

# Information Seeking and Mediated Searching. Part 2. Uncertainty and Its Correlates

T.D. Wilson, Nigel Ford, David Ellis,\* Allen Foster

Department of Information Studies, University of Sheffield, UK. E-mail: t.d.wilson, n.ford, a.e.foster@shef.ac.uk

Amanda Spink

School of Information Sciences and Technology, The Pennsylvania State University, 004C Thomas Building, University Park, PA 16801. E-mail: spink@ist.psu.edu

This article explores the relationship between the concept of *uncertainty* in information seeking, within a model of the problem-solving process proposed by Wilson (1999a) and variables derived from other models and from the work of Ellis and Kuhlthau. The research has involved longitudinal data collection in the United States and United Kingdom employing three interview schedules (incorporating self-completed questionnaires) used for pre- and postsearch interviews; and postsearch interviews with the information seeker and the search intermediary. In addition, the Sheffield team employed a fourth set of instruments in a follow-up interview some 2 months after the search. Related search episodes, with a professional search intermediary using the Dialog Information Service and other sources were audiotaped, and search transaction logs were recorded. The mediated search clients were faculty and research students engaged in either personal or externally supported research projects. The article concludes that the problem solving model is recognized by such researchers as describing their activities and that the uncertainty concept, operationalized as here, serves as a useful variable in understanding information-seeking behavior. It also concludes that Ellis's concept of "search characteristics" and Kuhlthau's information-seeking stages are independent of the problem stage, and that a set of affective variables, based on those of Kuhlthau, appear to signify a generalized positive or negative affective orientation towards the course of the information problem solution.

## Introduction

Briefly, two projects were brought together through the use of a common set of research instruments, which were

---

\* Currently Professor, University of Wales, Aberystwyth.

Received February 18, 2001; revised October 17, 2001; accepted January 25, 2002

© 2002 Wiley Periodicals, Inc.

employed in longitudinal studies of the information seeking and mediated search activities of 198 researchers (87 in the United States and 111 in the UK). The aim of the UK project was to investigate several aspects of information seeking and searching in the context of a theoretical model of the problem-solving process, with particular attention to the concept of uncertainty and to the theoretical perspectives of Kuhlthau (1993a) and Ellis (1989). One of the main aims was to operationalize the concept of uncertainty and to relate it to other aspects of the information-seeking and searching process.

## Uncertainty

Perhaps the earliest *formal* association of information and uncertainty was set out by Shannon and Weaver (1949), for whom *information* itself refers to the reduction in uncertainty about the state of an event after a message has been sent relative to the uncertainty about the state of the event before the message was sent. Since then, uncertainty has become a concept that is studied in a variety of fields, most notably in classical probability theory, but also in theories of decision making and in artificial intelligence research.

For example, the Association for Uncertainty in Artificial Intelligence ran its 16th annual conference in 2000. Much of this work concerns Bayesian probability theory and fuzzy logic, but some is of more direct relevance to information seeking and searching. For example, Henkind (1986) considers the kind of uncertainty that results from lexically imprecise words in medical expert systems; that is, words that may denote several similar, but not identical entities. The IR field is also directly concerned with uncertainty in terms of fuzzy logic and probabilistic theory, as the workshops on *Information retrieval, logic and uncertainty* (e.g., van Rijsbergen, 1998) stimulated by van Rijsbergen's (1986) suggestion that relevance in IR might be modeled as

an uncertain inference process. Ingwersen's (1996) cognitive theory for information retrieval interaction also seeks to relate logical uncertainty, the information space of IR systems, and the cognitive space of the information user in a concept of *polyrepresentation*, suggesting that different representations of an item in an IR system should be employed to improve retrieval.

Uncertainty also has a long history of being associated with decision-making research: as Harris (1998) notes:

Decision-making is the process of sufficiently reducing uncertainty and doubt about alternatives to allow a reasonable choice to be made from among them. This definition stresses the information gathering function of decision-making. It should be noted here that uncertainty is reduced rather than eliminated. Very few decisions are made with absolute certainty because complete knowledge about all the alternatives is seldom possible.

In this field, statistical and mathematical modeling are also employed to test theoretical models of the decision making process, with the aim, for example, of showing how optimal decisions might be arrived at (see, e.g., Moore & Thomas, 1988; Targett, 1996). An area of particular economic significance is that of decision making in currency markets (e.g., Bos, van Dijk, & Mahieu, 1999).

### Uncertainty and Information Seeking

Researchers in various fields have also been concerned with the relationship between uncertainty and information seeking; for example, Mignerey, Rubin, and Gorden (1995) show the relationship between uncertainty and information seeking in the case of the newcomer to an organization, and Borgers et al., (1993) show the relationship between *experienced uncertainty* and the information-seeking behavior of cancer patients.

In information science, the idea of uncertainty underlies all aspects of information seeking and searching, from Belkin's (1980) idea of the Anomalous State of Knowledge of the information seeker to Kuhlthau's (1993a) linking of affective states to information-seeking stages. Kuhlthau's (1993a) stage model of the information-seeking process links uncertainty to various stages in that process, specifically in the *Initiation* stage, when a person first becomes aware of a need for information, and during the *Exploration* stage, when the individual is seeking to establish the general field of the problem. Kuhlthau (1993b) has also proposed uncertainty as a basic principle for information seeking, defining uncertainty as: "...a cognitive state which commonly causes affective symptoms of anxiety and lack of confidence." and, drawing upon her research, noting that: "Uncertainty and anxiety can be expected in the early stages of the information search process . . . Uncertainty due to a lack of understanding, a gap in meaning, or a limited construct initiates the process of information seeking."

