Int. J. Man–Machine Studies (1990) **32**, 511–543

User models: theory, method, and practice

ROBERT B. ALLEN
2A-367, Bellcore, Morristown, NJ 07960–1910, USA

(Received 28 May 1987 and in revised form 20 March 1989)

While the technology of new information services is rapidly advancing, it is not clear how this technology can be best adapted to people's needs and interests. One possibility is that *user models* may select and filter information sources for readers. This paper examines the prospects and implications of automatic filtering of information, and focuses on predicting preferences for news articles presented electronically. The results suggest that the prediction of preferences can be straightforward when general categories for news articles are used; however, prediction for specific news reports is much more difficult. In addition, an effort is made to establish a systematic study of the effectiveness of information interfaces and user models. Fundamental issues are raised such as techniques for evaluating user models, their essential components, their relationship to information retrieval models, and the limits of using them to predict user behavior at various levels of granularity. For instance, prediction and evaluation methodology may be adopted from personality psychology. Finally, several directions for research are discussed such as treating news as hypertext and integration of news with other information sources.

1. Information interfaces, user models, and news articles

Information systems have been designed to present electronic textbooks (Weyer, 1982; Yankelovich, Meyrowitz & vanDam, 1985), dictionaries, encyclopedias (Koved & Shneiderman, 1986; Weyer & Borning, 1985), newspapers (Moghdam, 1978; Patten, 1986), news-related services (Lesk, 1982; Lippman & Bender, 1987; Stanfill & Kahle, 1986), and correspondence (Malone, Grant & Turbank, 1986). Ease of update, search, and controlled presentation of background material and details are among the advantages to electronic presentation. However because there can be so much detail in each electronic information source, the information presented to the reader may be of lower quality and less relevant than with traditional approaches. Therefore it seems natural to apply information processing power to the task of selecting items of interest and relevance. Indeed, the ability to select items for users may be essential to the viability of some services. This individualization may be provided by *user models.*

Thus, the research reported here has several goals. On one hand, the nature of news and its use is examined; on the other hand, issues such as strategies for evaluating information interfaces, how information is used by people, and the limitations of user models are considered. Many lines of research touch on the issue of providing effective user models and one goal of this paper is to compare ideas from these approaches. Although a relatively large number of information interfaces have been developed, almost no systematic studies have examined the use of these interfaces. Thus, the emphasis is on the user rather than on the interface, documents, or specific inference processes.

0020-7373/90/050511 + 33$03.00/0

News articles were chosen as a domain for studying information preferences and user models of those preferences. News has the advantage of usually being of interest to most readers. Beyond its inherent interest, news is an especially challenging type of material about which to make predictions. Moreover, news reading may have extremely varied objectives (Dozier & Rice, 1984), and both the reader and the subject matter are continually changing (see Section 3.1). On the other hand, the systematic study of news presents difficulties; for instance it is impossible to replicate the novelty of a news item and difficult to control exposure to competing information sources.

News may be defined as the descriptions of recent events that are new to an individual and also of some general interest. In this broad conception, much of human communication may be thought of as the transmission of news. However, because of the subjectivity of terms such as *new* and *general interest,* an operational definition of news is needed for the present research. The simplest operation definition of news is to define it as the material in a newspaper. However, few people would agree that news makes up all of the contents of a newspaper. Obviously, a newspaper also includes advertisements, background stories, obituaries, editorials, puzzles, and comics. A second operational definition would restrict news to what is present in conventional news articles such as those available from a news wire service. The first, more general definition is used (Section 4) to gain an overview of the issues involving news access, while the more restricted definition is used in studies of the effectiveness of fine-grained user models (Sections 5, 6).

2. User models

User models have been considered for applications as diverse as education, software engineering, help systems (Mason, 1986), expert systems (Sparck Jones, 1986), and natural language processing. Indeed, predicting the behavior of other agents is an important activity in the field of distributed artificial intelligence.

The term "user model" is employed in several senses. A user model may be a specific module with user knowledge which is part of an interface, or it may refer to the style of interaction that an interface engenders without reference to specific user knowledge. Indeed, the term may also be applied to the model that a user has of a computer program, although that usage is not relevant to this research. The terms *person model* or *individual model* may be more appropriate when the interaction does not involve obvious manipulation of a computer interface, e.g. the model a natural language interface might have of a person. Similarly, the term *agent model* may be more appropriate when a computer agent has a model of the capabilities, needs, and interests of other agents.

It is often difficult to determine the boundaries of a user model. In part, this is because of the difficulties of the definition. It might be argued that all programs have an implicit user model, but not necessarily an adaptive model. Indeed programs may be developed that serve one user's needs in only one situation, rather than a integrated more general program. Further, user models, especially those which involve complex inferences, may be seen as a collection of models rather than a single model. Given these problems, a functional definition is adopted. In this paper

a user model is the knowledge and inference mechanism which differentiates the interaction across individuals.

There are several scenarios in which a user model might be applicable. One is when an individual is not familiar with the necessary details needed to perform a task and needs advice. Another would be to save time. For instance, although a person could look through all possible news articles for a given day, it is obviously easier to delegate the selection to another human or computer agent.

While the term "user model" emphasizes the information about the person, it is obvious that a great deal of *situational, task,* or *environmental* information may be encoded in the model. Thus, just as situation and task have been recognized to be essential in recent personality models (e.g. Mischel, 1968; Snyder & Ickes, 1985), sometimes computerized user models might best be thought of as *user × situation models.* For instance, a help system might depend entirely on a classification of the user's current state and not on the user's demographics or the history of how the user got into that state. There may be a trade-off in the predictive utility of these components, since some situations are *strong* and have clear contingencies, while others are ill-defined (Snyder & Ickes, 1985).

2.1. RICH'S TAXONOMY OF USER MODELS

Rich (1979, 1983) has proposed a taxonomy of user models with three dimensions. First, the short-term/long-term dimension has to do with the collection and persistence of user information over time. Second, the explicit/implicit dimension concerns whether the user specifies the model directly or extracts it from ongoing user behavior. Finally, the individual/group dimension concerns whether separate models are developed for each user, or a single model is applied to everyone. While these seem to be important dimensions, it is often not clear which aspects of the models they refer to. For instance, a model may be long or short-term based on the history of interaction that it keeps about an individual, or in terms of the duration of any one interaction it modeled. Moreover, it is possible to identify a wide variety of other characteristics for user models. For instance, the short-term/long-term dimension suggests a possible distinction in predictions in the same situation versus across situations. Other dimensions might be the type of representation, the style or tone of the user interaction or the type of information modeled (user knowledge, user preferences, etc.); and other authors (e.g. Daniels, 1986; Kass & Finin, 1986) have suggested many other dimensions.

2.2. MODEL COMPONENTS

Rather than developing additional taxonomies, the approach here is to consider the structure of the models and the flow of information in user models. Thus, this section will consider independent variables (Section 2.2.1), representation (Section 2.2.2), and responses (Section 2.2.3). Since the models adapt across individuals and across time, feedback is also examined (Section 2.2.4). This approach also highlights the differences and similarities between various user models, simple prediction models, and models based on classical control theory; it is also similar to the three-layer network architectures of neural networks (Rumelhart, Hinton & Williams, 1986).

2.2.1. Independent variables

Some independent variables for user models locate a particular user in the space of possible users. For instance, demographic information about a user, such as responses to questions and recent behavior, may be employed to characterize the user, the user's goals, or the user's situation. Other types of input to the models must also be considered as independent variables. For instance, material which is filtered for presentation to a user would be a type of independent variable. In the language of control systems this material may be termed "controlled input". In addition, feedback may be considered a special type of input (Section 2.2.4).

The selection of good predictors for a given dependent variable is a major part of the design of user models. While general user dispositions (e.g. personality traits) might be considered effective user information, in fact, a great deal of research has shown that they are poor predictors of specific behaviors. However, as noted above there are many other types of information such as population demographics, norms, task demands, or statistical summaries which may be effective predictors of user behavior (see Section 2.5). Prediction effectiveness must be weighted against other factors such as the cost of gathering and storing information, the number of situations in which that information might be useful, and the penalty for making an error. For instance, conversation among people is generally successful despite misunderstandings, and in many cases, there are robust techniques for backup in a conversation.

2.2.2. Cassification and representation

As noted by Clancey (1984) user modeling is closely related to the process of classification. However, the more general view proposed here is that there is a continuum from classification through inference and that classification may be applied into fuzzily defined or even dynamic categories.

While the ultimate task for a user model is the prediction of behavior, attempting to identify a user's goals or *plan recognition* (Schmidt, Schridharan & Goodson, 1978) is a common approach in user modeling. Goals may be considered a type of classification, and in highly structured situations it is often easy to infer the user's goal from among a few alternatives. On the other hand, *stereotypes* in GRUNDY (Section 2.6) are a type of representation which involves classification without explicit goals. However, the tendency to attribute actions to goals is sometimes compelling and may lead to an effort to identify goals even when they are obscured, if present at all (Carroll, 1987).

The essential issue for representation is capturing input in a way that maintains salient information while minimizing unnecessary information. Many types of representations for user and situation knowledge are possible; for instance the representation might be frame-based (Bobrow, 1975), or statistical, such as neural networks (Allen, 1989; Rumelhart, Hinton & Williams, 1986). Representations are often subdivided into modules; for instance tutoring systems are often said to include student models, instruction models, and response models. Both the student model and instruction model are representations, while response models generate output (see Section 2.2.3). Representations may also be distinguished to the extent that general world-knowledge is incorporated, how the representation adapts due to feedback (Section 2.2.4), and the extent to which the classification dimensions are

meaningful; for instance whether specific psychological or task states are explicitly modeled.

