne, Duggan, and Neth [43], for example, conducted a series of experiments in which participants exhibited sensitivity to the continuous rate of gain while performing a task but also a tendency to switch tasks when the rate of gain (in terms of overall performance/reward across both tasks) dropped below some threshold. They related their results to foraging heuristics described by Stephens and Kreb's [33] for food foraging. These heuristic rely on one of four variables: time in patch, number of items encountered, giving-up time (time since last encounter with an item), and rate of encounter of items. Depending on the environment, all of these variables can be used as the basis for a simple threshold rule to guide patch-leaving decision making.

In a similar study, Payne et al. [43] found that their participants, although sensitive to factors such as rate of return, did not use a simple "giving-up" heuristic to determine when to leave a given task or information patch. Instead, they argued that their participants used a somewhat more sophisticated procedure, called Green's Assessment Rule [44], to determine when to switch between tasks. According to this procedure, if an organism is sensitive to the rate of gain across an entire patch, the organism keeps track of an estimate of the potential (expected total gain) of the patch. In this way, each encountered item increases the potential of the patch but that potential also decreases as a function of time. Green [44] presented the analogy of a clockwork toy which winds down steadily over time but gets wound up a little bit each time it encounters an item. If the rate of encounter is high, the toy will keep getting wound up and continue to run. However, when the encounter rate declines, the rate at which the toy gets wound up decreases until it is lower than the rate at which the toy runs down and the toy will eventually stop. In a similar manner, each encounter with a food item slightly increases an organism's estimate of the potential of the current patch but that estimated potential simultaneously decreases as a function of time. When the encounter rate falls low enough, the estimated potential will drop below a threshold and the organism will leave the patch.

Payne et al. [43] found that Green's Assessment Rule provided the best description of their participants' behaviours. This rule works well for unpredictable environments, which their participants encountered. Basically, participants set an amount time for performing a task and increased this by a fixed amount with each success. This rule predicted that participants would devote more time to the easier of two tasks even though they took longer to give up in the harder task.

In related research, Duggan and Payne [45] observed people using a satisficing heuristic in "skim reading." The objective in skim reading is to gain as much information as possible in the shortest amount of time. A satisficing heuristic accomplishes this by monitoring the rate of information gain while reading and setting a minimal acceptable threshold value. According to the heuristic, a reader reads a section of text until the information gain rate falls below that threshold then moves to the next section. For the satisficing strategy to work well, the patches (sections of text) must be differentially valuable to the reader and the initial part of each patch must be indicative of the value of the patch. If valuable information is randomly or uniformly distributed, satisficing will not work well.

Information foraging on the World Wide Web

The patch structure of the WWW

A great deal of work on IFT has explored its use for explaining how people perform information tasks on the WWW. The WWW is a good domain in which to study human information foraging because people use the WWW to acquire knowledge to facilitate ill-structured decision making and problem solving and thus are required to perform extensive information searches [22]. The two main problems posed by the WWW environment correspond to the primary issues of IFT, namely, deciding which links to follow based on available cues and deciding when to give up on a current Web locality and go to another. Interacting with the WWW involves costs, especially the opportunity cost of time involved in searching a given web site, and so favours rational foraging behaviour that maximizes the value of knowledge gained relative to cost of interaction.

The WWW exhibits several important structural regularities that affect how a person can forage for information. First, the WWW is generally arranged into hierarchical patches. Simon [46] showed that information systems tend to evolve toward hierarchical organizations because such organizations are robust and efficient. On the WWW, lower level information patches such as web pages and search result pages are collected into higher level information patches such as web sites and web portals. The web page can be considered a basic information patch because it collects content and a variety of other interactive hypermedia elements such as links, pull-down menus, etc. [17] (pp. 49-53). Web sites provide access to web pages (i.e., a site collects multiple pages) as well as other information technologies such as databases or consumer services [17] (pp. 49-53). Web portals act as central hubs, some of which (like *Google*) contain hundreds of millions of links hierarchically arranged by hundreds of thousands of semantic categories [17] (pp. 49-53). Web portals also use search engine technologies to index the WWW and dynamically generate links to WWW content relevant to a user's query.

Empirical studies of the link structure of the WWW reveal a patchy structure. Eiron and McCurley [39] investigated the distribution of links in web pages and found that the frequency of links decreased with distance from the source. They found that 75% of links from web pages go to other pages in same web site and 58% of the within-site links go to web pages in the same directory. The hierarchical structure of a web site's directory greatly affects the probability that a hyperlink will occur between pages. The probability of a hyperlink existing between two web pages decreases exponentially with their tree distance in the directory structure of web site host [39].

The link structure of the WWW bears some similarity to the structure of scientific literature [17] (pp. 49-53). Important review papers generally have higher than average number of citations (links) to other papers, which makes them useful to learn about topic. This type of paper is a node in a network with many links emanating from them. Papers that are heavily cited by others are considered important and influential and have many inbound links. Similar structures appear in the WWW where the term *hub* is applied to a node with many outbound links and the term *authority* is applied to a node with many inbound links [17](pp. 49-53).

The patchy structure of the WWW is also seen in the organization of content. Pirolli [17] (pp. 51-52) [22] notes that the WWW exhibits "gradients of relevance," such that content topics are located proximally according to their semantic relationships. More specifically, the similarity of web pages decreases with the link distance between those pages in a negative power function. This results in content on the WWW being organized into "topical patches" that contain pages with related topics. Pages sampled from a common directory have greater similarity on average than pages selected randomly from different directories.

The patchy information structure of the WWW means that web-based information search will follow the same principles as patchy search in other environments. In particular, it will generally be the case that a forager will experience diminishing returns when exploiting information within a patch [17] (p. 52). Results of search engine searches, for example, are ordered according to predicted relevance to the user's query, with the highest ranked results most likely to be relevant and relevance decreasing as rank decreases. Also, there will be redundancies among pages so relevance decreases at more than linear pace because latter pages will contain info already available in higher ranked pages.

Typical information search tasks

In addition to the organization of information, an information environment is defined by the tasks and goals of the forager. This is illustrated by survey conducted by Pirolli and his colleagues to identify and describe representative WWW tasks and create a taxonomy of web tasks [17] (pp. 53-56). The survey question asked respondents to describe (in enough detail for another person to visualize the situation) a recent instance in which they found important information on the WWW. The survey question was answered by 2,188 respondents who were generally frequent users of web with greater than average experience in performing web-based tasks. Based on these responses, it was learned that 25% of web-based tasks involved finding some fact, document, product, or software download, and 75% involved some more complex sensemaking task such as making a choice or comparison (51%) or understanding some topic area (24%). The most common method employed was the directed search for specific information items (25%) or multiple information items (71%). In terms of the content of web-based information searches, 30% of foraging was for information related to products, followed by medical information, people, and computers. Overall, the survey indicated that a majority of WWW tasks are ones in which the individual is goal-driven and seeking information to make a decision or perform a complex task.

To understand the cognitive functions of people performing web-based tasks in more detail, Card et al [47] performed a protocol analysis with participants performing several tasks on a standard desktop computer, using the Internet Explorer browser to locate information on the WWW. A program called WebLogger collected and time-stamped all user interactions with the browser (e.g., keystrokes, mouse movements, scrolling, use of browser buttons, pages visited) as well as all significant browser actions (e.g., retrieval and rendering of Web content). Detailed protocol analyses were done on protocols from four participants working on two problems: one, called the *Antz* problem, required the participants to find a particular set of movie posters, and the other, called the *City* problem, required participants to find the date on which a Second City Troupe was to perform and to find a photograph of the group to use in an advertisement.