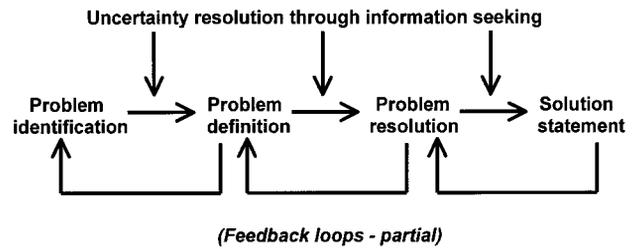


FIG. 1. The problem solving model.

Kuhlthau (1993a) proposes six corollaries to this fundamental principle: the *process corollary* (in the information-seeking process the user actively pursues understanding), *formulation corollary* (defining and extending understanding on the basis of information found in the course of a search), *redundancy corollary* (redundancy—the discovery of what we already know—decreases as uncertainty increases), *mood corollary* (the attitude of the information seeker that limits or extends the possibilities in the search for information), *prediction corollary* (based on expectations from past experience), and *interest corollary* (interest in the problem increases as uncertainty decreases).

Kuhlthau (1996) has also identified the relationship between task complexity and uncertainty and has explored this (Kuhlthau, 1997) in a case study of one securities analyst, where she examined the relationship between uncertainty, complexity of the task, the use of a wide range of resources and "extensive formulation." In a later paper (Kuhlthau, 1999) concludes that:

There is a marked contrast between information systems that are built on a principle of order and certainty, and the individual's profound experience of uncertainty in the early stages of the information search process. These findings indicate the need for considering uncertainty as a natural, essential characteristic of information seeking rather than regarding the reduction of uncertainty as the primary objective of information seeking. Uncertainty is a concept for information retrieval design that offers insight into the user's quest for a personal perspective within the process of information seeking, what I have called "formulating a focus." When uncertainty is accepted as an important concept many other related corollary concepts become apparent.

### Theoretical Framework

Wilson's (1999a) *problem-solving model* is used as the top-level concept to explain why people engage in information seeking (Fig. 1). This model suggests that information-seeking behavior (ISB) is *goal-directed behavior* with the resolution of the problem and, possibly, the presentation of the solution as the goal. In moving through each of the stages of *problem identification*, *problem definition*, *problem resolution*, and *solution presentation*, uncertainty must be reduced, and individuals are seen as engaging in interaction episodes with information sources (including people and other sources, as well as information retrieval systems) to resolve their uncertainty. More than one set of informa-

tion-seeking or searching activities may be necessary before the person is able to move on to the next stage of the process. The idea of stages, together with feedback, explains the phenomenon of successive searching (Spink, 1996), where information seekers are seen to carry out more than one search for information relating to their problem. Of course, attempts to resolve uncertainty may actually increase it and, therefore, the model provides for feedback within and between stages (Fig. 1 simplifies this process, because feedback is not necessarily between adjacent stages).

Kuhlthau's (1993a) model of the stages in information seeking may be seen as an elaboration of the problem-solving process, or we may view it as the stages a searcher proceeds through at any one of the problem-solving stages, because the possibility exists that the resolution of an information problem within any one of the problem stages may involve all of Kuhlthau's information-seeking stages. Ellis's (1989) search "characteristics" are seen as applying to either searching during the different problem stages or during any one of Kuhlthau's (1993a) stages.

The framework, therefore, seeks to bring together a number of different ideas from ISB research.

## Research Questions and Hypotheses

From the perspective of this article, the theoretical frameworks under review gave rise to the following research questions and associated hypotheses:

- (1) *Is the problem-solving stage model recognized by clients as appropriate for recording their progress on a project?* The problem-solving stage model had been suggested by Wilson (1999a) as a generalized basis for explaining the fact of successive searches by clients dealing with a particular issue, and it was of interest to the research team to determine whether it offered a way of categorizing clients that would provide a useful analytical variable. This research question does not have an associated hypothesis, other than in the very general sense that, if the problem-solving stage model *is* useful, the clients will be able to identify the stage in which they find themselves, whereas if it is *not* useful they will be unable to do so.
- (2) *Is the concept of uncertainty recognized by clients? Can they use a presented scale to indicate how certain/uncertain they are about their problem stage and about the availability of information to assist them in solving their problem?* This research question arises out of Wilson's (1999a) suggestion that the concept of "uncertainty" could be useful in categorizing the cognitive state of the client and that, if it could be operationalized, it might prove useful. Again, no formal hypothesis is proposed. The usefulness of the concept is determined by the extent to which clients could or could not identify their degree of uncertainty. However, one hypothesis (stated in null form in common with all of the hypotheses presented) links the two research questions:

**Hypothesis 1:** there is no difference in the degree of uncertainty expressed at the different problem stages.

We expect this hypothesis not to be supported, because it is proposed in the problem-solving model that uncertainty *decreases* as an information seeker proceeds through the problem stages.

The on-line search is an activity within the problem solving process, and the model suggests that uncertainty ought to be reduced following the search. This gives rise to:

**Hypothesis 2:** there is no difference in the level of uncertainty expressed by clients before and after the on-line search.

The central concepts of these first two research questions are also employed in formal hypotheses derived from the third research question:

- (3) How do the concepts of problem-solving stage and uncertainty relate to other variables identified in the models of information-seeking behavior proposed by researchers such as Ellis, Kuhlthau and Wilson? Wilson, in a series of models (1981, 1999a, 1999b), has suggested that (a) information-seeking behavior is stimulated by the recognition of a "problem" in the life-world of the individual; (b) that the resolution of the problem will involve a series of information-seeking activities (active search, passive search, passive attention, and on-going search) within the stages of problem solving; (c) that the extent to which information-seeking takes place will be affected by a number of demographic, psychologic, role-related, and other variables; (d) that risk/reward theory and self-efficacy theory may explain whether or not a person engages in information seeking; and, finally, that following the acquisition of information, information processing is required to apply the information found to the problem at hand.