2.2.3. Responses

A response to the user is generated for the user from the representation by the *response model.* This process is conceptually similar to using a personality classification to predict a person's behavior or proper clinical treatment (see Section 2.5). In some cases, the response model might be a direct function of the classification of the user or the user's state. In other cases, it might be modulated by additional user information. Indeed in complex models, a generate-and-test cycle might be employed in which the effects of certain interventions could be simulated and the best one selected.

2.2.4. Feedback

Not all user models are adaptive; however for those models which are adaptive, feedback is essential, and the models show a type of machine learning. A system may adapt within a session, with experience on a given individual, or across individuals. However, some care must be used because "adaptation" may be movements in space and not permanent changes (just generalization, Section 2.3.3). Even user models which are not adaptive in the narrow sense may be said to have feedback; specifically the designer of a system may tune the model to improve performance.

The choice of feedback involves many of the same issues as the selection of independent variables. Moreover, the interpretation of the *feedback variables* and the modification of the model based on feedback (i.e. *credit assignment*) are crucial design decisions. In some models inference about the feedback is rule based, and in other systems statistical measures are used. As in most control theory, feedback indicates a deviation from a target, such as expert performance.

2.3. EVALUATION OF INFORMATION INTERFACES AND USER MODELS

As mentioned earlier, a large number of information interfaces exist; however there is almost no data on their effectiveness. Obviously some systematic evaluation strategy is necessary, and often even simple data can be very useful. However because the dependent variables of the user model are usually not objective of the system, the evaluation must be indirect. Fortunately, there are many options in evaluation strategies (Allen, submitted). Of course, many aspects of a person's behavior may be modeled. Naturally, some of these are relatively straightforward, while others are complex. For example, it is easy to predict that a person will talk on the telephone when working in an office, but it would be much more difficult to predict word-for-word what will be said. Similarly, in some cases it may be simple to generate an effective response but very difficult in other cases.

2.3.1. Specifying parameters, simple tests, and in vivo *performance*

Many descriptions of user modeling in the literature lack enough detail to reproduce and study the performance of the model. Clearly, the descriptions of user models should specify the independent variables, the system architecture, and inference

procedure in detail. Of course both the training and testing conditions must be defined, for adaptive user models. Probably the ideal testing conditions would be to examine performance in a variety of natural field conditions and for a variety of individuals. However, this is generally impractical for testing a large number of variations. Thus simple testing conditions need to be devised for both scope and utility. Specific measures might include appropriateness, level of intelligibility, and completeness. Moreover there could be integration measures such as overall satisfaction.

2.3.2. Reliability and validity

Computerized user models may be related to personality and educational assessment tools such as tests of individual differences. Thus, for evaluation it seems reasonable to consider standard assessment criteria such as *reliability* and *validity* of the assessment technique. For instance, reliability would be indicated by the consistency by which different samples of a person's behavior produced the same classification. In the case of the news article prediction, we might ask whether similar news stories were classified similarly and at the model level whether similar terms generated similar models.

To the extent that a model involves categorization (Section 2.2.2), then the validity of the categories may be considered. Validity refers to the quality of the categories that are used. The ideal classification would show *convergent* validation from other behavior and *divergent* validation from other constructs. For instance, the classification of a person as a naive user ideally should be correlated with understanding of command languages and orthogonal to other measures such as mathematical ability. The optimal model would use the simplest set of independent variables and inference mechanisms to generate an effective classification. This property is termed *incremental validity* (Mischel, 1968) in personality tests. For instance, paper-and-pencil personality tests might be compared against simple prediction strategies such as past behavior, self-reports, social norms, and demographic data. Essentially this approach is pursued in the analyses in Section 8.

Of course the objective of personality classification or classification in user modeling is not the classification itself but the use of that classification to predict future behavior and effective treatments. Thus the *utility* (Mischel, 1968) of the assessment methodology, which includes user models, may be measured in addition to its validity and reliability. This implies evaluation of the classification in conjunction with performance measures.

2.3.3. Generalization across time, situation, and task

At some point validation becomes too dynamic to track and a model may become a self-adaptive learning system. As more experience is gained the model should perform better and testing could examine the evolution of the model through time. In some cases there is adaptation both "within" and "across" users and both of these may be evaluated. Evaluation of performance must then be measured by the quality of *transfer* to new tasks.

2.3.4. Turing controls

While the previous sections considered comparing performance between conditions, a different standard for performance would be to compare performance (of a

computerized user model or adaptive interface) with a human being engaged in a similar type of task. Because of the similarity of this type of comparison to the Turing test (Turing, 1950), it may be called a *Turing Control* condition. Although the Turing test is open ended, and was conceived as a measure of overall "intelligence" this control condition applies only to a limited aspect of behavior. Indeed, it seems possible that the model might perform better than the human being; thus the control condition provides a reference point and not an absolute target. There are several variations of this procedure; for instance, control subjects with different backgrounds may be selected, or friends of the subject (see Section 6.3) might provide special insight. On the other hand, friends might also have biases, and there are potential methodological problems in any procedure which involves non-random selection of subjects.

2.4. INFORMATION RETRIEVAL AND PREDICTION

Information retrieval (IR) typically involves obtaining documents which are relevant to rather specific user queries (e.g. Salton & McGill, 1983; VanRijsbergen, 1979). Thus, IR is closely related to the problem of user modeling with the primary emphasis on classification of documents rather than individuals or their states. Of course IR most often focuses on cases in which the user queries are extremely brief, the documents are relatively lengthy, and the document collection extremely large. Increasingly, however, IR is being extended to cases which violate these traditional limitations.

Moreover, many suggestions make the IR approaches close to user models. For instance, feedback on the relevance of a document to a user's needs has been found to be useful (e.g. Croft & Thompson, 1984). In addition, Selective Dissemination of Information (SDI) systems (Luhn, 1958) match a stored user query against new documents as they arrive. Typically SDI systems are keyword driven, and much recent work has emphasized enhancements to these keyword interfaces (e.g. Leggate, 1975; Oddy, 1977). The information Lens (Malone, 1987) is especially close to SDI systems in that stored patterns may be matched to incoming messages. Of course, these approaches still emphasize generalization across documents, rather than across characteristics of individuals. However some systems have been proposed which incorporate user knowledge with information retrieval interfaces (e.g. Brajnik, Guida & Tasso, 1987; Brooks, Daniels & Belkin, 1985) to aid in the disambiguation of queries and the generation of responses.

2.5. PSYCHOLOGY AND PREDICTION

The categorization and prediction of human behavior is one of the fundamental tasks of psychology. For instance, operant learning theory attempts to describe how behavior is controlled by reinforcement. Likewise, cognitive psychology makes predictions about forgetting and information processing which may determine later behavior. However, the fields that are most often associated with the prediction of human behavior are social and personality psychology. In social psychology attitudes are often measured to predict behaviors such as voting. While personality theory is perhaps best known for psychodynamic models, many other approaches such as trait models have been considered. Moreover, personality models have a well-developed methodology for evaluating the effectiveness of predictions.

As noted by Rich, there is a similarity in predictions of future behavior and the "implicit personality judgements" made on the basis of the informal impressions of observers (Section 2.6). However, there is also a strong link between social/personality psychology and user models seen in the possibility of applying formal personality assessment methodology (see Sections 2.4.2, 8) to user modeling. For instance, the individual/group dimension (Section 2.1) also may be to be related to the traditional distinction from personality psychology between nomethetic (i.e. generally across individuals) or idiographic (i.e. tailored to the individual) models. Of course, even formal models are only moderately successful as predictors of human behavior, just as attitudes sometimes do not correlate with behavior (Fishbein & Ajzen, 1975; Ajzen & Fishbein, 1977). Many of the limitations of these predictions are understood. For instance, better prediction may sometimes be accomplished by using other independent variables such as demographics and social norms (Mischel, 1968) (see Sections 2.4.2, 8). Moreover, as with recent personality research (Bowers, 1973; Snyder & Ickes, 1985), prediction in user models would probably benefit from incorporating situational variables.

Classification/treatment models have also been proposed in educational psychology. For instance, *aptitude-by-treatment interaction* (ATI) models suggest that individuals may have differentiated aptitudes and that effective treatments could then be applied to them. However this has not proven an effective model with traditional aptitude tests (Cronbach & Snow, 1977).

2.6. STEREOTYPES AND GRUNDY

Rich (1979, 1983) has proposed the use of "stereotypes" for user modeling and applied them in the GRUNDY system which suggests books to people they would like to read. By stereotypes she means a cluster of characteristics which are likely to occur together, and the task of the user model is to assign or classify the user to these clusters. Further, the classification is updated with each new input, much like the inferences made in social cognition. Other approaches to stereotypes might use a more efficient classification technique; for instance, the entire history could be reconsidered when each new inference was required.

The independent variables for GRUNDY are self-descriptive adjectives and the abstracts which are considered. Although the stereotypic reasoning in GRUNDY seems to be based on an analogy to social cognition, there is no attempt to demonstrate the psychological validity of the ways in which the stereotypes were employed. Moreover, data on the performance of GRUNDY are sketchy. It was shown to perform better than chance, averaged across a large number of its predictions; however, there was no test of whether the adaptive mechanism improved performance. Some control conditions for GRUNDY are reported in Section 8.

2.7. PROS AND CONS OF COMPUTERIZED USER MODELS

Would people really want a personal assistant such as a user model for selecting news articles? Indeed, there are several objections that may be raised about the application of user models.

2.7.1. Perceived freedom

People may have a bias against being controlled (Christie, 1981) or at least, against the belief that they are being controlled. While it remains to be clearly demon-

strated, this could be a problem for user models. On the other hand people may be more willing to accept models which are simple enough that they can be easily controlled. These could be called *transparent user models.* For instance, although keyword models may not be as effective as other techniques complexes, they may be preferrred to complex user models in which the user is unable to understand or control the action of the model.