The participants' data suggested that their web foraging behaviour was structured by three problem spaces. A "link problem space" consisted of the information patches (typically web pages) the participant searched in addition to the basic operators involved in moving from one patch to another (e.g., by clicking on links or the back button). The "keyword problem space" consisted of alternative queries to search engines and the operators involved in formulating or editing the search queries. Finally, a "URL problem space" consisted of states that are legal URLs to be typed into the address bar of a web browser and the operators involve formulating and editing URL strings.

Card et al. [47] visualized participants' actions with web behaviour graphs (WBG) that illustrate the states and operators of the user in the three problem spaces. Analysis of WBGs indicated several phenomena. First, the Antz task was more difficult than City task and WBGs for the Antz task showed more branches and backtracking. Participants in the City task moved very directly to target information whereas participants in the Antz task followed more unproductive paths. The greater difficulty of the Antz task might have resulted in the poverty of cues that clearly led to desired target information. Second, participants visited multiple web sites but tended not to flit from one site to another. Participants exhibited more transitions within a site than between sites. By looking at the pages visited within a site and the order in which participants visited them, Card et al. [47] found that participants were very sensitive to cues about the content of web pages that could be used to guide information search. Generally, participants tended to remain at a web site as long as these cues promised a sufficient likelihood of finding additional useful information.

Information scent

The study by Card et al. [47] illustrated the role of "sensory" processes in information foraging. Although optimal foraging strategies can be devised solely on the basis of the stochastic distribution of food in the environment, virtually all organisms use some form of sense organs to enhance their foraging success. Sensory information allows an organism to improve its foraging performance by reducing the time it takes to locate food items and to distinguish desirable from undesirable items more quickly.

Just as organisms use sensory cues to decide where to forage for food, Card et al.'s [47] study shows that people can use cues embedded in information sources to decide on search paths through information patches. Pirolli [17] (p 68) devised the concept of *information scent* to describe the process by which information foragers make use of contextual cues to enhance their information foraging performance. Information scent refers to cues associated with navigation options that provide a forager with some indication of the nature and value of the content available by these options [17] (p.68)[22]. Rather than sensory cues, however, information scent generally refers to semantic context associated with navigation options for getting information.

The concept of information scent has become very important in understanding peoples' navigation on the WWW. Studies have shown that navigation choices are largely driven by how well the link labels semantically match the user's search goal [48]. Almost universally, web pages contain labeled navigation links from one web page to another (e.g., [22]). Web page designs have evolved to associate small snippets of text and graphics with such links (i.e., information scent cues). For example, most search engines return a list of results that are indicated by a title, phrases from the result containing words from the query, and a URL. These associated text and/or pictures are intended to convey the nature of the content available by following a link.

Whereas the kinds of sensory information usable by organisms in food foraging are generally readily evident it can be difficult to operationally measure information scent. To do so, we need a way to measure information scent in a wide variety of information environments. A key concept related to the value of information and possible the usefulness of information scent cues is that of relevance. The relevance of information is determined very specifically by the goals of the information user.

Budiu, Royer, and Pirolli [48] have suggested that semantic similarity as a potential measure of information scent. Semantic similarity can be assessed objectively by the co-occurrence of words in text, based on the assumption that words that co-occur in the same page of content or the same context are related to one another. A number of ways exist to measure semantic similarity, including Latent Semantic Similarity (LSA), Pointwise Manual Information (PMI), and Generalized Latent Semantic Analysis (GLSA), of which the latter two may be the more effective [48].

Modeling information scent

To explore the concept of information scent, and make it useful in predicting human information search behaviour, [17] (pp 69-81) performed a rational analysis of the use of contextual cues in surfing the WWW. Briefly summarized, this analysis comprises three parts: a) a Bayesian analysis of the expected relevance of a distal source of information associated with information scent cues, b) a mapping of the Bayesian model onto a spreading activation model (e.g., [49]), and c) a model of rational choice that uses spreading activation to evaluate the utility of alternative navigation options (see also [22]). The objective was to produce a model that could be applied to actual web-based information structures to generate optimal foraging strategies given the information scent cues available.

The Bayesian analysis of information scent is predicated on the assumption that an information forager makes predictions about the value of different navigation options based on available cues and selects the option that has the greatest expected value [22]. Pirolli [38] suggests that Brunswik's Lens Model offers a way to understand how information scent cues are used. In this framework, distal sources of information (i.e., sources that can be accessed by navigation actions such as link-following) have certain characteristics (link summaries, node labels, etc.) that serve as proximal cues. These cues can be used by a forager to make an inference about whether it is worthwhile to pursue that information source. The quality of the information scent cues is a function of their validity (i.e., the probability that the value of the cue leads to a correct decision). The quality also depends on the forager having an appropriate internal representation of the cues and their predictiveness.

The concept of spreading activation has been widely used to understand human cognition. A thorough account of spreading activation models is beyond the scope of this report but it can be said that this kind of model represents the human cognitive system in terms of a large network of highly inter-connected nodes [49]. Nodes represent semantic concepts and the connections between them weight the transmission of "activation" that can be thought of as cognitive activation or attention. The weights of connections attenuate the spread of activation that generally tends to decay as it spreads. Iterative activation of nodes representing external inputs and activation from the spreading of activation through connections represents the cognitive manipulation of semantic content.

The Bayesian rational analysis of information scent is mapped onto a spreading activation model by the formula [17] (pp. 78-79):

$$A_i = B_i + \sum_j W_j S_{ji}. \tag{19}$$

Here A_i refers to the activation of a node i, B_i is the base level of activation of that node, S_{ji} is the association strength between associated node j and node i, and W_j reflects attention on node j. Thus, the activation of node i is a function of its pre-existing level of activation plus the weighted sum of activation propagated from all associated nodes. The basic idea is that attention to information scent cues activates the nodes corresponding to the cues and causes activation to spread to associated nodes in a spreading activation network. The most strongly activated nodes are the ones that correspond to information elements that the forager expects to encounter by following the scent cues.

When the Bayesian model is mapped onto a spreading activation network, [17] (pp. 79-80)[22] applied a model for translating activation in the network into an expression of the expected utility of the activated nodes. This is done so the forager can understand the extent to which activated chunks in the network correspond to desired information. For this purpose, Pirolli [17] used the Random Utility Model (RUM), which is grounded in microeconomic theory. RUM provides a formula by which the predicted utility of distal info content is computed on basis of summed activation of all goal features plus a random variable error term that reflects a stochastic component of utility. The information forager can then select a navigation option if it has greater utility than all other options. The stochastic error component allows the model to account for random human error/variability.

This basic modelling framework is instantiated specifically to predict web-based navigation in the Scent-based Navigation and Information Foraging in the ACT architecture (SNIF-ACT) [17] (pp. 90-104). SNIF-ACT is a spreading activation model based on a large associative network that represents the web-user's linguistic knowledge. The spreading activation network is assumed to have an organization that reflects the statistical properties of the user's actual linguistic environment, which can be modeled by a very large corpus of words in the user's language. SNIF-ACT was designed to make two kind of predictions: 1) link selection behaviours (i.e., choices by an information forager about which links to click), and 2) decisions to stop following a particular path and try another [17] (p. 93). The basic architecture is described in Figure 4.