We do not intend here to explore each of these elements; however, hypotheses are presented that relate to various issues raised in the models.

First, the contextual situation of the information seeker is explored by reference to the discipline within which he or she works, by the state of their existing knowledge of the problem area, the range of information-seeking activities in which they have engaged prior to seeking a mediated search, and their perception of the nature of the problem, measured here by the degree of comprehensiveness they believe to be necessary in the search.

Various researchers have explored the relationship between various aspects of information-seeking behavior and the discipline within which the person works (e.g., Mote, 1962; Palmer, 1991). Such research, however, has looked at the relationship between discipline and various information-seeking activities and we felt that it would be useful to determine whether a relationship existed at a more general level. If, for example, researchers in different disciplines were to present themselves for searches at different stages of the problem solving process, it would be important for the design of systems to discover why this is the case and what other, if any, information-seeking activities had been engaged in before they presented themselves.

The hypotheses that emerge from this analysis of the general and immediate context of the search are:

**Hypothesis 3:** There is no difference in the level of uncertainty expressed in relation to the problem stages by clients in different disciplines.

**Hypothesis 4:** there is no difference in the overall level of uncertainty expressed by clients with different levels of knowledge of their domain.

**Hypothesis 5:** There is no difference in the overall level of uncertainty expressed by clients with range of information-seeking activities engaged in before the search.

**Hypothesis 6:** there is no difference in the overall level of uncertainty expressed by clients requiring different levels of comprehensiveness in the search.

Among the variables that intervene between the recognition of a problem and the decision to engage in information seeking, that have been examined by previous researchers are demographic variables. For example, in relation to seeking health information, both age and sex have been identified as of interest (e.g., Connell & Crawford, 1988; Slevin et al., 1988). Both of these variables are examined in this research:

**Hypothesis 7:** there is no difference in the overall level of uncertainty expressed by the different sexes.

**Hypothesis 8:** there is no difference in the overall level of uncertainty expressed by clients in different age groups.

In the case of Kuhlthau's (1993a) model of the stages in information seeking, one of the key features is the use of the idea of an affective dimension in the information-seeking process. On the basis of qualitative research, Kuhlthau postulates that the different stages in the process are accompanied by the experience of different feelings. Here, two aspects of the stage model were explored: first, the relationship between uncertainty (as defined in this research) and the stages in Kuhlthau's model was examined:

**Hypothesis 9:** there is no difference in the overall level of uncertainty expressed by clients engaged at different stages of the Kuhlthau model.

The relationship between the problem-solving stages of Wilson's model and Kuhlthau's information-seeking stages is of interest. Our proposition is that the problem-solving model is more general, because it is intended to apply to all actions in problem solving; for example, it would be possible to analyze decision-making actions within such a stage model, as well as information-seeking action, while Kuhlthau's stages are intended to apply to information problems. A further point is that we believe that, for some complex cases, a problem solving stage may involve *all* of Kuhlthau's stages; for example, an information problem may consist of seeking to determine whether or not a possible research problem is genuinely such. In other words, the person is engaged in the problem definition stage. Having moved through Kuhlthau's stages in arriving at a satisfactory answer to that problem, the person is then faced with a new information problem or, indeed, set of information problems associated with the problem resolution phase.

Second, an attempt was made to relate the feelings identified by Kuhlthau to uncertainty:

**Hypothesis 10:** there is no difference in the feelings expressed by clients having different levels of uncertainty.

Among Kuhlthau's affective variables is the feeling of uncertainty, although, as noted earlier, she actually defines

uncertainty as a *cognitive* state. In this research uncertainty was related to the problem-solving stages and was understood as a cognitive variable. Hypothesis 11, therefore, was an attempt to examine the relationship between uncertainty as defined here and the total set of feelings variables postulated by Kuhlthau.

Ellis's (1989) behavioral activities are not presumed to constitute a stage process, in the contexts of problem solving and uncertainty and, if Ellis's proposition that these characteristics are common to all information seekers, we would not expect there to be any variation in the level of uncertainty expressed by those who specify that they are engaged in different activities, nor should there be any variation in activities in the different problem stages. This leads to the hypotheses:

**Hypothesis 11:** there is no difference in the overall level of uncertainty expressed by clients engaged in different activities as expressed in Ellis's behavioral model.

**Hypothesis 12:** there is no difference in the activities engaged in, as expressed in Ellis's behavioral model, at the different problem stages.

## Research Design

### *Data Collection*

All of the data reported here were collected in the course of interviews with those who requested an on-line search, in response to the project being advertised within the Universities of Sheffield and North Texas. The interviews were carried out before the on-line search, to determine the nature of the client's problem and to collect other information, after the search had been carried out, and, in the case of Sheffield only, some 2 months after the search. The interviews, which lasted between 20 and 90 minutes) employed a combination of open questions, for example, on the nature of the research problem; closed questions which presented a set of alternative response to a question; and instruments completed by the interviewee in the presence of the interviewer to collect data on, for example, the Ellis "characteristics" of the search process. A pilot test of the research instruments, involving 22 clients, was carried out in Sheffield at the beginning of the project and some changes were made to the instruments as a consequence. The results of the pilot test have been described elsewhere (Wilson, 1999b).

## Results

### *Problem-Solving Stage*

Following the presearch interviews, in which extensive accounts of problem situations were elicited,<sup>1</sup> clients were asked, "What stage are you at in terms of defining or resolving the problem, or in presenting the answer?" and

<sup>1</sup> An example of a problem description is given as Appendix 3.