Still another problem related to perceived freedom is that people may feel self-conscious about the articles selected based on their interests, if for instance the user model frequently suggested sensational news articles and the reader had objections to that type of material. Conceivably, this could cause a rejection of the news articles and of the system itself. On the other hand, in some cases interaction with a computer has been found preferable to interaction with a human being. Specifically Lucas (1977) reports that a computer interview for hospital admittance may be preferred to an interview with an overworked human being.

2.7.2. Serendipity and diversity

Another concern is that user models may be too mechanistic, potentially too narrow, and readers might miss the advantage of serendipitously useful information. For instance, many people report that a seemingly irrelevant news item has become of great importance sometime after they read it and remembered it. A related argument concerns the importance of exposure to a diversity of opinions about a topic. However the severity of these concerns seems difficult to predict without functioning systems. Indeed, concrete evidence for the occurrence of serendipity is weak. The first step would be to determine how often serendipity occurs and under what cirumstances. On the other hand, it might easily be argued that exposure to articles selected by some criterion produce more "serendipity" than randomly selected articles.

2.7.3. Privacy and security

User models may contain a great deal of personal information. If this could become available to others, there would be a serious concern for privacy (e.g. Parker, 1979; Maurer, Rozsenich & Sebestyen, 1984). Moreover, much more complex scenarios are possible (see Section 9.2.6) in which the user model is a responsive agent, and these pose even greater complications for privacy and security. These are substantial concerns and they must be dealt with; whether effective solutions can be found remains an open and important technical question. However, such tests are costly and it is often difficult to find stable testing conditions. Moreover, even small changes in a large program can require large scale tests. Thus, it often makes sense to consider the performance of components separately or with subjects in some controlled testing environment.

3. Predicting new preferences

3.1. THE CHALLENGE OF NEWS ARTICLES

While the previous section describes issues of prediction in general, Sections 3–8 focus on prediction of preferences for news articles. News is undoubtedly the most widely accessed type of non-fiction material; however, the prediction of preferences

for news articles is especially challenging. First, the document collection is continually changing. Thus, unlike most ordinary document retrieval tasks, the retrieval is always about different documents than those already in the archive. Indeed, there may be a penalty for retrieving documents similar to those previously accessed. Moreover, the archive includes many unusual terms (e.g. proper names), and it often includes large numbers of both overlapping and very different articles.

In addition, the reader's preferences are probably complex and frequently changing. In some cases a reader may want to know all available details about a news item; in other cases only a broad overview of news, and in still other cases news may be read primarily for entertainment. This parameter may be termed the *scope* of the user's interest. For instance, in some areas the reader may want only specific questions answered (e.g. stock quotations), in other areas quick summaries of news, or a great deal of background information (perhaps with only the very tangential or redundant information removed). Finally, there are methodological complications in studying news articles. For instance, many news articles are rapidly outdated; thus it is difficult later to recover the original impact a given item may have had.

3.2. MOTIVATIONAL ANALYSIS: MEDIA USES AND GRATIFICATIONS

If some clear motives for media access could be determined, perhaps then news articles could be chosen which matched those motives (Section 2.3.2). Indeed in the field of media and communication, many suggestions have been made for motives of media access. Among the main factors which have been proposed are goal satisfaction (Brajnik, Guida & Tasso, 1987), the social importance of media (Stepenson, 1967), and entertainment and play (Katz, Blumler & Gurevitch, 1974). Unfortunately the analysis of uses and gratifications of media has not led to clear predictions. Thus, it is not clearly useful in predicting information preferences. However based on this literature, Dozier and Rice (1984) noted that early interfaces to electronic newspapers were highly structured and might be more acceptable if they supported browsing, as would be indicated by a play theory of news reading.

3.3. INTEREST AS PREDICTOR OF NEWS PREFERENCES

A common explanation that people give for reading the newspaper is that they are *interested* in its contents. Interest has been proposed as a critical variable in determining the processing of sentences (Schank, 1979) and in story comprehension (Anderson, Shirley, Wilson & Fielding, 1987; Hidi & Baird, 1986). While "interest" is an intuitively appealing explanation, it immediately raises the problem of assessing what is interesting to a person. Schank has proposed that certain topics such as sex and violence and factors such as personal relevance make text interesting. However, no data are presented to support this claim. Even if interestingness is correlated with the likelihood of reading a newspaper, it is reasonable to ask whether it is a causal or correlated factor. Thus, it is possible to argue that the fact of being interesting is merely an *attribution* which people apply afterwards to justify their choices (Allen, submitted).

4. Informal studies of traditional newspaper content and access

Two brief studies considered the content and uses of a typical newspaper. Both of these studies are informal and have methodological flaws; however, they may be useful for a general orientation to news.

4.1. WHAT'S IN A NEWSPAPER

Examination of the front section of a large urban/suburban newspaper† showed that there were about 23 news articles of which two had two pictures each and four had a single picture. These articles averaged 586 words and they covered about 26% of the surface. About 5% of the area was pictures, and most of the remainder was advertisements. Almost none of the pictures presented information that was essential to the news story; rather the pictures provided a context or emotional background for the news. The only table or chart was the weather map.

TABLE 1
Informal classification of temporal relations in sentences in news text

Time relation	Proportion of sentences
Future	0·14
Previous 24 h	0·49
Previous week	0·04
Earlier than previous week	0·15
Indeterminate past	0·07
No temporal information	0·10

In Table 1 sentences from the front page of a newspaper on two different days are classified by the temporal information they convey. Not surprisingly, news articles are concerned with recent events. However, approximately half of the content is background. as well as notices of events to come. The sentences were also classified on another dimension, as shown in Table 2, which is roughly the illocutionary force (Searle, 1969) of the sentences. It appears that the largest number of sentences contributed background factual knowledge. Another interesting aspect of the writing in news articles is the style. For instance, it frequently appears to violate the "given-new contract" (Haviland & Clark, 1974). While most writing is characterized by stating the known information (given) and then moving to new information, news writing frequently moves from the new information (news) to background material.

† The *Newark Star-Ledger* was used for the studies involving paper copies of a newspaper. This is the largest circulation newspaper in New Jersey and includes articles from several major news wires, including AP. Of course other types of newspapers may have substantially different distributions of pictures, graphics, and different writing styles. The data for the various conditions were collected from June 1985 to March 1986. Except as noted, data were not collected during periods in which extraordinary news stories were reported. For instance, no data were collected at the time of the explosion of the Shuttle Challenger.

TABLE 2
Informal classification of news text

Content	Proportion of sentences
Gives general world knowledge	0·43
Reports opinion or quotation	0·26
Gives background specific to event	0·23
States conflict	0·09

4.2. HOW PEOPLE READ NEWSPAPERS

Some general questions may be addressed simply by watching people read the newspaper, for instance, whether people are consistent about the sections they read. In Section 5, detailed records are kept on which news articles people read; in this section some general aspects of how people read conventional newspapers are studied.

4.2.1. Procedure

Five subjects‡ from the local community were observed while they read the newspaper on Monday and Wednesday of the same week. These subjects were asked to spend about 15 min reading the newspaper and they were told that although the Experimenter would be making notes about what they did, they were not being evaluated and should try to be as natural as possible. The Experimenter noted how long the subjects spent reading each article.

4.2.2. Results and discussion

Four of the five subjects started on the first page of the paper and read through to the last page. As noted in other media research the physical structure of a newspaper is a strong determinant of what is accessed. The Experimenter categorized the articles into 13 groups. The categories, the mean time the subjects spent reading articles in each category for each of the two days, and the correlations across the five subjects are shown in Table 3. The average percentage of time spent on each category stayed about the same across the two observations with the exception of the time spent on advertisements and on the entertainment pages. Many supermarkets advertise specials on Wednesday and the subjects spent considerable time with those. The large positive correlations suggest that across almost all categories there was consistency in the choices of individual subjects as compared to the variability across subjects. The large correlations for less frequent categories (e.g. weather) suggest that only a few subjects read those, but those subjects were consistent for the two days. For instance, only one subject read the obituaries.

With the two exceptions noted above, subjects seemed consisent in the amount of

‡ Throughout these studies two types of subjects participated. One was recruited from the local community; generally, these subjects were female college graduates and they were paid $\$5\,h^{-1}$ plus transportation expenses for their participation. A second set was composed of volunteers from the Bellcore staff, and several of these subjects had advanced degrees. Both types of subjects seemed frequently to read the newspaper for pleasure. Except as noted, no subjects participated in more than one study.

TABLE 3
Time spent on several news categories across 2 days

Category	Proportion of Session Monday	Wednesday	r
National/international news	0·20	0·14	0·68
State/local news	0·16	0·14	0·90
Advice columns	0·11	0·11	0·95
Business/finance	0·10	0·03	0·00
Editorials/columnists/letters	0·09	0·11	0·87
Advertisements	0·06	0·17	0·63
Obituaries	0·05	0·03	1·00
Gossip	0·04	0·07	0·25
Sports	0·04	0·02	0·76
Comics	0·04	0·02	0·41
Entertainment	0·03	0·10	0·40
Bridge	0·02	0·01	1·00
Weather	0·01	0·01	1·00

Note: The correlations were calculated across the five subjects. The categories used here seemed reasonable, but they were not validated; however, the effects are so large that even some subjectivity would probably not affect the conclusions.

time they spent on general categories of the news. It appears easy to make predictions about subjects' behavior at this level of granularity. Moreover, even the two categories which showed substantial shifts might show similar shifts in later weeks. In addition to the implications of the data in Table 3 for constructing user models of news preferences, they also provide evidence for a micro-economics of information usage (Ferguson, 1972; King, Roder & Olsen, 1983), and of how subjects allocate their time to read news stories. In particular, the subjects appeared to substitute some categories between Monday and Wednesday while the others remained relatively constant.