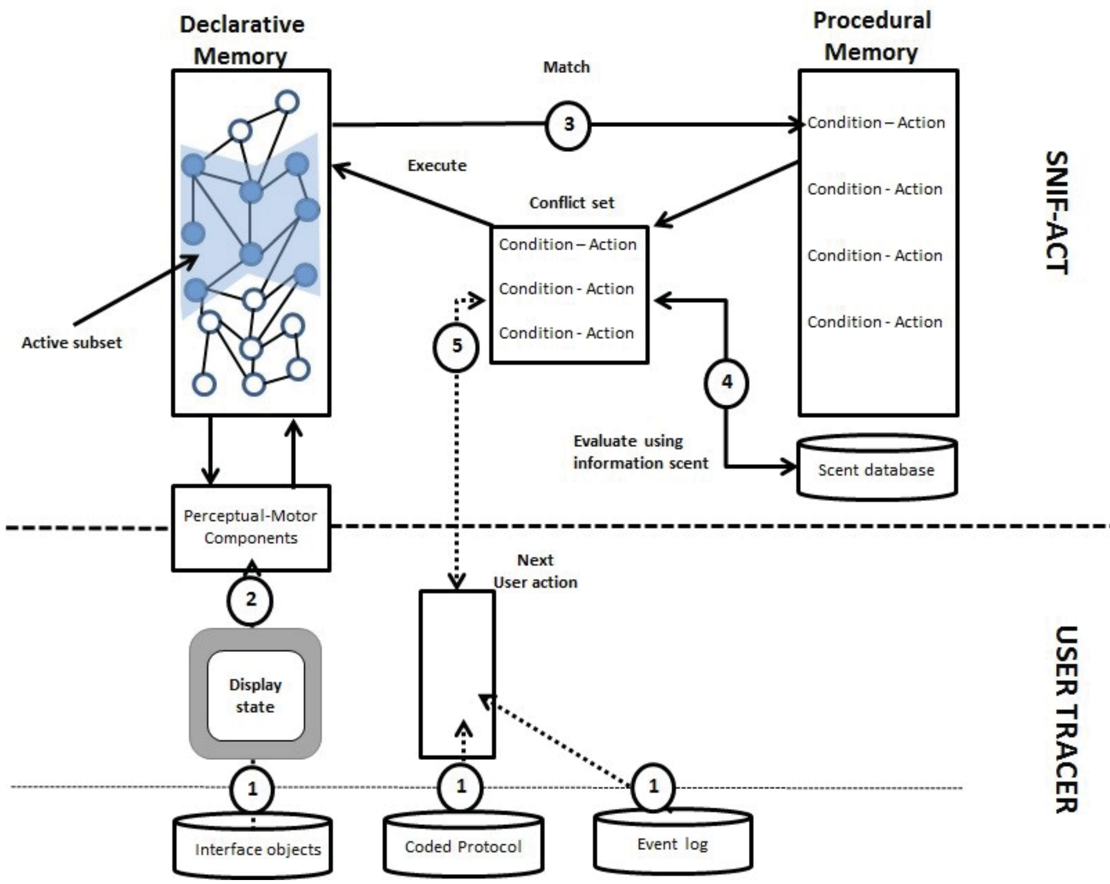

Figure 4: *SNIF-ACT architecture. (Adapted from Figure 5.1, Pirolli [17], p. 94, By permission of Oxford University Press, USA).*

In SNIF-ACT, declarative memory, storing general knowledge, in particular that pertaining to the content to be found on web links, functions of browser buttons, etc., is represented by nodes in the spreading activation network [17] (pp. 90-104). Procedural memory contains knowledge of how to do things represented by production rules; i.e., condition-action pairs [17] (pp. 90-104). These define what action a person will perform under a given set of conditions, although a conflict resolution mechanism can be called upon when conditions are consistent with more than one action.

Information scent is modeled in terms of the forager's task goals and the proximal cues that exist for available navigation options [17] (pp. 90-104). A task goal is represented as set of nodes in declarative memory corresponding to the information needed to perform the task. Proximal cues are also represented by nodes. When cues are encountered during the forager's interaction with a web page, the nodes representing the cues are activated and that activation spreads along semantic links. To the extent the cues are related to the task goal, activation will spread from cue nodes to goal nodes. The amount of activation accumulating on the goal nodes is an indicator of the mutual relevance of cues and goal and is matched to production rules that evaluate and select navigation choices.

A version of the model, dubbed SNIF-ACT 1.0, was developed by Pirolli and Fu [50] to fit the sequences of navigation actions (i.e., behavioural traces) observed of individual web-users. The SNIF-ACT 1.0 model was tested against detailed protocol data from Card et al. [47] that recorded the navigation actions of users attempting to perform several tasks that required accessing information using the WWW. The model predicted what link-following action a web-user should take, modelled by three production rules: *Attend-To-Link*, *Click-Link*, and *Leave-Site*. The utilities of these productions were determined by information scent computations based on a previous analysis of the task [17](Ch. 4). Utility was defined as the sum of all activation received by nodes representing the user's goal from proximal cues associated with the link, plus some stochastic noise. In general, if the scent associated with a link strongly activated goal-related nodes in the model's semantic network, that cue would have a high utility.

SNIF-ACT 1.0 predicts two major kinds of actions, 1) which links on a web page a person will click on, and 2) when a person decides to leave a page. The data logs from Card et al. [47] contained a record of human participants' navigation decisions against which the model's decisions could be compared for the same web content. Analysis showed that link-following actions were strongly predicted by scent-based utilities of navigation choices, as SNIF-ACT 1.0 predicts [50]. Analysis also showed that site-leaving actions were strongly predicted by the expected utility of further link following on the web page. Thus, on a given page, when the mean information scent of available links tended to be high users tended to click on links. Right before users left a web site, however, the mean information scent dropped, indicating that people tended to leave a site when the aggregate information scent of available choices dropped below general average value of foraging.

The accuracy of information scent

Users of the WWW seem to prefer to follow links over other means of web navigation [17] (p. 69). For the user, however, there is uncertainty about the relation of proximal cues to linked information resources and whether a link will lead to the desired information. In the complex network organization of the WWW, small perturbations in the accuracy of information scent can cause qualitative shifts in the cost of browsing [17] (p. 69). In this instance, accuracy refers to the precision of the expected navigation outcome that can be inferred from information scent. Information scent cues provide a probabilistic indication of what content is expected to be found by following a link. A forager who uses information scent to judge that a link will likely lead to relevant information may be disappointed because the information scent cue did not provide a certain prediction.

Figure 5 shows an hypothetical example of the impact of information scent inaccuracy on the time it takes to locate desired information [22]. The figure displays the search cost in terms of the average number of pages that must be viewed before arriving at a desired page. This is plotted against the desired page's "depth" in the network structure. Depth refers to the distance, in navigation points or links that must be navigated, from the current to desired web page. Several functions are indicated, corresponding to different false alarm (*f*) rates that indicate the probability that the information scent cue will lead the forager away from the desired information rather than toward it. When that false alarm rate is relatively low (10% and less), information scent makes navigation very efficient across all depths. When the false alarm rate grows higher, however, the search is much more inefficient, with inefficiency expanding with the depth of the desired page. Thus, the forager's search cost changes with depth very little with a false alarm rate of 0.015 to 0.100 but changes dramatically as the false alarm rate becomes greater than 0.100 [17] (p. 74) [22].