TABLE 1. Problem-solving stages.

Problem Stage	Frequency	Valid percent
<b>Identification</b>	18	9.7
Mid-point	1	.5
<b>Definition</b>	74	39.8
Mid-point	8	4.3
<b>Resolution</b>	71	38.2
Mid-point	2	1.1
<b>Presentation</b>	12	6.5
Total	186	100.0
<b>Missing</b>	12	
Total	198	

given a list of the problem-solving stages with definitions (Appendix 1). Very few clients had any difficulty in fitting the progress of their work into one or other of the stages offered. All were given the opportunity to select an alternative position on the scale—generally between two of the stages—and only 11 did so. The results are shown in Table 1. As may be seen from the table, and not surprisingly (because we might expect the greatest intensity of information seeking to occur at these stages), the majority of clients located themselves in either the problem definition stage or the problem resolution stage.

The problem stage can also be used to show the progress of clients as they move through the information-seeking and use process. Figure 2 shows the presearch and follow-up stages of the clients who were involved in the study at both of these phases of the research. The anticipated shifts are shown to take place, that is, fewer clients were in the early stages and more were in the later at the time of the follow-up.

*Uncertainty*

Again, the vast majority of clients were able to use the scale presented to identify the state of their certainty/uncertainty for each stage of the problem-solving process and for

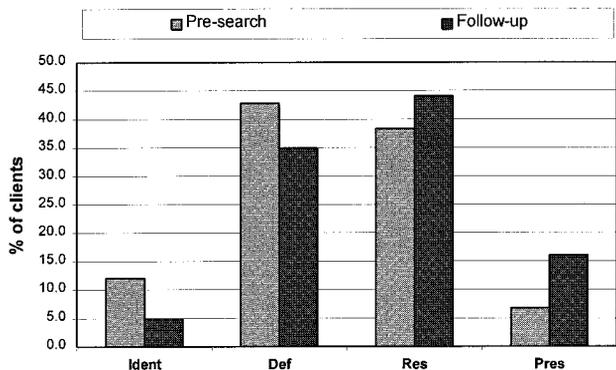


FIG. 2. Changes in problem stage.

TABLE 2. Distribution of results for uncertainty.

Score	Frequency	Valid percent
Relatively uncertain (0 to 3.9)	25	13.3
Relatively certain (4-8)	163	86.7
Missing	10	
Total	198	100.0

the likely availability of information sources (Appendix 2). For example, Table 2 shows the distribution of results for uncertainty about the definition of the problem. Appendix 2 shows the questions used to elicit certainty/uncertainty.

**Hypothesis 1:** *There is no difference in the state of uncertainty expressed at the different problem stages.*

This hypothesis proved difficult to test because of the highly skewed nature of the distribution—the greater majority of people were relatively certain about the progress of their work. However, when the mean value of the median uncertainty scores of people in different stages of the problem-solving process was calculated, the result shown in Figure 3 was obtained.

In the diagram, the higher the score, the more certain the individuals are, and it can be seen that a greater degree of uncertainty is experienced by those at the problem recognition stage, and a greater degree of certainty by those at the presentation stage.

Although the null hypothesis cannot be accepted or rejected on statistical grounds (because of the skewed nature of the data), the picture presented in Figure 3 allows us, perhaps, to suggest that uncertainty may be used to discriminate among information seekers.

**Hypothesis 2:** *There is no difference in the level of uncertainty expressed by clients before and after the on-line search.*

Clients were asked to state their level of uncertainty at three stages in the research—at the presearch interview, following the search, and in the later follow-up interviews. The result is best expressed graphically, as in Figure 4. As

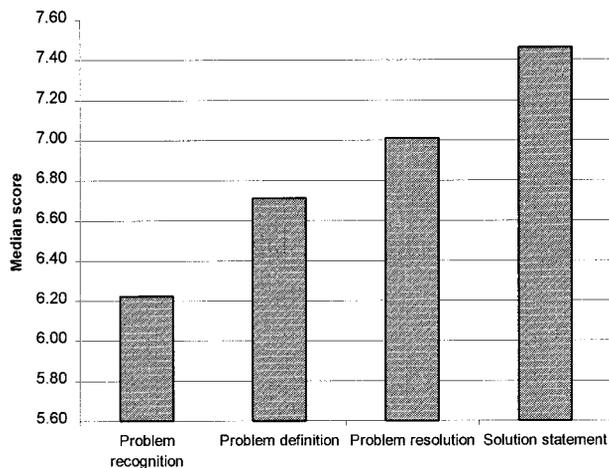


FIG. 3. Uncertainty and problem stage.

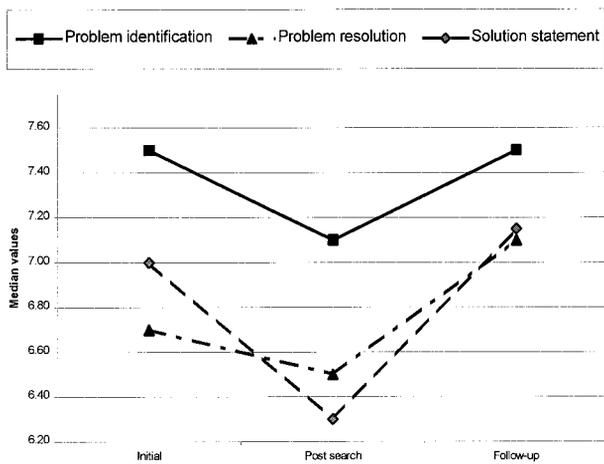


FIG. 4. Changes in uncertainty measures.

may be seen, the level of uncertainty actually increases immediately following the search and then falls again at the follow-up stage (usually 2 months following the search), when clients have had a chance to acquire and use some of the material drawn to their attention during the search. This effect is most pronounced for *problem identification*, *problem resolution*, and *solution statement*, and less so for *problem definition*, and *information availability*. The last two variables have been removed from the graph to show the effect more clearly. However, despite the apparently dramatic shift in the medians, the only ones that are statistically significant are those relating to Problem Identification, where both the shift between Initial interview and Postsearch interview and between Postsearch and Follow-up uncertainty are significant (Wilcoxon = -2.264 sig. = 0.014; and -3.216 sig. = 0.001). Although, in general, therefore, this hypothesis is supported, the data suggest that further exploration may be interesting.