5. Choices for specific news stories

Although the previous studies suggest that it should be possible to predict the general categories of news a person will read, they do not show whether it is possible to predict specific news stories that the person would read. In this study, subjects used a limited electronic newspaper which generated detailed records about which stories they read. The procedure and basic results are presented in this section, several control conditions are described in Section 6, and an extended model is described in Section 7.

5.1. PROCEDURE

5.1.1. News source: the AP news wire services

The Associated Press (AP) news service (French, Powell & Angione, 1980; Newspaper Center, 1984) distributes articles at several grades of service. The basic

service used in these studies was a portion of the "A" service which includes news articles, human interest stories, and a few regular features. Limiting the research to only parts of the AP wire makes the prediction task difficult because it eliminates classifications of news stories, such as sports and financial, for which user preferences might be readily determined (Section 4). The "A" wire transmits about 300 stories per day; however this was reduced to about 200 stories by a pre-processor, which deleted story revisions (write-thru's), financial news, notes to the editors, and so forth. Each article included a header which was five to 15 words in length (see Figure 2 for examples) and the articles averaged 2537 words.

5.1.2. Procedure

Five subjects, who were employees of Bellcore, volunteered for an extended study of reading news. Each subject completed 11 two-session cycles. On the first day the subjects chose and read articles from the electronic display as described below. On the alternate days the subjects were asked to make ratings of how interested they would be in reading 15 news articles. The sessions were completed only on days when both the subjects and Experimenter were available; typically about one week elapsed between one "choice session" and the next.

5.1.3. A limited electronic newspaper

Programs were developed to display a list of story headers on a window presented on a Sun workstation, as shown on the left part of Figure 1. About 40 were shown at any one time. Additional headers could be presented by moving the cursor below the bottom of the window, which caused the entries on the window to scroll up, or moving the cursor above the top of the window, which caused the entries to scroll down. Stories were chosen by positioning the cursor over a header and clicking the appropriate mouse button. This caused a new window to be created with the text of the news story. The article remained on the screen until the button was released. For each news article the subject chose, the title of the article and the time spent reading it (i.e. how long the mouse button was held down) were recorded. On some trials, perhaps as the result of an accidental button press, the subject would not read an article that had been selected. Therefore the analyses reported below included only articles held for more than 10 s. Because the objective of the experiment was to determine whether preferences would be apparent in a relatively unconstrained news interface, even the order of presentation of the headers was randomized across subjects.

5.1.4. Ratings of 15 news articles

On alternate sessions, ratings of the subjects' interest in certain news stories were collected. To provide a relatively large sample of text the 15 longest articles were selected, except that stories which essentially duplicated stories in the set were replaced with other stories. The subjects were asked to rate from 0–100 how interested they would be in reading the articles if they had seen them in the newspaper. There are both advantages and disadvantages to the use of ratings as a dependent variable. On one hand, the ratings may be considered an indication of the strength of the persons' interest about certain articles. On the other hand they

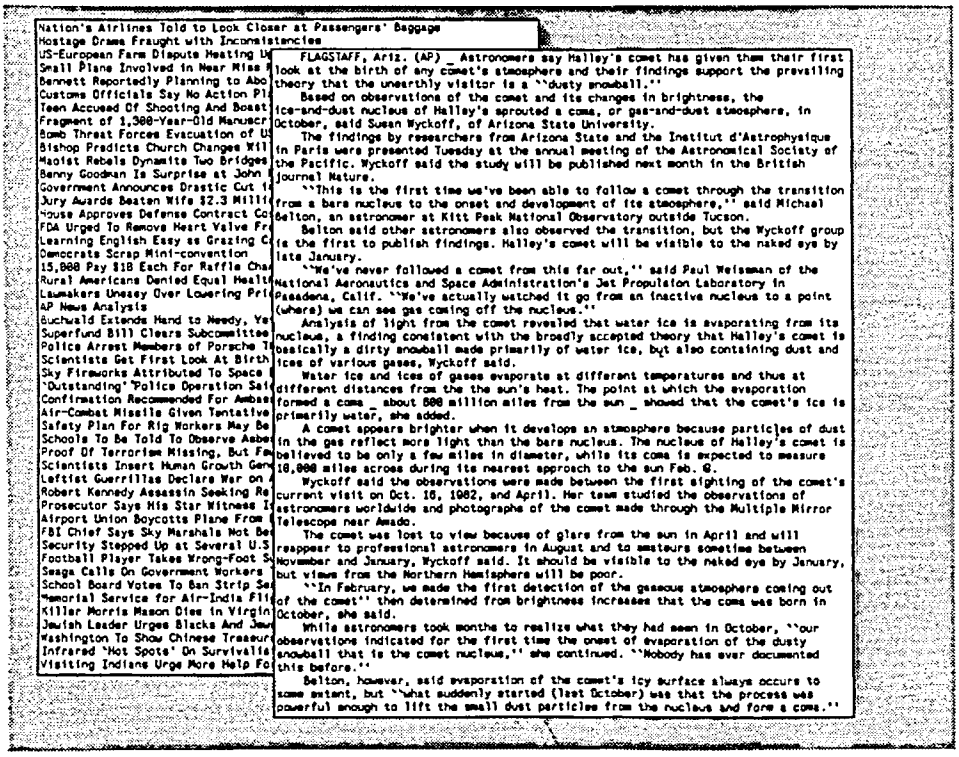

FIGURE 1. Window interface for news articles.

are a form of self-report and self-report must be interpreted with caution (see Section 8.1).

5.2. RESULTS

5.2.1. Descriptive

As shown in Table 4, there were substantial differences among the subjects in the number of articles they read and the amount of time they spent reading them. The headings for the stories read by one subject (no. 3) on two successive sessions are shown in Figure 2. For comparison, Figure 3 shows a random selection of headings for stories which were not read by that subject on the same days as those shown in Figure 2.

5.2.2. Inter-subject consistency in article choices and ratings

An important issue for making predictions is the extent to which the behavior and decisions of other people can be useful predictors (Kochen & Wong, 1962). A test of the overlap was constructed in which the number of overlapping stories was obtained for each pair of subjects on each of the days when both subjects read news articles. The number of overlapping stories compared to the number expected by chance, and no difference was found. In addition, as shown in Table 5 there was

TABLE 4
Number of articles and time spent on each article per session

Subject	Articles read	s/article
1	10·0	89
2	10·5	46
3	16·5	41
4	39·8	43
5	7·7	54
M	16·9	55

relatively little overlap in the ratings the subjects gave to articles and indeed most of the correlations are negative. These results suggests that simple base-rate predictions (Section 8.2) are inadequate for this type of material. Perhaps, effective predictions could be made by combining ratings from several other individuals; however, that was not tested. Moreover, informal impressions from a local online interface to the news articles suggest that in some situations, e.g. when there is a breaking news story, individuals often do access the same articles (see Section 9).

Willis Says FEC Brushed Off Her Complaint About Rajneeshees
Weinberger Says Cutting Industry Security Clearances Will Be Tough
Diablo Canyon Gets Full Power License for 2nd Unit
Crane Lifting Hot Tub Tips Over, Damaging Three Homes
Man Drowns At Party for 100 Lifeguards
Brooklyn Library Reverses Book List Recall
Stocks Shrink, Cars Torched As Strike Continues
Scientists Find Evidence That Other Viruses Sometimes Help Trigger AIDS
Judge Rules Against United Air Lines Inc.
Son, Daughter-in-Law of Carl Sagan Beaten With Hammer
Mail Order House Accused of Fake Billing
University Team To Beam Video Images To Shuttle
10th Anti-Freeze-Tainted Wine Identified, 6 Arrested, Artificial Wine Found
Bank Searches For Missing Shareholders
Good Life Evades Maine Indians 5 Years after Land-Claims Settlement

Atheists Reveal Bible's "Steamier" Sections in Special Edition
Says Will Build High Speed DMC-12 in Ohio
Chemical Defgnse Backfires For Sierra Willow; Attracts Hungry Beetles
'Sis-Boom-Ba' Turns To 'Hiss-Boo-Bah'
Highway Shut, Evacuations In Sierra Nevada
Discovery of Thicker Tissues in Left-Handers May Shed Light on Brain Function
Says Many 'Stars Wars' Critics Motivated by Ideology
Downtown Mall Billed as World's Largest Mobbed at Opening
Reactor Begin Operations, Authorities Say
Detroit Edison Resumes Fermi II Testing

FIGURE 2. Headings of news articles read by one subject (no. 3) on two successive sessions.

Government Safety Agency Sharply Criticized
Judge Visits Woman, Refuses to Order Surgery
Disinfectant Leaks Into Dialysis System At Hospital
Philadelphia Agrees to Allow Women to Guard Male Inmates
Bike Theft Rides The Crest of Dutch Crime
"Incredible Shrinking Man" Actor Grant Williams Dies at Age 54
Shiite Leader Denies Hostages Held By Fundamentalist Family
Nkomo's Homes Raided
Customs Officers Seize Nearly 2000 Pounds Cocaine After Ocean Chase
Dalai Lama's Library Opened To Public in Tibet

Money Supply Down, First-Time Unemployment Claims Up
Middle Tennessee Expects Its Boom to Roll On
Weinberger Defends MX Missile—Says Nation Needs More Than 50
Use of Private Guards At Navy Bases Criticized In Internal Report
19 Recording Companies Agree to Voluntarily Warn of Explicit Lyrics
Parent Union Drops Support of Brown & Sharpe Strike
Man Who Delivered Body to Jail Charged with First-Degree Murder
Reagan Steps Up Effort to Reduce 'Intrusive' Regulations
Speakes Says Doctors Haven't Explained Why They Gave Him Wrong Information
Negotiators Continue Search for End to Car Haulers Strike

FIGURE 3. Random selection of story headings *not* read by the same subject and on the same days as the headings shown in Figure 2.