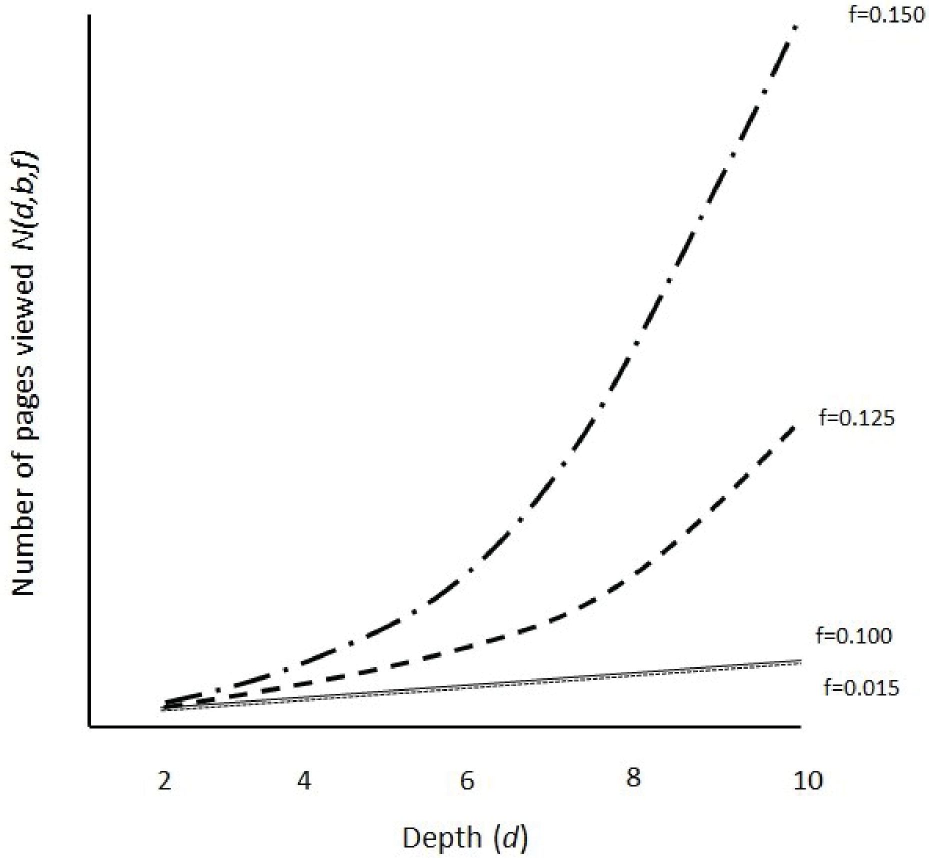

Figure 5: *Impact of information scent accuracy on web navigation (Adapted from Figure 3, Pirolli [22]).*

Applying information foraging theory to ntelligence analysis

This section lays out a plan for the application of IFT to the military intelligence domain. It begins with a discussion of the process of describing the task environment in which analysts work and moves to the issue of defining key IFT concepts in that environment. Perhaps the main challenge to applying IFT to intelligence analysis will be determining definitions for key IFT concepts such as resource value. To work, an IFT model must allow for quantification of the basic "currencies" of intelligence analysis. Currently, however, there is no objective standard for measuring benefits and costs that can be directly applied in the modeling effort.

The section ends with a discussion of ways the application of IFT may benefit military intelligence analysis. An IFT model can guide development of procedures and systems to support intelligence analysts in information search. Several examples of how IFT has aided information search in other domains are provided.

Roadmap for IFT research

Deriving practical solutions to the problems of information overload and severe time restrictions will require a research project focused on modeling the military analysis domain from the perspective of IFT. This project will seek to improve the effectiveness and efficiency of information search by intelligence analysts by creating models of intelligence analysts' information search strategies and comparing them to optimal strategies determined on the basis of the actual information environment.

This review of the scientific literature pertaining to information foraging serves as a starting point. The next step is to create a model of the intelligence analysis work domain. This is essential in order to relate analysts' information search strategies to underlying constraints that determine the relative efficiency of those strategies. Comparison of actual to optimal strategies will allow the development of training materials and decision support concepts.

Describing the intelligence environment

In applying IFT to the information search behaviour of intelligence analysts, we rely on the analogy to biological organisms' search for food in the environment, which casts analysts as "informavores" who use adaptive strategies to locate and make use of information. For this analogy to be useful, however, we must understand the task and information environments in which analysts work in a way similar to which researchers of animal behaviour understand the physical and biological environments in which organisms live.

Intelligence analysis is a complex domain that requires analysts to employ a wide range of different information types that are obtained from a variety of sources [1]. Thus, it is to be expected that a well-planned and thorough approach will be needed to characterize the task environment. One such approach is to conduct a Cognitive Task Analysis (CTA) in conjunction with a Work Domain Analysis (WDA).

CTA comprises a variety of interview and observational methodologies to create a representation of the knowledge and decision strategies employed by people performing a specified task or set of tasks [51]. This representation aids in the development of training and decision support systems. With a wide variety of primary methods available, CTA is readily tailored to the specifics of the particular work domain and has been widely applied in many fields [51].

WDA distinguishes between the environment in which a person works and that person's behaviours [52][53]. Rather than attempt to characterize the way someone performs a task, as in CTA, the goal of WDA is to identify the constraints that limit what actions a person may perform [52] (p. 19). In identifying all these constraints, WDA creates a representation of the space in which a person can select actions to complete tasks. As a result, WDA is a technique that allows one to characterize the scope of possible work-related activities as opposed to current or prescribed practices [53].

Combining these techniques to examine intelligence analysis will yield a definition of the information space in which analysts forage as well as descriptions of information search strategies employed by analysts. The information space will identify all possible information sources that analysts can access (including sources that could be accessed even if analysts currently do not make use of that source) and the distribution of different information types within those sources. Moreover, the results will indicate how analysts formulate information needs, which determine the relative values of different types of information and different items within each information type.

Parts of these kinds of analyses have already been performed for intelligence work. Hutchins, Pirolli, and Card [13][54], for example, conducted a CTA of military intelligence analysis with students enrolled in the U.S. Naval Postgraduate School. They were able to identify key analyst activities and information needs but focused on the cognitive challenges that make intelligence analysis a challenging process, including extreme time pressure, high cognitive load, and need to combine information from multiple sources. It is possible to provide a more systematic description of the analysis information environment with a focus on what are possible information sources rather than what are currently used sources. More importantly, analyses that concentrate on the Canadian perspective are needed to identify specific task goals that will determine information needs and how analysts evaluate information.

Defining key concepts

The first step in identifying optimal information foraging strategies is the same as the first step in identifying optimal food foraging strategies, namely to identify what the forager is attempting to maximize [30]. It can seem as though defining resource value for information foraging is a daunting enterprise, as information itself, with the interplay of different kinds of physical media and the almost infinite potential cognitive structures of different people, is not a simple concept. How can an outside observer hope to determine the value of information to a specific individual analyst? Winterhalder [30] points out that this problem is not unique to IFT but exists for researchers studying food foraging of organisms. In biology, the fundamental evolutionary property for an organism is reproductive fitness and so all foraging activities should be evaluated with respect to that. It is, however, impossible to directly observe reproductive fitness, especially on the level of an individual, and so researchers generally have to resort to various proxy measures – observable and measurable qualities that are assumed to be directly related to reproductive fitness. Thus, in OFT, it is assumed that the energy value of food items, which can be determined based on biochemical analysis, serves as a suitable proxy. Food energy directly contributes to metabolism and activity for the organism, which in turn contributes to reproductive fitness.