**Hypothesis 3:** *There is no difference in the level of uncertainty expressed in relation to the problem stages by clients in different disciplines.*

This hypothesis is supported: no significant associations were found between the client's discipline and the level of uncertainty expressed for the problem-solving stages (as measured by "problem stage uncertainty," which was the mean value of the uncertainty scores for each of the stages) (chi squared = 1.796 sig. = 0.616). There was a slight, but not statistically significant, tendency for clients in pure science, medicine and engineering to express greater uncertainty about the availability of relevant information for their problem than those in the humanities and social sciences. Given the findings of researchers mentioned earlier (Mote, 1962; Palmer, 1991) it may be that the size of the groups, or their composition, or the specificity of the discipline affected our ability to discern more significant differences, and this may be a topic for continuing research.

**Hypothesis 4:** *There is no difference in the overall level of uncertainty expressed by clients with different levels of knowledge of their domain.*

Knowledge of the domain was examined in two questions: one asked: On the scale below how would you rank the state of your knowledge in relation to the broader domain to which your problem is related?

While the second asked: *On the scale below please indicate how you would rank the specific knowledge or expertise you feel you possess in relation to the problem at hand?*

In both cases a visual analog scale (an 8-cm line upon which the client marked a cross) was used: this type of scale was used for the majority of questions in the study.

This hypothesis is not supported: the Spearman correlation coefficient (rho) for the relationship between the level of knowledge of the broader domain and the sum of the uncertainty measures for the four problem stages was 0.313, which is significant at better than the 0.01 level. In other words, as the client's knowledge of the field increased, so did the certainty they expressed about the problem-solving stages. For specific knowledge of the problem area Spearman's rho was 0.366, which, again, is significant at better than the 0.01 level.

However, the relationship between level of knowledge and uncertainty about the probable availability of relevant information was *not* significant.

**Hypothesis 5:** *There is no difference in the overall level of uncertainty expressed by clients with range of information-seeking activities engaged in before the search.*

This hypothesis is also supported. Clients were presented with a list of information-seeking activities and asked to check each that they had engaged in: Did the searching involve (checklist for interviewer—client can indicate as many as necessary):

- \_\_\_(1) Web-based database(s) or information sources
- \_\_\_(2) on-line database(s), searching done by a librarian (intermediary)
- \_\_\_(3) on-line database(s), searching done on your own
- \_\_\_(4) printed index(es)
- \_\_\_(5) library catalog(s)
- \_\_\_(6) library collection without use of a catalog
- \_\_\_(7) own collection
- \_\_\_(8) colleagues' collection
- \_\_\_(9) other, please specify:

The responses were grouped and crosstabulated against two categories of uncertainty, high and low. No significant association was found between the variables. The chief reason for this finding was the relatively high degree of certainty expressed by most clients.

**Hypothesis 6:** *There is no difference in the overall level of uncertainty expressed by clients requiring different levels of comprehensiveness in the search.*

This hypothesis was not supported. Clients were asked to indicate on a visual analog scale the extent to which they needed the search to be comprehensive or selective (i.e., an indication of the level of recall required). The scores were grouped and crosstabulated against level of uncertainty. The chi-squared test was significant at better than the 0.05 level (i.e., 0.046), showing a moderate association—the tendency

TABLE 3. Feelings reported.

Feelings reported as	Feelings reported as . . .						
	Uncertain/ certain	Pessimistic/ Optimistic	Confused/Clear	Frustrated/ Relieved	Doubtful/ Confident	Dissatisfied/ Satisfied	Disappointed/ Pleased
Uncertain/certain	1.000	.585	.679	.618	.597	.606	.682
Pessimistic/Optimistic	.585	1.000	.342	.434	.546	.534	.613
Confused/Clear	.679	.342	1.000	.496	.435	.517	.535
Frustrated/Relieved	.618	.434	.496	1.000	.522	.507	.601
Doubtful/Confident	.597	.546	.435	.522	1.000	.648	.701
Dissatisfied/Satisfied	.606	.534	.517	.507	.648	1.000	.831
Disappointed/Pleased	.682	.613	.535	.601	.701	.831	1.000

was for those who were most certain to require greater comprehensiveness in the search (i.e., higher recall). We may assume that those who are most certain about their problem and progress through the problem-solving stages will wish to ensure that they do not miss critical information, whereas those who are less certain will look for high precision to enable them to clarify their understanding of the issue.

It may be thought that this proposition is in conflict with the findings of others (e.g., Spink, Bateman, & Greisdorf, 1999; Vakkari & Hakala, 2000) that those who are not familiar with a task or topic tend to accept more references than those who are more knowledgeable. However, those investigations deal with relevance judgements made at the time of the search, whereas this research involved a statement of preference for comprehensiveness *before* the search had been undertaken. Clearly, although a preference for comprehensiveness may exist, the actual search may result in few items being considered relevant, that is, the search has resulted in as high a degree of comprehensiveness *relative to the problem* as could be attained. It should also be noted that there are problem types, such as patent searches like that quoted in Appendix 3, where comprehensiveness is vitally necessary, regardless of the level of knowledge already held by the enquirer.