Note: The text is reproduced as received.

TABLE 5
Mean intercorrelations of subject's ratings on days of overlap

	1	2	3	4	5
1	—				
2	−0·15	—			
3	−0·05	−0·06	—		
4	−0·17	−0·33	−0·03	—	
5	−0·05	+0·37	+0·19	−0·23	—

Note: Each of the correlations reported in Table 5 was calculated only for days on which both subjects in a pair made ratings ($5 < N < 11$).

6. Control conditions and subsidiary experiments

6.1. AFFECTIVE CONTENT AS A PREDICTOR

It is commonly asserted that people prefer sensational, highly emotional news articles. For instance, affective factors were central to Schank's (1979) account of interestingness, and the success of some sensational periodicals suggests that apparently some people prefer that type of reporting. As a test of whether this is an effective level of prediction which could be used to differentiate news stories that the subjects in Section 5 read, an independent group of subjects rated articles on a set of 15 terms. These terms, shown on the left of Table 6, were selected as having various sensational attributes. Most of these terms were derived from the suggestions of

TABLE 6
Factor loadings of terms

	Factors			
	1	2	3	4
Murder	0·91	0·18	0·02	−0·002
Aggression	0·90	0·16	−0·09	0·01
Death	0·87	−0·08	0·06	0·25
Destruction	0·86	−0·29	−0·13	−0·12
Chaos	0·82	0·17	−0·11	−0·12
Danger	0·70	−0·38	−0·30	0·14
Unexpectedness	0·55	−0·33	0·22	−0·08
Ambition	0·09	0·76	0·004	−0·002
Money	−0·46	0·70	0·04	−0·14
Power	0·36	0·65	−0·37	0·03
Wealth	−0·18	0·61	0·49	−0·07
Romance	−0·10	−0·03	0·75	0·13
Sex	0·02	0·39	0·72	0·15
Disease	0·001	−0·06	0·08	0·97
Dilemma	−0·03	0·15	−0·53	0·09
VP	5·01	2·47	1·92	1·12

Schank (1979), as factors affecting "interestingness" in natural language processing (see also Hidi & Baird, 1986). Four subjects made ratings of 30 articles on these scales; articles were chosen from three different days on which subjects in the main study had made ratings. Because many of the terms seemed to overlap, a factor analysis reduced the amount of data and examined the underlying dimensions on which these ratings were made. Four factors were extracted, and the factor loadings are shown in Table 6; the factors seem to be reasonably well characterized by the terms with the highest loading: *Death, Ambition, Romance* , and *Disease.* Scores for each article on each factor were obtained by multiplying the original ratings for the articles times the factor loadings and then summing. These scores were then used as predictors of the interest ratings of the original subjects in several different analyses. Ideally, the analyses would show generally large correlations and consistency within subjects as to the sign of the correlation.

It is unclear how the different components of emotion would be combined in predicting interest ratings. Initially, two values were employed; the maximum, in which the emotion with the largest component was correlated with the ratings, and the sum, in which the total of the four emotion components was correlated with the ratings.

The results of these analyses are shown in the upper two panels of Table 7. Neither analysis yielded a clear pattern of correlations; additional sets of correlations were then generated between the scores on the individual emotion factor and the ratings, and these are shown in the lower part of Table 7. Examination of the correlations for these additional tables suggests that only the correlations for *Romance* are both relatively large and consistent. While there appears to be an effect for at least one of the factors, it should be replicated; and at best, only one of

TABLE 7
Correlations of various factor scores with ratings across 3 days

	Maximum			Factor 2 (Ambition)		
	1	2	3	1	2	3
1	−0·02	−0·49	−0·27	0·61	−0·29	−0·31
2	0·12	0·23	−0·03	0·31	0·24	−0·16
3	x	0·15	0·07	x	0·17	−0·19
4	−0·31	0·05	0·53	0·03	−0·65	0·09
5	0·46	0·05	x	−0·25	0·40	x
	Sum			Factor 3 (Romance)		
	1	2	3	1	2	3
1	0·11	−0·43	−0·15	0·29	0·56	−0·40
2	−0·01	0·31	−0·03	−0·08	−0·43	−0·18
3	x	−0·01	−0·21	x	−0·43	−0·18
4	−0·28	−0·08	0·40	0·14	0·15	−0·35
5	0·28	0·08	x	−0·33	−0·31	x
	Factor 1 (Murder)			Factor 4 (Disease)		
	1	2	3	1	2	3
1	−0·16	−0·39	−0·06	−0·40	0·13	0·31
2	−0·01	0·25	0·09	0·13	0·43	0·19
3	x	0·07	−0·05	x	−0·41	−0·13
4	−0·28	0·09	0·32	−0·29	0·16	0·11
5	0·43	−0·01	x	0·18	0·03	x

the emotional factors was predicted. These results suggest that affective content is not a strong predictor of which news articles a person will read. Anderson, Shirley, Wilson, and Fielding (1987) have recently proposed some other factors that many determine interest (novelty, life theme, and concrete detail) for children's stories.

6.2. READER-GENERATED TERMS AND PHRASES AS PREDICTORS

Up to this point, the evaluation of the models has involved comparison with other models. Obviously another important standard of comparison is how well a person could perform a similar task; this is described in Section 2.4.4 as Turing Control conditions. The original five subjects were each asked to generate terms and short phrases which described the types of news articles which they found either interesting or not interesting. These descriptions were given to other subjects with the 15 articles that had been rated, and these new subjects tried to use these terms and short phrases to predict the ratings. As shown in the left column of Table 8, the correlations of the predictions are modest.

6.3. FRIENDS' RATINGS

In the previous section the subjects in the control condition had relatively little information about the preferences of the person for whom the predictions were to

TABLE 8
Correlation of data from two control conditions and final ratings

	Brief descriptions	Friend
1	0·48	0·65
2	0·43	−0·44
3	0·17	0·26
4	−0·36	0·13
5	0·18	0·25
M	0·19	0·19

Note: For the means of the correlations in this table and elsewhere, a standard $r \rightarrow Z$ transformation was applied.

be made. By contrast in this study friends of the subjects were selected to make predictions about which news articles they thought the subjects would like. The subjects were asked to suggest friends who would be likely to cooperate with the experiment. These results are shown on the right side of Table 8; four of the five correlations were positive. Later questioning of subject no. 2 and his friend suggested that they had been friends for only a few weeks. Perhaps even better predictions could have been obtained from friends with some exposure to the articles which subjects had chosen previously.

7. Predicting from articles previously read

Because attempts at predicting news preferences with simple approaches such as described in the previous sections were not very successful, this section explores several techniques based on analysis of articles which had been previously read. In psychological research (Section 2.5) past behavior has often been found to be an effective predictor of future behavior. During news reading the richest data source of the user's behavior is the text of the articles. Thus the models developed here combine text analysis with user choices.

7.1. DOCUMENT SIMILARITY SCORES

Previous work which has looked at the classification of news stories by topic (DeJong, 1982) may not be relevant to the problem of modeling the preferences of individual users, because the user's categories may not match those established by the program. While similarity is generally desirable for document retrieval, it may not be a useful predictor for the selected news articles. News articles are more redundant than scholarly documents, and newspaper readers may be less tolerant of redundancy than are most users of information retrieval systems. Nonetheless, because the other techniques examined above were poor predictors, it is seriously worth considering this approach.

Each article from the main experiment (Section 3) was passed through a program which identifies parts of speech, and all terms marked as nouns were then passed

through a simple stemming program. The resulting list was sorted and saved. Across the stories read by one subject (no. 1) in the 11 choice sessions there were 4258 unique nouns identified by this method and of these 1558 appeared in more than one article. The similarity of pairs of documents could then be determined with a cosine overlap measure (VanRijsbergen, 1979). Two simple hypotheses involving the similarity measures were examined. All pairs of inter-document similarity scores were calculated for each subject between the articles read on the final session and those articles read on each of the 10 earlier sessions. A second set of scores was calculated between the articles read on the last session and those articles *not* read during the 10 earlier sessions. While it might be expected that more recent articles would show greater similarity than older articles, and that there would be greater similarity between articles read than between articles not read, an analysis of variance on these scores showned no significant effect; indeed, the means are almost identical.

7.2. NOUN-OVERLAP AND RATING MODELS

7.2.1. Choice–choice

Models may be developed which record the predictiveness of individual terms for the subject's interest. To the extent that this technique is sensitive to the frequency of occurrence of terms, it is related to established information retrieval techniques (Sparck Jones, 1986). However, it is quite different in that models are calculated for each subject. This type of model has the advantage that *combinations* of terms which have not appeared in any of the training material may still be identified. The articles were passed through Parts (Cherry, 1982), a program which assigns parts of speech to text passages. The nouns from articles the subject had read were collected and used to predict ratings and the future articles the subject would read.

$$P = \frac{\sum_{t=1}^{N} \left(\frac{H_t}{A_t} \times T_t \right)}{\sum_{t=1}^{N} T_t}.$$

The prediction, P, is a function of H_t (the number of times a relevant term was "hit" in previously read articles), A_t (the number of times the term appeared in the previously read articles), and T_t (the number of times the term appeared in the target article).