To create an IFT model of intelligence analysis, what is needed is an appropriate proxy measure for the impact that individual information items have on the usefulness of the end intelligence product. Various metrics have been proposed, such as information accuracy, user confidence, and degree to which an information item changes an analyst`s understanding of the current problem [15][55]. One potential measure is information relevance in terms of the degree to which an item provides some useful contribution to the analysis process [56]. The problem with this measure is that relevance can be difficult to define and even more difficult to operationally assess.

The difficultly in operationalizing the concept of information value is that information itself can be defined in at least two ways, as a property of some external medium or as a property of the human mind that interacts with that medium. Communication researchers distinguish two ways of defining media: an attribute-based approach in which content is related to objective characteristics or attributes of the media itself, and an effect-based approach in which content is defined in terms of the psychological states that the media creates in users of the media [57]. The attribute-based approach assumes that message content and features of media are associated with specific cognitive and/or emotional responses in the users. Thus, one can talk about specific media possessing specific meanings. The effect-based approach, on the other hand, does not assume that the media itself must be associated with any specific mental state within the user. Rather, content is entirely constructed by the user through interaction with the media and the media itself cannot be said to possess or contain any specific content.

The choice between an attribute-based and effect-based concept of information determines whether relevance can be assessed with respect to objective qualities of external objects (media) or must be an exclusively subjective quality. Another complication is the speed with which the information needs of an analyst can change while gathering information. Relevance must be assessed with respect to the task goals of the analyst but these are unlikely to be stable for extended periods of time as the analyst develops a dynamic mental model of the task. We can talk about situational relevance as the relationship between media content and the analyst`s mental model at a particular time or point in the analysis [58].

Borlund and Ingwersen [58] contrast two types of relevance measures: situational relevance and topicality. As described above, situational relevance measures are subjective and related to the internal representation of information need based on the analyst`s current mental model. Topicality is an objective or system-based measure of relevance that assesses how well the topic of information matches the topic of relevance. The topic of a medium (document, video, etc.) is considered a property of the actual media and so topicality is an attribute-based measure, whereas situational relevance is an effect-based measure.

The most sophisticated approach to assessing the usefulness of information in the context of military intelligence analysis has been developed by Hammell and colleagues [59][60][61]. They developed a measure termed *Value of Information* (VoI) as a way to improve support to data collection. This measure is based on an annex to NATO STANAG (Standard Agreement) 2022 and Appendix B of the U.S. Army FM-2-22.3, both of which describe a procedure for assigning alphanumeric ratings of the confidence (or trustworthiness) and applicability (or truthfulness) of information items by analysts [59]. Although these procedures are clearly specified, they rely on subjective assessments and can be time-consuming to perform manually. Moreover, the doctrine does not indicate how analysts should take into account the specific mission context in making assessments [60]. For this reason, Hammell and colleagues developed a procedure for automatically assigning VoI values to information items.

The NATO STANAG (Standard Agreement) 2022 and Appendix B of the U.S. Army FM-2-22.3 distinguish the reliability of an information source from the accuracy or truthfulness of the content, deeming both important in assessing the value of information. Thus, they establish separate 6-point rating scales on which analysts can assess *source reliability* and *information content*, with both being used in judging the value of the information. The VoI approach uses these scales as bases for an automated assessment procedure.

Hammell and colleagues automated this assessment process using Fuzzy Associative Memory (FAM) structures [60]. FAM structures are multidimensional tables in which each dimension corresponds to an input or measure of some external quantity. In this case, the dimensions correspond to ratings of source reliability and information content, and categories of mission context. A FAM then serves as a lookup table in which values in the cells of the table are determined by the row and column vectors and a particular combination of input values ratings will yield a particular output value.

The main inputs to the model are analysts' ratings of individual pieces of information being considered for analysis. The analyst must make a quantitative assessment of the reliability, truthfulness, and timeliness of each individual piece of information. Ratings of source reliability and information content are fed into what is termed the Applicability FAM [60]. The Applicability FAM computes a single value representing the level of relevance of the information to the analysis. The value outputted from the Applicability FAM is fed into a second FAM, the VoI FAM, along with a categorical indicator of the timeliness of the information. Hammell et al. [59] employ a 3-level scale for timeliness. The output of the VoI FAM is a single value representing the value of the information.

The VoI model is one way to address the issue of defining and quantifying the concept of resource value for an IFT model of intelligence analysis. If VoI computations were to be performed for all information items assessed for analysis, it would be possible to map the distribution of information value across different information sources and estimate the average value associated with different information patches.

Potential impact of IFT on support to intelligence analysts

A major part of intelligence analysis is the search for information. Among the main challenges of intelligence analysis are information overload and severe time restriction. In addition there are risks associated with intelligence analysis that could impair the quality of intelligence. Not surprisingly, issues related to information search and sensemaking have been prominent in efforts to enhance intelligence analysis [7]. Intelligence analysis is a domain that could benefit from application of IFT.

IFT has the promise to help intelligence analysts reduce the amount of time they spend searching for information, presumably allowing more time for sensemaking activities essential to gaining true information superiority [2]. The main way IFT can enhance intelligence analysis is by providing a quantitative model on which to base systems, procedures, and decision support. An IFT model will define the analysis process in terms of the actors, resource currency, and task constraints, and identify what decision strategies are optimal within the task environment.

Automated analysis of analyst goals

One way in which IFT might be applied to intelligence analysis is through automated identification of analyst goals. This is an approach that is well-suited to computerized information systems and the WWW. Web sites, for example, generally record user interactions in some way, such as links clicked, time spent on page, etc. [62]. These data can be used to summarize and analyse user behaviour from which it is possible to infer the kinds of information the user is seeking or will need to accomplish some goal. It is then possible to provide some form of guidance to the user to assist in locating relevant information [63].

Bloodhound and Lumberjack are two systems designed to infer users' goals based on their behaviour as represented in Web log files [17] (pp. 175-177). Lumberjack, for example, analyzes navigation behaviour (e.g., links chosen at each page, amount of time spent at each page) and web site features (e.g., hyperlink structure of web site, content of web site pages) to construct user profiles (see also [64]). User profiles are some representation (e.g., vector of word association strengths) that can be taken to represent user's goals. Profiles may be submitted to analysis techniques such as clustering to group together users with similar goals.

Automated parameter tracking

It may possible to enhance analyst performance by giving them objective feedback on key aspects of their foraging behaviour. Automated parameter tracking refers to the systemic monitoring of variables such as average time taken to shift from one data source to another, average time viewing a particular source, and the relevance of obtained results. These variables correspond to variables in IFT models: between-patch time, within-patch time, and resource value. Yet, information systems often fail to provide any indication of which sources are most promising, how best to search within an information source, or how individual items can be rank ordered in terms of value [65].

Automated parameter tracking, in conjunction with analysis of users' goals, can also be used to automate portions of the information search process. The Time Bounded Reasoning (TIBOR) agent is an example of a decision support system designed to aid foragers in information search [66]. The agent uses an intelligent user interface to assist an analyst in sensemaking by gathering information to validate hypotheses and eliminate incorrect hypotheses. It uses interactive visualisations to enable an analyst to gather and sift large amounts of evidence in reasonable time and to collaborate with others. TIBOR employs an AI blackboard system and resource-bounded control mechanisms. It handles three types of decisions: gathering of large scale, high dimensional data from a variety of sources, determining the type of processing to extract data from these sources, and determining appropriate interactive visualisation of these data.