Moving to the demographic variables that may affect information-seeking behavior:

**Hypothesis 7:** *There is no difference in the overall level of uncertainty expressed by the different sexes.*

This hypothesis is supported: there is no significant association between the sex of the client and the overall degree of uncertainty expressed (chi squared = 4.096 sig. = 0.251).

**Hypothesis 8:** *There is no difference in the overall level of uncertainty expressed by clients in different age groups.*

This hypothesis is supported: there is no significant association between the age of the client and the overall degree of uncertainty expressed (chi squared = 0.087 sig. = 0.957).

As noted earlier, Kuhlthau's (1993a) model of the stages of information seeking is very important in the literature—it has become one of the most used and most cited models in the field. Consequently, it was of interest to examine the

relationship between this model and the problem-solving model proposed by Wilson (1999a). It must be borne in mind that Kuhlthau adopted a qualitative, interpretative research strategy for her work, while this research seeks to operationalize the variables she isolated in a quantitative study. Consequently, this work should not be taken as a test or attempted validation of Kuhlthau's work but as an attempt to discover relationships when the variables are used as here, and rather than those variables being derived from interviews with our clients, they were predetermined and listed for selection by them.

**Hypothesis 9:** *There is no difference in the feelings expressed by clients having different levels of uncertainty in relation to the problem-solving stages.*

This was an attempt to relate the measure of uncertainty at different problem stages to a set of affective variables, based on those identified by Kuhlthau as accompanying the search stages she had defined. A problem was experienced here in that a high level of intercorrelation of all of the feelings terms existed in the data: this is shown in Table 3.

Table 3 shows that the Uncertain/Certain variable is rather closely intercorrelated with all of the other feelings variables, and most of these are also highly intercorrelated with each other. The conclusion we draw is that all of the feelings variables are measuring essentially the same underlying variable. Taking the sum of the correlations, we find that the most highly intercorrelated variable is *disappointed/*

TABLE 4. Pre-search problem-stage uncertainty and Kuhlthau's feelings.

Component Matrix	Component	
	1	2
Pre Search Problem stage 1 uncertainty	.432	.638
Pre Search Problem stage 2 uncertainty	.537	.635
Pre Search Problem stage 3 uncertainty	.417	.473
Pre Search Problem stage 4 uncertainty	.413	.640
Feelings on Progress—Uncertain/certain	.809	—
Feelings on Progress—Pessimistic/Optimistic	.663	-.198
Feelings on Progress—Confused/Clear	.746	—
Feelings on Progress—Frustrated/Relieved	.717	-.287
Feelings on Progress—Doubtful/Confident	.754	-.181
Feelings on Progress—Dissatisfied/Satisfied	.814	-.291
Feelings on Progress—Disappointed/Pleased	.840	-.309

TABLE 5. Post-search problem stage uncertainty and Kuhlthau's feelings.

Component Matrix	Component	
	1	2
Post Search Problem stage 1 uncertainty	.475	.626
Post Search Problem stage 2 uncertainty	.536	.402
Post Search Problem stage 3 uncertainty	.517	.543
Post Search Problem stage 4 uncertainty	.600	.660
Feelings on Progress—Uncertain/certain	.774	-.251
Feelings on Progress—Pessimistic/Optimistic	.689	—
Feelings on Progress—Confused/Clear	.662	-.373
Feelings on Progress—Frustrated/Relieved	.655	-.365
Feelings on Progress—Doubtful/Confident	.777	-.206
Feelings on Progress—Dissatisfied/Satisfied	.781	-.261
Feelings on Progress—Disappointed/Pleased	.835	-.215

*pleased*, and we interpret the intercorrelations as expressing a *general affective state* that varies between positive and negative.

This proposition finds support from a factor analysis that was carried out on the data. The analysis explored the relationship between problem-stage uncertainty (PSU) and the Kuhlthau feelings variables. Two analyses were run using the PSU measures taken before and after the search. In the analysis of the presearch values, nine factors emerged, but only two of these had Eigenvalues greater than 1.0 and these two explained 61.0% of the variance. Table 4 shows the loadings on the two factors (components). These suggest the existence of two groups of clients: group 1 (signified by component 1) consists of those who load positively on all the PSU measures *and* positively on all of the feelings variables; group two consists of those who load positively

(and more highly) on the PSU measures, but negatively on five of the seven feelings.

Three factors emerged from the analysis of *postsearch* PSU and feelings, two of which, cumulatively, explained 61.5% of the variance. The third factor explained only 9.2%, had an Eigenvalue of only 1.01, and is therefore excluded from the analysis. The factor analysis shown in Table 5 again reveals a very similar situation to that described above. The values are very similar in both tables and the only differences are that in the postsearch data, group 2 clients load negatively on the uncertain/certain feeling, and not at all on the pessimism/optimism feeling.

To explore further the relations between feelings and uncertainty in the problem stages, the sum of scores across the four stages of the problem-solving model was obtained and correlated with the Kuhlthau uncertainty variable. This showed a high degree of correlation—Spearman correlation of 0.414, significant at the 0.01 level. This suggests that Hypothesis 9 is not supported.

However, the correlations between the level of uncertainty expressed at the *different* problem-solving stages and the other feelings were not always significant, as Table 6 shows.