The parameters of the model were fit for each subject based on the choice data from the first 10 cycles and then used to predict choices on the 11th session. The ranks, on the final session, of the articles the subjects actually read as a proportion of the entire set of articles, were calculated. The mean of these proportion ranks for each subject is shown in the left column of Table 9. While all of the ranks are less than 0·50, and thus the effect appears to be real, the predictions are not strong. Moreover another analysis, which examined how many of the top articles selected by the model were also read by the subject, showed that while the model's predictions were better than chance, the accuracy was low.

TABLE 9
Correlations of two models with ratings for final session

Subject	Mean rank	Choice-rating r	Rating-rating r
1	0·21	0·73	0·12
2	0·36	0·37	0·24
3	0·25	0·36	0·18
4	0·40	0·19	−0·06
5	0·47	0·44	−0·20
M	0·34	0·44	0·10

7.2.2. Choice-ratings

One explanation for poor prediction in the choice-choice model above is the poor resolution in using the choices as the independent variable. In this section the ratings gathered after the 11th session were used as independent variables and the choices from all 11 sessions were included in the prediction model. Correlations between the predictions of the model based on the articles read and the ratings of the subjects are shown in the center column of Table 9. These results are somewhat encouraging; for instance Subject 1 had a correlation of 0·73. However additional data analyses, such as different weighting formulas, suggested that even this result was not robust. The use of adjectives, in addition to nouns, yielded better models for some subjects but poorer models for others.

7.2.3. Rating model

Just as the ratings improve the resolution of the independent variables, they might also improve the dependent variables. Moreover, while the choices may have been affected by factors such as competition with other articles, that would not be a factor with the ratings. A model was developed from the average of the ratings received for each noun. The correlations between the rating model generated for the first 10 sessions and the ratings made on the final (11th) session are shown in the right-hand column of Table 9. It can be seen that these scores are poor predictors. Moreover, a combined choice and rating model did no better than the simple choice model.

7.3. CHARACTERISTICS AND LIMITATIONS OF THE OVERLAP MODELS

The models described above grew to be quite large; for instance after the first day, the model for Subject 3 reached more than 7000 terms, and by the 11th session it had reached 27,000 terms. Examination of this model does show some promising regularities. Table 10 lists the terms that have the highest ratio of H_t/A_t given that the term appeared in at least 20 different articles. This is the ratio of the number of times the terms appeared in articles which the subject read, over the number of times the terms appeared in corpus of articles which the subject had a chance to read, but didn't. Most of these terms are related to science and technology, and this seems to give a strong indication of the subject's interest. On the other hand, it was not a clear enough trend to determine predictions in the entire corpus (Section 7.2); nor, did the other subjects show such strong clustering.

TABLE 10
Twenty most predictive terms from choice model for subject no. 3

Rank	Ratio	Term	Rank	Ratio	Term
1	0·800	comet	11	0·375	leak
2	0·750	young	12	0·368	venture
3	0·500	theory	13	0·368	shuttle
4	0·500	gene	14	0·364	reactor
5	0·455	species	15	0·350	collection
6	0·444	crane	16	0·333	uranium
7	0·424	technology	17	0·333	sun
8	0·407	method	18	0·333	screen
9	0·400	irs(IRS)	19	0·333	cox
10	0·375	valve	20	0·320	scientist

It seems difficult to make detailed predictions about preferences for specific news articles. The most successful model predicted ratings on the final session from choices (Section 7.2.2). Of course there were substantial limitations in the procedure that might have allowed better generalization. The 11 sessions reported here provide only a limited baseline for making predictions of this type. Possibly with a much longer baseline, more stable predictions could have been observed. In addition, the predictions do not demonstrate the *strength* of preferences for the articles which the subjects chose; for instance, whether the subject would have been almost as satisfied with one set of articles as with the other. However limited evidence against this interpretation may be obtained from the ratings, which showed no apparent discontinuities.

7.4. HUMAN PREDICTION *VS* COMPUTER USER MODELS

The previous sections examined the performance of the model in comparison to an absolute standard. To evaluate the performance of the computer model, human readers were given material similar to that available to the computer program. Of course, the subjects in this procedure had at least three advantages over the computer program; they understood natural language, they had general world knowledge with which to interpret the news, and they had some model of how other people would interpret the news.

An independent group of five subjects was paired one-to-one with the five subjects from the main experiment (Section 5). These "yoked" subjects were asked to form an impression of the interests of the original subjects by studying a carefully constructed set of articles. This set included two sets of all the articles selected by the subject during two of the choice sessions in the main experiment and additional unread articles randomly taken from the same day's set of articles making a total of 40 articles. The additional subjects were then asked to make predictions of how the original subjects would rate 15 articles from one of the rating sets. The correlations between the ratings of the original subjects and the predictions of the model based on exactly the same set of 40 articles are shown in Table 11 (rather than the complete set of terms). Given this comparison, the subjects do much better than the limited model. While comparisons are difficult, the predictions appear to be slightly

TABLE 11
Correlations for predictions of model and "yoked" subjects based on 2 days of background information compared with ratings of subjects in the main experiment

	Yoked subject	Limited rating model
1	0·51	−0·02
2	0·31	0·19
3	0·32	−0·04
4	0·35	0·36
5	0·13	−0·25
M	0·33	0·05

lower than those obtained when a model based on the full data set is used (see Table 9).

8. Predicting book preferences

The studies predicting news articles indicate it is difficult to make predictions at a detailed level. This is in contrast to the apparent ease that GRUNDY had in predicting preferences for fiction, primarily using adjectives as independent variables. Of course, books are different from news articles both in their length and in being self contained. Thus, it is somewhat difficult to compare directly the success of predictions for the two types of material. To understand more about the reasons for GRUNDY's performance, four additional studies were conducted. These studies also illustrate a hierarchical evaluation set of predicting independent variables.

8.1. YES/NO RESPONSES

In the studies reported above, the subjects were asked to make ratings about the extent to which they would like to read given news stories. However for GRUNDY, subjects made yes/no responses about whether they would like to read a book. It seems likely that subjects' verbal reports might not correspond to actual behavior. Thus, on the day after the 7th cycle, the subjects of the main news study (Section 5) were presented with 20 randomly selected articles from AP newswire and asked to state yes/no whether they would like to read the article. The subjects responded "yes" to 47% of the articles. However on the following session with the electronic newspaper, they read only 7% of the available news stories. One explanation is that the subjects had different thresholds for articles that they expressed interest in and those they took the time to read. An alternate explanation is that subjects found articles of equal interest and they chose randomly from that collection.

8.2. BASE-RATE PREDICTIONS

A second issue for prediction is the extent to which individual differences need to be included in a model. As noted in Section 8, predictions of complex models need to

be compared with the predictions of minimal-information models. For the news articles, it seems unlikely that base-rate or minimal-information predictions would be satisfactory; as observed in Section 5.2.2 there was little overlap in the stories which people actually read. However, this might be a factor in predicting book preferences. The popularity of best-sellers suggests that there may be substantial overlap.

A test was made whether predictions of subjects' stated preferences for books could be based simply on base-rate information. At the end of the human-prediction study (Section 8.4) the 12 subjects in that study were asked to state whether or not they would like to read the 20 books from the "random" set. The three books which got the highest ratings and the three books which got the lowest ratings were selected and given to 20 additional subjects. Overall, the three high base-rate books were selected 39 times and the low base-rate books were selected 21 times. The difference between the high and low base-rate was significant, $t(19) = 3{\cdot}00$, $p < 0{\cdot}01$.

8.3. MINIMAL-INFORMATION PREDICTIONS

Assuming that individual differences contribute additional resolution to the model beyond base-rate predictions (Section 8.2), the question arises as to what is the simplest independent variable needed to make effective predictions (see Section 8). Rich observed that male/female differences in reading preferences appeared as a component of GRUNDY's stereotypes. The brief study reported here considered whether that component alone, without the complexity of the rest of the model, could be used to make effective predictions.

From the set of 40 book descriptions (see Section 8.4), five were selected by the Experimenter as books which would be expected to appeal more to males and five others as books which would be expected to appeal to females. These were given to 10 male and 10 female subjects. The original rating of the book was an effective predictor of preference; overall, the "male" books were preferred by males and the "female" books by females, $F(1, 36) = 4{\cdot}76$, $p < 0{\cdot}05$.

8.4. HOW SUCCESSFUL ARE PEOPLE AT SUGGESTING BOOKS FOR OTHERS TO READ

As a final control condition, a human's predictions were compared to the kinds of predictions which GRUNDY made. 12 female college graduate subjects from the local community participated in one of two conditions. In one condition, the six subjects chose those adjectives which they felt applied to them from a list of 20 adjectives, similar to those used in GRUNDY. The six remaining subjects chatted with the Experimenter for about 5 min. Forty cards were prepared, each giving the title and a short summary of one book. These were similar to the materials prepared by Rich and in some cases they were taken verbatim from the materials she reports. The cards were divided into two groups of 20. After either chatting with the subjects or reading the list of adjectives, the Experimenter chose cards for six books from one group of books based on the impression she had of the subject. In addition, cards for six books were chosen at random from the second set of 20 cards. The 12 cards were then merged and the subject looked through each summary and stated whether she would like to read the book. Finally, the subject made a decision on each of the remaining 14 books in the random set. As shown in Table 12, there was

TABLE 12
Preference for random book predictions vs Experimenter's predictions

	Adjectives	Conversation
Random > selected	3	1
Random = selected	1	1
Random < selected	2	4

no difference between the random selections and the impression-based selections. However, for the subjects where impressions were based on conversation with the Experimenter, the predictions made by the Experimenter were consistently better than the random predictions. This difference was close to statistical significance, $t(10) = 1{\cdot}81, p < {\cdot}06$. In this case, adjectives alone were ineffective predictors. However, in most cases reasonable predictions could be made from brief conversations.