Hoare and Sorensen [65] have also proposed the use of “recommender” systems that present suggested actions based on the user’s previous behaviours or the behaviours of some collaborative group. These suggestions can be made specific and context-related by using parameters derived from the user’s foraging activities.

Enhancing scent cues

Typically, WWW users access content by following links from one page to another. The content of the pages associated with links are usually presented to the user by some snippets of text or graphics to allow the user to predict what content will be encountered by following the link. Such information scent cues are imperfect as is the subjective perception of the value and cost of information sources obtained from proximal cues. Thus, another avenue to enhancing intelligence analysis is to somehow make better use of information scent to enhance information foraging performance.

One potential way to do this is to develop some kind of automated scent cue generator. Varying the length and content of the text snippets associated with links can dramatically affect their usefulness as scent cues [67]. Given the use of some user-goal inference devise, it may be possible to generate user-specific text snippets that are more goal-related so that they convey more and better information about the content of a web page.

Chi et al. [62] developed the Web User Flow by Information Scent (WUFIS) algorithm to predict WWW-users’ navigation decisions based on information scent cues. This algorithm analyses the quality of scents cues in relation to the page content they lead to and generates a probability that a user will navigate that link. Experiments performed by Chi et al. [62] indicated that WUFIS works well to predict the relevance of web pages based on the proximal scent cue associated with pages and user’s goals.

The ScentTrails system [17] (pp. 179-180) [68] is an approach to link navigation that modifies the rendering of link information to enhance info scent cues that are predicted to be particularly useful given users’ goals. When a user indicates an information goal, ScentTrails identifies a set of relevant pages at a web site. Using a graphical representation of the link topology of the web site, ScentTrails initializes nodes representing those relevant pages with some score corresponding to their relevance (scent). These scent values are spread through the graph from relevant target pages, flowing backward along links (opposite the direction the links would be browsed). At each link, some scent is lost, so scent diminishes exponentially as a function of link distance. This process spreads scent back from the target pages through intermediate web pages at a site in a way that reflects the cumulative scent from paths emanating from a page. The amount of scent is used to scale the highlighting of links, so links with greater scent are larger and more salient.

It may even be possible to enhance the use of information scent by maintaining records of past information searches. Wexelblat and Maes [69] proposed that information systems maintain “interaction histories,” or records of the interactions of multiple users with an information system. These records can then serve to guide subsequent searches for similar information by later users.

Visualisation techniques.

Information foraging can also benefit from the science of visual analytics, which examines how human reasoning can be facilitated by interactive visual interfaces [70]. Interactive displays are those that can be adjusted to present information in ways that are consistent with the user's task and goals. Hoare and Sorensen [65], for example, propose that people can more effectively forage for information in a 2-dimensional representation of information space. A list of options, such as generated by a search engine, is a 1-dimesnional representation. A list can be ordered by one factor, such as relevance to a search query. A 2-dimensional representation, however, allows for the similarity among search results to be depicted.

Hoare and Sorensen [65] have proposed their own 2-dimensional information map called SolonEvo, which indexes document collections and makes them searchable. This tool makes use of visual clustering of search results to help users evaluate results efficiently, presenting results in a visualizable set of patches for foraging. Similarity is represented by distance with more similar items being close to one another in the display. Clustering items in this way makes it easier to evaluate the relevance of clumps of items. If one item in a cluster is rejected, the user does not have to waste time looking at others in that cluster. Hoare and Sorensen [65] argued that combining recommender systems with 2-D visualisations can greatly enhance search efficiency and minimize between-patch time and to let user know when all relevant info in patch has been consumed.

Conclusion

Military intelligence analysis comprises a complex and demanding set of activities. One of the most important of these is information search, which currently consumes a great amount of analysts' time [5][12]. The nature of military intelligence is such that analysts have a continual need for new and updated information and must search many diverse sources to obtain that information. This means that analysts look through a huge volume of information in search of specific items of relevance to their particular topics. Despite the amount of work involved in information search, analysts typically work under severe time restrictions [2][7]. Consequently, there is a serious need to maximize the value of the information analysts obtain through search in whatever time is available.

IFT offers a way to examine intelligence analysis from this perspective of adaptive efficiency. As a theoretical perspective, IFT takes advantage of a mature area of research in OFT, which has been successful in predicting the food foraging behaviour of numerous species. OFT does this by examining an organism's foraging strategies with respect to the statistical structure of its environment. The kinds of things that can be done with OFT include predicting an organism's preferences for food items (diet models), predicting how an organism will exploit its environment, in particular how it allocates time to food patches (patch models), and predicting how an organism's sensory mechanisms enhance foraging (use of cues). In drawing the analogy between information search and food foraging, IFT adapts the modeling techniques of OFT to do the same kinds of things in the domain of information search [17][18].

Despite the promise of IFT, a great deal of work needs to be done to apply it to the military intelligence analysis domain. The first step will be to describe the "information environment" in which analysts operate. The information environment is analogous to the physical environment in which organisms forage for food and it sets the constraints on action to which analysts must adapt in order to achieve optimality in their search behaviours. Describing the information environment entails examination of specific analyst roles as the specific goals and areas of responsibility of each analyst are critical in determining what information they seek and what resources they can use. The description of an information environment will take the form of a comprehensive record of tasks, information sources, types of information needed, information technologies, and information search strategies associated with an analyst's role.

With a description of the information environment, it is then possible to define the key concepts needed by IFT. Precisely defining information as a resource and quantifying its value will be more challenging than defining resource value for food foraging. Organisms possess broadly similar metabolic systems that convert organic matter to energy. Information users, however, seek and use information for a wide range of purposes and it is likely that information value must be defined in a very context-specific manner. Rather than attempting to tackle the whole of intelligence analysis, it will be important to limit initial modeling to a limited set of analyst positions to ensure the modeling process is a tractable problem.

The description of the information environment is the basis for modeling the statistical structure of that environment. The statistical structure refers to the distribution of resource value in the environment, which is assumed to be non-random and non-uniform. This structure determines the constraints that one must satisfy to forage optimally. Thus, it is possible to define optimal foraging strategies for the described environment (it is, of course, critical to accurately describe the environmental constraints to ensure that the optimal strategies actually apply). The foraging strategies actually used by analysts can then be compared to optimal strategies and actions devised to reduce discrepancies between them. A wide range of options exist to remediate sub-optimal information foraging.

This page intentionally left blank.

References

[1] Gouin, D., & Lavigne, V. (2011). *Trends in human-computer interaction to support future intelligence analysis capabilities*. Paper presented at the 16th International Command and Control Research and Technology Symposium. Quebec City, Canada.

[2] Rudner, M. (2002). The future of Canada's defence intelligence. *International Journal of Intelligence and CounterIntelligence, 15*, 540-564.

[3] Barber, J. (2001-2002). An intelligence, surveillance and reconnaissance (ISR) vision for the Canadian Forces. *Canadian Military Journal, Winter*, 41-46.

[4] Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993, May). The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 269-276). ACM.

[5] Puvathingal, B. J., & Hantula, D. A. (2012). Revisiting the psychology of intelligence analysis: From rational actors to adaptive thinkers. *American Psychologist, 67*, 199-210.

[6] Pirolli, P., & Card, S. (2005, May). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis* (Vol. 5, pp. 2-4).