The highly significant correlations (i.e., at the .01 level) are:

- (1) Uncertainty at the *problem recognition* stage is highly correlated only with *disappointed/pleased*;
- (2) Uncertainty at the *problem definition* stage is highly correlated with *confusion/clarity*, and *doubtful/confident*;
- (3) Uncertainty at the *problem resolution* stage is highly correlated with *pessimism/optimism* and *doubtful/confident*; and

TABLE 6. Correlations between Kuhlthau's feelings dimensions and uncertainty at problem stages.

	Pessimistic/ Optimistic	Confused/ Clear	Frustrated/ relieved	Doubtful/ Confident	Dissatisfied/ Satisfied	Disappointed/ Pleased
<b>Problem recognition uncertainty</b>						
Spearman's rho	0.208*	0.189	0.153	0.216*	0.230*	0.272**
Sig. (2-tailed)	0.045	0.069	0.146	0.038	0.027	0.009
N	93	93	91	93	92	90
<b>Problem definition uncertainty</b>						
Spearman's rho	0.189	0.361**	0.164	0.320**	0.246*	0.238*
Sig. (2-tailed)	0.071	0.000	0.123	0.002	0.019	0.025
N	92	92	90	92	91	89
<b>Problem resolution uncertainty</b>						
Spearman's rho	0.339**	0.228*	0.099	0.329**	0.220*	0.209*
Sig. (2-tailed)	0.001	0.029	0.353	0.001	0.037	0.050
N	92	92	90	92	91	89
<b>Solution statement uncertainty</b>						
Spearman's rho	0.205*	0.347**	0.148	0.299**	0.277**	0.302**
Sig. (2-tailed)	0.050	0.001	0.165	0.004	0.008	0.004
N	92	92	90	92	91	89

\* Correlation is significant at the .05 level (2-tailed).

\*\* Correlation is significant at the .01 level (2-tailed).

- (4) Uncertainty at the *solution statement* stage is highly correlated with *confusion/clarity*, *doubtful/confident*, *dissatisfied/satisfied*, and *disappointed/pleased*.

It may be that we have here two different ideas of uncertainty: *affective uncertainty*, (which is associated in this study with the other affective dimensions, such as pessimism/optimism); and *cognitive uncertainty*, associated with more rational judgements about the problem stages, with which it was associated in this study. This suggestion appears to have support from the fact that Confusion and Doubt may be more closely related to cognitive uncertainty than to the other feelings and hence have the higher correlations shown in Table 6.

**Hypothesis 10:** *There is no difference in the overall level of uncertainty expressed by clients engaged at different stages of the Kuhlthau model.*

This hypothesis is generally supported: the only significant relationship was between the *Presentation* stage and the problem stage uncertainty measure (PSU) (Spearman's  $\rho = -.228$  sig. = 0.005). Kuhlthau's concept of uncertainty as a feeling was also tested against the Kuhlthau stages and, again, the only significant association was with the *Presentation* stage (Spearman's  $\rho = 0.275$ , sig. = 0.008). We have suggested that the Kuhlthau stages may fit within the problem solving stages proposed by Wilson, in that they may describe a series of activities engaged in to complete a particular phase of the problem solving process. If this argument holds, we would *expect* the hypothesis to be supported.

Turning to the activities or characteristics of the search process described by Ellis:

**Hypothesis 11:** *There is no difference in the overall level of uncertainty expressed by clients engaged in different activities as expressed in Ellis's behavioral model.*

This hypothesis is supported: no highly significant associations between level of uncertainty in problem solving and the activity in which clients were currently engaged (browsing, chaining, etc.) were found, using Spearman's  $\rho$ . There was an association between those engaged in the *Verifying* activity and summed problem-stage uncertainty, but only at the 0.05 level.

**Hypothesis 12:** *There is no difference in the activities engaged in, as expressed in Ellis's behavioral model, at the different problem stages.*

Clients were asked: Which of the following activities is important in your information seeking at this stage?

- (1) Following chains of citations or other forms of referential connection between documents.
- (2) Browsing or semidirected searching in an area of potential interest.
- (3) Differentiating sources of information on the basis of the nature and quality of the material examined.
- (4) Maintaining awareness of developments in relation to this topic through the monitoring of particular sources.
- (5) Systematically working through a particular source to locate material of interest.
- (6) Verifying or checking the accuracy of information.

This hypothesis was generally supported: clients identified all of the behaviors shown in Ellis's model as important in all of the different stages. However, there were *some* differences.

Using the mean as a ranking element, most significance was attached at the first stage of the problem-solving process to *Browsing* and *Maintaining awareness*; in the second stage, to *Browsing* and *Differentiating*; in the third stage, to *Maintaining awareness* and *Chaining*; and in the fourth stage, to *Maintaining awareness* and *Browsing*. Within the problem-solving stages there were very few significant differences between the pairs of means: in fact, in phases two and four, no significant differences. In phase 1, the mean score for the importance of *Maintaining awareness* was significantly different from that for the importance of *Verifying* ( $t = 2.706$ , sig. at the 0.02 level), while in phase 3, the mean score for the importance of *Differentiating* was significantly different from that for the importance of *Verifying* ( $t = 2.437$ , sig. at the 0.02 level) and the mean score for the importance of *Maintaining awareness* was also significantly different from that for the importance of *Verifying* ( $t = 2.024$ , sig. at the 0.05 level).

The support for these two hypotheses suggests that people need to engage in this set of activities if they are seeking information, regardless of their state of certainty/uncertainty, thereby supporting Ellis's proposition that they are general characteristics of the search process.

## Discussion

This research has shown that the concept of uncertainty can be operationalized so that information seekers can express the degree of uncertainty they have regarding the stages of the problem-solving process in which they are involved, and in relation to their assessment of the likelihood of information being available. It has also shown that the idea of a stage model of the problem-solving process is readily understandable for academic researchers—although some prefer to use the word *issue*, rather than *problem*.