9. Additional research topics

9.1. NEWS AS HYPERTEXT AND INTEGRATION OF NEWS WITH OTHER INFORMATION SOURCES

News stories are often interrelated and they might be thought of as linked together forming a hypertext document. It is easy to imagine moving through a web of related articles or perhaps from one concept within one article to related concepts within other articles. Moreover, it seems natural to make links from news articles to other information sources such as encyclopedia articles or perhaps a picture archive. Table 2 suggests that most of the content of even front-page news articles consists of general knowledge. Indeed, newspapers may be thought of as a continually updated encyclopedia.

The electronic presentation of news articles suggests the need for new computer tools and technologies. For instance, the writer of a news article might use an authoring workstation which had access to a wide variety of background and reference materials. Hooks to standard reference materials could be built into the text and those references could then be retrieved by the user's system when they were requested. Non-standard material could be made available across the network as needed. In some cases static hyperdocuments such as electronic books may have links crafted by the authors. However, for large active document collections such as news text intended for a heterogeneous audience, the links might be made automatically. It is natural to assume that a user model might generate appropriate links. However, this application of user models may have many of the same limitations observed in the work described here.

9.2. ADDITIONAL TOPICS: USER MODELS

Of course an important issue for study is whether by collecting additional data, much better predictions could be achieved for news preferences. Beyond that

possibility, several other directions for research are also suggested. While this paper has focused on user models for information interfaces, user models may be applied in many other situations. For instance they have been widely discussed in tutoring systems and expert systems, although their contribution has not been systematically studied.

9.2.1. Modeling human editors and news writers

Human newspaper and magazine editors are generally successful at selecting appropriate articles for their readers. Moreover, magazine and newspaper editors have implicit or even explicit (American Society of Newspaper Editors, 1984) models of their audience; thus, a natural investigation to examine how editors choose articles could be studied systematically. One direction would be to interview news editors and perhaps present them with test articles. A second approach would be to attempt to model the articles that appear on the front page of a newspaper over a period of time. Of course there is variability among editors and it seems, if one looks at two different newspapers, that different editors may choose very differently. A related issue would be to investigate how reporters write articles. Clearly, many "editorial" decisions are made before and during the composition of a specific news article.

9.2.2. Predicting individual information importance

While much of this paper has emphasized the prediction of information preferences, there are also many topics of interest in predicting user information *needs* (Dervin, 1983) especially when those needs do not correspond to a specific task the user is engaged in. For instance, consider whether a news article about a hospital would be of interest to a reader. To answer this, this system must have more than knowledge of the user and an understanding of the semantics. It would have to have general world knowledge about the likelihood of certain events occurring and detailed knowledge of the user, the user's family, acquaintances, and possessions.

9.2.3. Individualizing news text

Although the approach taken in this paper is to look at entire news articles, it is also instructive to consider the structure and function of specific news articles, paragraphs within articles, and even individual sentences. This level of prediction might be important if a system were designed to compose individualized text for news stories. News articles may be viewed as expository text (e.g. Britton & Black, 1985). Typically only a few facts are presented in a news article, and a much greater portion of the news writing is spent in providing background and in putting the information in context for the reader (Section 4.1).

9.2.4. Generality of user models

On the assumption that effective user models can be developed (Sections 4, 5), it is reasonable to wonder whether these can be applied beyond news. Presumably the model that is developed from news stories reflects important aspects of the users' interests and could be applied in other areas such as ICAI or the automatic selection of material from an encyclopedia. Would the same models be useful for other types

of selection such as books and video? The general question is when are models applicable across time, situation, and perhaps across "similar" individuals.

9.2.5. User-user prediction

While Table 5 (Section 5.2.1) indicated little overlap among the information preferences of the participants in the news-article access study, extensions of that concept might yield more promising results. For instance, rather than looking for overlap on specific stories it would be possible to look for overlap across groups of related stories. Another approach might be to find groups of users who shared interests. Moreover, it should be noted that although the procedure employed in these studies showed few overlaps when no feedback was available, it is possible that when such information is available it would be a strong determinant of behavior.

9.2.6. User agents

While user models are typically thought of as filters for incoming information, the concept of user models might be extended to incorporate *user agents* (Waterman, 1978). Several levels of responsiveness might be considered for user agents. First, unlike simple user models, user agents could respond to inquiries directed from other human and computer agents. For instance, Greif and Sarin (1987) report work on distributed personal databases such as those employed in the calendar scheduling procedures. Second, user agents might attempt to produce the responses that the user would typically give. Third, information could be gathered in anticipation of a user's needs. Fourth, user agents might be able to engage in negotiation on behalf of the individual and make commitments for that individual. Thus, detached-active user agents could *gather information* and *initiate* negotiation on a user's behalf.

9.2.7. Ultimate user control

Users may want the ability to bias preferences in the user model or to completely override some of its assumptions. This may be termed *ultimate user control over the user model.* While this is conceptually simple, its implementation may be difficult. For instance, one issue is interpreting the intent of a user's instruction. A second problem is that adaptive models may "adapt around" the constraints imposed by the user.

9.2.8. Detailed user feedback

If the user model compiles information about the user in some systematic way, that information might itself be of interest to the user. Thus, the model might provide feedback to the user about preferences or performance in relation to others. Indeed, a user might be interested in using the model for projections of his/her own reactions.

9.3. ADDITIONAL TOPICS: CHARACTERISTICS AND USES OF NEWS

9.3.1. Quick impressions of news: scanning the headlines vs *reading articles*

These studies have been based on the scenario that people select a given news article, read it in depth, and then pass on to another news article. Casual

observation suggests that this may not always be valid and that people often scan a newspaper quickly. This might be examined in much greater detail by analyzing the eye movements of a person who was reading a page from a newspaper. Beyond merely noting the occurrence of the news skimming, its purpose might be considered. For instance, is it primarily to obtain an overview of the news, or simply filtering topics looking for one which might be of interest? Finally, if as seems likely, this news scanning serves as a quick overview of the news, facilitation by computer interface might be considered.

9.3.2. Stability and generality of information preferences

In the studies reported here, the information (news) was clearly changing and presumably, there were short-term changes in the individuals such as acquiring information from other news sources. An individual's underlying information preferences may shift across time and this could be studied with more extensive testing. Moreover, it would be of interest to isolate and identify the factors which initially shaped the underlying news preferences.

9.3.3. Dimensions of news

Perhaps if the dimensions along which news articles varied were known, prediction would be straightforward. One strategy for exploring the dimensions would be to have subjects make similarity judgements of pairs of news articles. Statistical techniques such as multi-dimensional scaling could be applied to attempt to determine dimensions of similarity. These dimensions could then be tested for predictive utility.

9.3.4. Memory for and uses of news

While a brief attempt was made in Section 4.2 to study systematically how people read newspapers, clearly many other questions remain concerning how people read, react to, remember, and use news articles. For instance, subjects might be asked to explain current events and the answers might be used to understand the subjects' mental organization of information.

9.3.5. Contribution of pictures to news

A substantial portion of a newspaper may include pictures (see Section 4.1) and it is reasonable to ask what contribution they make. As with news itself (Section 3.2), it is difficult to isolate the functions of photographs. On one hand, pictures and graphics may simply provide information. On the other hand, as demonstrated in Section 4.1 much of the purpose of pictures seems to be emotional.

9.3.6. Multiple information sources

Of course, written news articles are not a person's only source of information and there are complex interactions of these various information sources. One possibility is that people may be less interested in one medium if they have found the information in another medium. On the other hand there may be complementary effects. For instance, a person might want to see pictures (either news photos or television) of news that had been on the radio.

10. Conclusion

This paper has focused on user models for predicting information preferences. Section 4 provides some evidence that user models are possible for gross categorizations. However the research in Sections 5, 6 and 7 makes it clear that fine-grained prediction is difficult. It is likely that additional data collection could improve performance of the fine-grained models for prediction of news preferences such as those developed here. However, it remains an open question whether sufficient accuracy can be achieved to make much difference in the design of effective news presentation systems. Presumably this would also be true in other areas of application for user models. For instance in education, grade level might be used to guide an ICAI system and that might be satisfactory for many cases. It remains an open question whether fine-grained user models in other areas such as ICAI, will be feasible and useful. It is intriguing to speculate on whether there is an inherent unpredictability in human behavior or whether with sufficient information all behavior could be predicted.

Beyond the specific results which we have reported on predicting preferences for news articles, this paper has attempted to establish a *systematic study of information interfaces and user models.* This effort covers a broad range of topics. Many of the issues that have been encountered in this paper are related to social psychology. Indeed, it may be proposed (Allen, submitted) that social psychology is an important paradigm for computing research, which includes many of the most active research topics in computer science, speech acts, group interaction, categorization of behavior, and belief systems. Certainly, the selection of newspaper articles is relatively unconstrained, and unconstrained behaviors have long been recognized by social psychologists as difficult to predict (Ajzen & Fishbein, 1977; Snyder & Ickes, 1985).

Ultimately, new conceptions of information media will probably emerge. For instance, the historical events in a textbook or encylopedia are in some sense, simply different views on material that might once have appeared in newspapers. In any case, while the prediction of news articles appears difficult, some sort of information filtering seems inevitable because of the large amount of material. Thus the refinement of user models will remain an important challenge for future research.

I thank those subjects who volunteered for various conditions in these studies and especially the subjects who participated in the main study. I also thank Selma Kaufman for serving as the Experimenter in these studies; her care and interest contributed greatly to the research.

References

AJZEN, I. & FISHBEIN, M. (1977). Attitude-behavior relations: a theoretical analysis and review of the empirical literature. *Psychological Bulletin,* **84,** 888–918.