[7] Badalamente, R. V., & Greitzer, F. L. (2005, May). Top ten needs for intelligence analysis tool development. In *proceedings of the 2005 international conference on intelligence analysis*.

[8] Fingar, T., Fischhoff, B., & Chauvin, C. (2011). Analysis in the US intelligence community: Missions, masters, and methods. In B. Fischhoff and C. Chauvin (Eds.), *Intelligence analysis: Behavioral and social scientific foundations* (pp. 3-27). The National Academies Press: Washington, DC.

[9] Moore, D. T. (2011). *Sensemaking: A Structure for an Intelligence Revolution.* Washington, DC: National Defense Intelligence College.

[10] Bandali, F., Bruyn, L., Vokac, R., Keeble, R., Zobarich, R., Berger, N., Rehak, L., Lamoureux, T. (2007). *CF Procedures and Practices Involving Information Aggregation*. DRDC Toronto Contractor Report (CR 2007-049). Defence R&D Canada, Toronto. Toronto, Canada.

[11] Hastie, R. (2011). Group processes in intelligence analysis. In B. Fischhoff and C. Chauvin (Eds.), *Intelligence analysis: Behavioral and social scientific foundations* (pp. 169-196). The National Academies Press: Washington, DC.

[12] Ayoub, P. J., Petrick, I. J., & McNeese, M. D. (2007, October). Weather Systems: A New Metaphor for Intelligence Analysis. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 51, No. 4, pp. 313-317). SAGE Publications.

[13] Hutchins, S. G., Pirolli, P., & Card, S. (2003, October). Use of critical analysis method to conduct a cognitive task analysis of intelligence analysts. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 3, pp. 478-482). SAGE Publications.

[14] Pfautz, J., Fichtl, T., Guarino, S., Carlson, E., Powell, G. & Roth, E. (2006). Cognitive Complexities Impacting Army Intelligence Analysis. Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting (pp. 452-456). Sage.

[15] Scholtz, J., Morse, E., & Hewett, T. (2004). In depth observational studies of professional intelligence analysts. In *Human Performance, Situation Awareness and Automation Technology Conference*.

[16] Cook, M. B., & Smallman, H. S. (2008). Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence. *Human Factors, 50*, 745-754.

[17] Pirolli, P. (2009). Information foraging theory: Adaptive interaction with information. New York, NY: Oxford University Press.

[18] Pirolli, P., & Card, S. (1999). Information foraging. *Psychological review*, *106*(4), 643.

[19] Anderson, J. R. (1989). A rational analysis of human memory. Varieties of memory and consciousness: Essays in honour of Endel Tulving, 195-210.

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

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

[22] Pirolli, P. (2005). Rational analysis of information foraging on the web. *Cognitive Science, 29*, 343-373.

[23] Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L., & Wakefield, J. C. (1998). Adaptations, exaptations, and spandrels. *American psychologist*, *53*(5), 533.

[24] Wells, V. K. (2012). Foraging: An ecology model of consumer behaviour? *Marketing Theory, 12*,117-136.

[25] Mantovani, G. (2001). The Psychological Construction of the Internet: From Information Foraging to Social Gathering to Cultural Mediation. *CyberPsychology & Behavior, 4*, 47-56.

[26] Mobus, G. E. (1999). *Foraging Search: Prototypical Intelligence*. Paper presented at The Third International Conference on Computing Anticipatory Systems, Liege, Belgium.

[27] Davies, N. B. (1977). Prey selection and the search strategy of the spotted flycatcher (*Muscicapa striata*): A field study on optimal foraging. *Animal Behaviour*, *25*, 1016-1033.

[28] Krebs, J. R., Ryan, J. C., & Charnov, E. L. (1974). Hunting by expectation or optimal foraging? A study of patch use by chickadees. *Animal Behaviour*, *22*, 953-964.

[29] Mittelbach, G. G. (1981). Foraging efficiency and body size: a study of optimal diet and habitat use by bluegills. *Ecology*, *62*(5), 1370-1386.

[30] Winterhalder, B. (1981). Optimal foraging strategies and hunter-gatherer research in anthropology: Theory and models. *Hunter-gatherer foraging strategies: Ethnographic and archaeological analyses*, 13-35. Chicago, IL: University of Chicago Press.

[31] Holling, C. S. (1959). Some characteristics of simple types of predation and parasitism. *The Canadian Entomologist*, *91*(07), 385-398.

[32] Stephens, D. W., Couzin, I., & Giraldeau, L-A. (2012). Ecological and behavioral approaches to search behavior. In P. M. Todd, T. T. Hills, and T. W. Robbins (Eds.), *Cognitive search: Evolution, algorithms, and the brain* (pp. 25-45). Cambridge, MA: The MIT Press.

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

[34] Mukherjea, S., & Foley, J. D. (1995). Visualizing the world-wide web with the navigational view builder. *Computer Networks and ISDN Systems*, *27*(6), 1075-1087.

[35] Pirolli, P., & Card, S. K. (1998, May). Information foraging models of browsers for very large document spaces. In *Proceedings of the working conference on Advanced visual interfaces* (pp. 83-93). ACM.

[36] Fu, W. T. (2012). Information Foraging on the Internet. In P. M. Todd, T. T. Hills, and T. W. Robbins (Eds.), *Cognitive Search: Evolution, Algorithms, and the Brain* (pp. 283-299). Cambridge, MA: MIT Press.

[37] McCart, J. A. (2009). *Goal Attainment On Long Tail Web Sites: An Information Foraging Approach* (Doctoral dissertation, University of South Florida).

[38] Pirolli, P. (2006k). The use of proximal information scent to forage for distal content on the World Wide Web. In A. Kirlik (Ed.), *Working with Technology in Mind: Brunswikian Resources for Cognitive Science and Engineering* (pp. 247-266). New York, NY: Oxford University Press.

[39] Eiron, N., & McCurley, K. S. (2003, May). *Locality, hierarchy, and bidirectionality in the Web.* Paper presented at the Workshop on Algorithms and Models for the Web Graph, WAW 2003, Budapest, Hungary.

[40] Davison, B.D. 2000. Topical locality in the Web. Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Athens, Greece, 272–279.

[41] Manning, C. D., & Schuetze, H. (1999). Foundations of statistical natural language processing. Cambridge, MA: MIT Press.

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

[43] Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary Task Interleaving: Heuristics for Time Allocation in Cognitive Foraging. *Journal of Experimental Psychology: General, 136*, 370-388.

[44] Green, R. F. (1984). Stopping rules for optimal foragers. *American Naturalist, 123* 30-43.

[45] Duggan, G. B., & Payne, S. J. (2009). Text Skimming: The Process and Effectiveness of Foraging Through Text Under Time Pressure. *Journal of Experimental Psychology: Applied, 15*, 228-242.

[46] Simon, H. A. (1962). *The architecture of complexity.* Paper presented at the Proceedings of the American Philosophical Society, volume 106 (pp. 467–482).

[47] Card, S. K., Pirolli, P., Van Der Wege, M., Morrison, J. B., Reeder, R. W., Schraedley, P. K., & Boshart, J. (2001, March). Information scent as a driver of Web behavior graphs: results of a protocol analysis method for Web usability. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 498-505). ACM.

[48] Budiu, R., Royer, C., & Pirolli, P. (2007, May). Modeling information scent: A comparison of LSA, PMI and GLSA similarity measures on common tests and corpora. In *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)* (pp. 314-332). Le Centre de Hautes Estudes Internationales d`informatique documentaire.

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

[50] Pirolli, P., & Fu, W. (2003). SNIF-ACT: A model of information foraging on the World Wide Web. In P. Brusilovsky, A. Corbett, & F. de Rosis (Eds.), *User Modeling 2003, 9th International Conference, UM 2003* (Vol. 2702, pp. 45-54). Johnston, PA: Springer-Verlag.

[51] Clark, R. E., Feldon, D., Van Merrienboer, J., Yates, K., & Early, S. (2008). Cognitive task analysis. *Handbook of research on educational communications and technology*, *3*, 577-593.

[52] Naikar, N. (2013). Work domain analysis: Concepts, guidelines, and cases. CRC Press.

[53] Vicente, K. J. (1999). Cognitive work analysis: Toward safe, productive, and healthy computer-based work. CRC Press.

[54] Hutchins, S. G., Pirolli, P. L., & Card, S. K. (2004). *A new perspective on use of the critical decision method with intelligence analysts*. 2004 Command and Control Research and Technology Symposium.

[55] Morse, E., Steves, M. P., & Scholtz, J. (2005, May). Metrics and methodologies for evaluating technologies for intelligence analysts. In *Proc. Conference on Intelligence Analysis*.

[56] Elm, W., Potter, S., Tittle, J., Woods, D., Grossman, J., & Patterson, E. (2005, September). Finding decision support requirements for effective intelligence analysis tools. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 3, pp. 297-301). SAGE Publications.

[57] Taol, C-C., & Bucy, E. P. (2007). Conceptualizing Media Stimuli in Experimental Research: Psychological Versus Attribute-Based Definitions. Human Communication Research, 33, 397-426.

[58] Borlund, P., & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. The Journal of Documentation, 53, 225-250.

[59] Hammell, R. J., Hanratty, T., & Heilman, E. (2012, June). Capturing the value of information in complex military environments: A fuzzy-based approach. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on* (pp. 1-7). IEEE.

[60] Newcomb, E. A., & Hammell II, R. J. (2012). Examining the Effects of the Value of Information on Intelligence Analyst Performance. In *Proceedings of the Conference on Information Systems Applied Research ISSN* (Vol. 2167, p. 1508).

[61] Hanratty, T., Heilman, E., Dumer, J., & Hammell II, R. J. (2012, April). Knowledge Elicitation to Prototype the Value of Information. In *Midwest Artificial Intelligence and Cognitive Science Conference* (p. 173).

[62] Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001). *Using information scent to model user information needs and actions on the web*. SIGCHI'01, Seattle, WA.

[63] Chi, E. H., Pirolli, P., & Pitkow, J. (2000, April). The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 161-168). ACM.

[64] Chi, E. H., Rosien, A., & Heer, J. (2002, July). *LumberJack: Intelligent discovery and analysis of Web user traffic composition*. Paper presented at the ACM-SIGKIDD Workshop on Web mining for usage patterns and user profiles, WebKDD, 2002, Edmonton, Canada.

[65] Hoare, C., & Sorensen, H. (2005). Enhancing Information Retrieval Interfaces with Information Foraging. *AICS'05*, 319.

[66] Liu, D., Raja, A., & Vaidyanath, J. (2007, November). Tibor: A resource-bounded information foraging agent for visual analytics. In *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology* (pp. 349-355). IEEE Computer Society.

[67] Cutrell, E., & Guan, Z. (2007, April). What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 407-416). ACM.

[68] Olston, C., & Chi, E. H. (2003). ScentTrails: Integrating browsing and searching on the Web. *ACM Transactions on Computer-Human Interaction, 10*, 177-197.

[69] Wexelblat, A., & Maes, P. (1999, May). Footprints: history-rich tools for information foraging. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 270-277). ACM.

[70] Shrinivasan, Y. B., & van Wijk, J. J. (2008, April). Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1237-1246). ACM.

List of symbols/abbreviations/acronyms/initialisms

CAF	Canadian Armed Forces
CTA	Cognitive Task Analysis
f	False Alarm
FAM	Fuzzy Associative Memory
GEOINT	Geospatial Intelligence
GLSA	Generalized Latent Semantic Analysis
HUMINT	Human Intelligence
IFT	Information Foraging Theory
LSA	Latent Semantic Similarity
MVT	Marginal Value Theorem
OFT	Optimal Foraging Theory
OSINT	Open-Source Intelligence
PMI	Pointwise Manual Information
RUM	Random Utility Model
SA	Situation Awareness
SIGINT	Signal Intelligence
SNIF-ACT	Scent-based Navigation and Information Foraging in the ACT architecture
TIBOR	Time Bounded Reasoning
U.S.	United States
URL	Uniform Resource Locator
VoI	Value of Information
WBG	Web Behaviour Graph
WDA	Work Domain Analysis
WUFIS	Web User Flow by Information Scent
WWW	World Wide Web

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security markings for the title, abstract and indexing annotation must be entered when the document is Classified or Designated)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g., Centre sponsoring a contractor's report, or tasking agency, are entered in Section 8.) Defence Research and Development Canada – Toronto 1133 Sheppard Avenue West P.O. Box 2000 Toronto, Ontario M3M 3B9	2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.) UNCLASSIFIED 2b. CONTROLLED GOODS (NON-CONTROLLED GOODS) DMC A REVIEW: GCEC DECEMBER 2012

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Information foraging theory: A framework for intelligence analysis

4. AUTHORS (last name, followed by initials – ranks, titles, etc., not to be used)

Bryant, D. J.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
November 2014	56	70

7. DESCRIPTIVE NOTES (The category of the document, e.g., technical report, technical note or memorandum. If appropriate, enter the type of report, e.g., interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Scientific Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence Research and Development Canada – Toronto
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC-RDDC-2014-R115	

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Information Foraging Theory (IFT) is proposed as a framework in which model information search in the military intelligence analysis domain. Information Foraging Theory explains human information search and exploitation as adaptations to the informational structure of the environment and has been used to model peoples' preferences for information types, rules for exploiting discrete information sources, and the use of semantic cues to enhance the search process. A plan for the application of Information Foraging Theory to the military intelligence domain is described, beginning with the process of describing the task environment in which analysts work and moving to the issue of defining key Information Foraging Theory concepts in that environment. The report ends with a discussion of ways application of IFT may benefit military intelligence analysis, such as automated goal analysis and parameter tracking, enhancing information scent cues, and information visualisation techniques.

La théorie du butinage des renseignements (TBR) est proposée comme cadre et modèle de recherche des renseignements dans le domaine de l'analyse du renseignement militaire. La théorie du butinage des renseignements explique la recherche et l'exploitation humaine des renseignements comme des adaptations à la structure informative de l'environnement et a été utilisée pour modéliser les préférences des individus selon les types de renseignements, les règles pour l'exploitation de sources de renseignements discrètes, et l'utilisation d'indices sémantiques pour améliorer le processus de recherche. Un plan pour l'application de la théorie du butinage des renseignements dans le domaine du renseignement militaire est décrit, en commençant par le processus de description de l'environnement de travail dans lequel les analystes évoluent et abordant la question de la définition des concepts clés de la théorie du butinage des renseignements dans cet environnement. Le rapport se termine par une discussion sur les façons dont la TBR peut bénéficié l'analyse du renseignement militaire, comme l'analyse automatisée des objectifs et le suivi des paramètres, l'amélioration des renseignements des signaux olfactifs et les techniques de visualisation des renseignements.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g., Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Information foraging; intelligence analysis; information search