ALLEN, R. B. (1989). Developing agent models with a neural reinforcement technique. *IEEE Systems, Man and Cybernetics Conference,* November, Cambridge, MA.

ALLEN, R. B. Social processes and computing, submitted.

AMERICAN SOCIETY OF NEWSPAPER EDITORS. (1984). *Relating to Readers in the 80's* (ASNE, Washington, DC).

ANDERSON, R. C., SHIRLEY, L. L., WILSON, P. T. & FIELDING, L. G. (1987). Interestingness of children's reading material. In R. Snow & M. Farr, Eds. *Aptitude, Learning and Instruction: Cognitive and Affective Process Analyses.* Hillsdale, NJ: Lawrence, Erlbaum Associates.

BOBROW, D. G. (1975). Dimensions of representation. In D. G. Bobrow & A. Collins Eds. *Representation and Understanding,* New York: Academic Press, 1–34.

BOWERS, K. S. (1973). Situationalism in psychology. *Psychological Review,* **80,** 307–336.

BRAJNIK, G., GUIDA, G. & TASSO, C. (1987). User modeling in intelligent information retrieval. *Information Processing and Management,* **23,** 305–320.

BRITTON, B. K. & BLACK, J. B. (1985). *Understanding Expository Text.* Hillsdale, NJ: Lawrence Erlbaum.

BROOKS, R. M., DANIELS, P. J. & BELKIN, N. J. (1985). Problem descriptions and user models: developing an intelligent interface for document retrieval systems. *Advances in Intelligent Retrieval, Proceedings of Informatics,* **8.**

BURGOON, J. K., BURGOON, M. & ATKIN, C. K. (1982). What is news? Who decides? And how? (A preliminary report conducted for the American Society of Newspaper Editors.)

CARROLL, B. (1987). Expert systems for clinical diagnosis: Are they worth the effort? *Behavioral Science,* **32,** 274–292.

CHERRY, L. (1982). Writing tools. *IEEE Transactions on Communications, COM-30,* 100–105.

CHRISTIE, B. (1981). *Face-to-File Communication.* Chichester: Wiley.

CLANCEY, W. J. (1984). Classification problem solving. *Proceedings of the National Conference on Artificial Intelligence,* August, Austin, TX, 49–55.

CROFT, W. B. & THOMPSON, R. H. (1984). The use of adaptive mechanisms for selection of search strategies in document retrieval systems. In C. J. van Rijsbergen Ed. *Research and Development in Information Retrieval,* pp. 96–110. Cambridge: Cambridge University Press.

CRONBACH, L. J. & SNOW, R. E. (1977). *Aptitudes and Instructional Methods.* New York: Irvington.

DANIELS, P. J. (1986). Cognitive models in information retrieval—an evaluative review. *Journal of Documentation,* **42,** 272–304.

DEJONG, G. (1982). An overview of the FRUMP system. In W. G. Lehnert & M. H. Ringle Eds. *Strategies for Natural Language Processing.* Hillsdale, NJ: Lawrence Erlbaum.

DERVIN, B. (1983). Information as a user construct: the relevance of perceived information needs to synthesis and interpretation. In S. A. Ward & L. J. Reed, Eds. *Knowledge Structure and Use: Implications for Synthesis and Interpretation,* pp. 153–183. Philadelphia, PA: Temple University Press.

Dozier, D. & Rice, R. E. (1984). Rival theories of electronic newsreading. In R. E. Rice *et al.* Eds. *The New Media,* pp. 103–127. Beverly Hills: Sage Publications.

FERGUSON, C. E. (1975). *Micro-economic Theory.* IL: Irwin, Homewood.

FISHBEIN, M. & AJZEN, I. (1975). *Belief, Attitude, Intention, and Behavior.* Reading, MA: Addison-Wesley.

FRENCH, C. W., POWELL, E. A. & ANGIONE, H. (1980). *The Associated Press Stylebook and Lible Manual.* Reading, MA: Addison-Wesley.

GREIF, I. & SARIN, S. (1987). Data sharing and group work. *ACM Transactions an Office Information Systems,* **5,** 187–211.

HAYILAND, S. E. & CLARK, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior,* **10,** 512–521.

HIDI, S. & BAIRD, W. (1986). Interestingness—a neglected variable in discourse processing. *Cognitive Science,* **10,** 179–194.

KASS, R. & FININ, T. (1986). *Modeling the User in Natural Language Systems.* University of Pennsylvania Technical Report, MS-CIS-86-87.

KATZ, E., BLUMLER, J. & GUREVITCH, M. Utilization of mass communication by the individual. In: J. Blumer and E. Katz Eds. *The Uses of Mass Communication,* pp. 19–34. Beverly Hills: Sage Press.

KING, D. W., RODER, N. K. & OLSEN, H. A. Eds. (1983). *Key Papers in the Economics of Information.* White Plains, NY: Knowledge Industry Publications.

KOCHEN, M. & WONG, E. (1962). Concerning the possibility of a cooperative information exchange. *IBM Journal,* 270–271.

KOVED, L. & SHNEIDERMAN, B. (1986). Embedded menus: selecting items in context. *Communication. ACM,* **29,** 312–318.

LEGGATE, P. (1975). Computer based current awareness services. *Journal of Documentation,* **31,** 93–115.

LESK, M. E. (1982). Combining data bases: National and cartographic files. *Proceedings AFIPS Office Automation Conference* (San Francisco, April) 415–426.

LIPPMAN, A. & BENDER, W. (1987). News and movies in the 50 megabit living room. Globecom (Tokyo, November) 50.1.1–50.1.6.

LUCAS, R. W. (1977). A study of patients' attitudes to computer interrogation. *International Journal of Man–Machine Studies,* **9,** 69–86.

LUHN, H. P. (1958). A business intelligence system. *IBM Journal of Research and Development,* **2,** 314–319.

MALONE, T. W., GRANT, K. R., TURBANK, F. A., BROBST, S. A. & COHEN, M. D. (1987). Intelligent information-sharing systems. *Communications of the ACM,* **30,** 390–402.

MALONE, T. W., GRANT, K. R. & TURBANK, F. A. (1986). The information lens: an intelligent system for information sharing in organizations. *Proceedings of SIGCHI '86* (Boston, April) 1–8.

MASON, M. V. (1986). Adaptive command prompting in an on-line documentation system. *International Journal of Man–Machine Studies,* **25,** 33–51.

MAURER, H., ROZSENICH, N. & SEBESTYEN, I. (1984). Videotext without 'Big Brother.' *Electronic Publishing Review,* **4,** 201–214.

MISCHEL, W. (1968). *Personality and Assessment.* New York: Wiley.

MOGHDAM, D. (1978). *Computers in Newspaper Publishing.* New York: Marcel.

NEWSPAPER CENTER, (1984). *Wire Service Transmission Guidelines,* Report 1312, June 15.

ODDY, R. N. (1977). Information retrieval through man-machine dialogue. *Journal of Documentation,* **33,** 1–14.

PARKER, D. B. (1979). *Ethical Conflicts in Computer Science and Technology.* Arlington VA: AFIPS Press.

PATTEN, D. A. (1986). *Newspapers and New Media.* White Plains, NY: Knowledge Industry Publications.

RICH, E. (1979). *Building and Exploiting User Models.* CMU Dissertation.

RICH, E. (1983). Users are individuals: individualizing user models. *International Journal of Man–Machine Studies,* **18,** 199–214.

RUMELHART, D. E., HINTON, G. E. & WILLIAMS, R. J. (1986). Learning internal representations by error propagation. In: D. E. RUMELHART & J. L. MCCLELLAND Eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Cambridge, MA: Bradford Books/MIT Press.

SALTON, G. & MCGILL, W. J. (1983). *Introduction to Modern Information Retrieval.* New York: McGraw-Hill.

SCHANK, R. C. (1979). Interestingness. *Artificial Intelligence,* **12,** 273–297.

SEARLE, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

SCHMIDT, C. F., SHRIDHARAN, N. S. & GOODSON, J. L. (1978). The plan recognition problem: An intersection of psychology and artificial intelligence. *Artificial Intelligence,* **11,** 45–83.

SNYDER, M. & ICKES, W. (1985). Personality and social behavior. In G. LINDZEY & E. ARONSON Eds. *The Handbook of Social Psychology.* vol. 2, 3rd ed. pp. 883–991. New York: Random House.

SPARCK JONES, K. (1986). Issues in user modeling for expert systems. In: A. G. COHN & J. R. THOMAS Eds. *Artificial Intelligence and its Applications,* pp. 183–195. Wiley: Chichester.

STANFILL, C. & KAHLE, B. (1986). Parallel free text search on the Connection Machine system. *Communication. ACM,* **29,** 1229–1239.

STEPENSON, W. (1967). *The Play Theory of Mass Communication.* Chicago: University of Chicago Press.

TURING, A. (1950). Computing machinery and intelligence, *Mind,* **59,** 433–460.

VANRIJSBERGEN, C. J. (1979). *Information Retrieval.* London: Butterworths.

WATERMAN, D. A. (1978). Exemplary programming in RITA. In D. A. WATERMAN & F. HAYES-ROTH, Eds. *Pattern-Directed Inference Systems,* pp. 261–279. New York: Academic Press.

WEYER, S. (1982). The design of a dynamic book for information search. *International Journal of Man–Machine Studies,* **17,** 87–107.

WEYER, S. & BORNING, A. (1985). A prototype electronic encylopedia. *ACM Transactions on Office Information Systems,* **3,** 63–88.

YANKELOVICH, N., MEYROWITZ, N. & VAN DAM, A. (1985). Reading and writing the electronic book. *Computer,* **18,** 15–30.