The results presented here suggest that differences in the level of uncertainty experienced by individuals are not related to sex, age, or discipline, nor does discipline affect the problem-solving stages identified by clients. Nor does the extent of prior information seeking appear to relate to the level of uncertainty expressed. Significant differences have been found, however, in the relationship of level of uncertainty to the problem stage in which the client is engaged, as well as to the knowledge the client has of the domain he or she is exploring. The problem-solving model forecasts the first of these significant relationships, as it suggests that as the client proceeds through the stages, s/he will engage in specific information-seeking acts to reduce *uncertainty*.

The relationship between *uncertainty* and *knowledge of the domain*, is also expected, if we think of *cognitive uncertainty*—the greater the knowledge of the domain pos-

sessed by the client, the less likely s/he is to express uncertainty, especially uncertainty about problem identification.

The measures of uncertainty were strongly intercorrelated in both the initial interview and in the reassessment following the search. In fact, the intercorrelations were higher in the second stage than in the first, and in both cases were significant at the 0.01 level. The only exception was that, at the initial interview, the correlation between uncertainty at the problem identification stage and uncertainty about the availability of information was not significant. However, it *was* strongly significant at the follow-up stage.

The suggestion made above that *uncertainty* may have both affective and cognitive dimensions is of importance in clarifying the concept and perhaps in carefully defining one's use of the term in research investigations. Realization of the difference arose in this analysis as a result of the use of the term in two different questions, which had different contexts—one affective, associated with the other affective terms used by Kuhlthau, and one cognitive, associated with (presumably) rational judgements on the progress of one's research.

In general, as Wilson (1981) suggested and as Kuhlthau (1993a) confirmed, feelings may be associated with information-seeking behavior: Wilson had suggested that affective factors might give rise to needs that could be satisfied (at least in part) by information-seeking behavior, while Kuhlthau showed that affective variables could be associated with the stage of the search engaged on by the user. We have not discovered the same degree of correspondence between feelings and stages as suggested by Kuhlthau, but that may be a function of the way the variables were employed in the investigation. At least it is now possible to refine the instruments presented here, perhaps in more sensitive ways, to investigate the cognitive and affective relationships more closely. These ideas have not been fully explored in this article but the results reported in the discussion on Hypothesis 9 above suggest that further exploration may be worthwhile. Further analysis of the present data from this perspective is now being undertaken.

## Conclusion

Information seeking and searching involves a complex set of activities on the part of the individual seeking information. Much light has been cast on the activities involved through largely qualitative investigations and useful models of the overall process that are proposed by Kuhlthau, Wilson, Ellis, and others. Such research, however, generally involves rather small numbers of people, and the researchers generally make no claim that the results can be generalized across the population at large.

It is important, however, to try to move to a position where a more general understanding of human behavior is achieved, because the design and development of information systems is generally dependent upon information of this kind.

We have tried to operationalize some of the variables evolved in qualitative investigations and to apply them in a largely quantitative investigation. Inevitably, some of the richness of interpretative research is lost but, at the same time, the value of that research is increased if it is shown that the variables can be used in studies involving more people.

It can be claimed, on the basis of this work, that the concepts of problem-solving stages and uncertainty are sufficiently robust and sufficiently discriminating among individuals to be employed in further investigations.

## Acknowledgments

Our thanks to Professor Peter Willett and anonymous referees for reading and commenting on an earlier draft of this article. We would also like to thank the British Library Research and Innovation Centre for its support for the Sheffield study.

## References

- Belkin, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5, 133–143.
- Borgers, R., Mullen, P.D., Mertens, R., Rijken, M., Eussen, G., Plagge, I., Visser, A.P., & Blijham, G.H. (1993). The information-seeking behaviour of cancer outpatients—A description of the situation. *Patient Education and Counseling*, 22(1), 35–46.
- Bos, C.S., Van Dijk, H.K., & Mahieu, R.J. (1999). Relevance of alternative modelling decisions for hedging currency risk. *Proceedings of the Conference on Inference and Decision Making*, Erasmus University, Rotterdam, 17–19 June 1999.
- Connell, C.M., & Crawford, C.O. (1988). How people obtain their health information: A survey in two Pennsylvania counties. *Public Health Reports*, 103, 189–195.
- Ellis, D. (1989). A behavioural approach to information retrieval design. *Journal of Documentation*, 46, 318–338.
- Harris, R. (1998). Introduction to decision making. Home page: <http://www.vanguard.edu/rharris/crebook5.htm>. [Visited 14 October 2000].
- Henkind, S.J. (1986). Imprecise meanings as a cause of uncertainty in medical knowledge-based systems. In J.F. Lemmer & L.N. Kanal (Eds.), *Uncertainty in artificial intelligence 2* (pp. 35–41). Amsterdam: North Holland.
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 3–50.
- Kuhlthau, C.C. (1993a). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.
- Kuhlthau, C.C. (1993b). A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4), 339–355.
- Kuhlthau, C.C. (1996). The concept of a zone of intervention for identifying the role of intermediaries in the information search process. In *Global complexity: Information, chaos and control*. ASIS 1996 Annual Meeting, Baltimore, October 19–24. Electronic proceedings.
- Kuhlthau, C.C. (1997). The influence of uncertainty on the information seeking behaviour of a securities analyst. In P. Vakkari, R. Savolainen, & B. Dervin (Eds.), *Information seeking in context: Proceedings of an international conference on research in information needs, seeking and use in different contexts*, 14–16 August, 1996, Tampere, Finland. London: Taylor Graham.
- Kuhlthau, C.C. (1999). Accommodating the user's information search process: Challenge for information retrieval system designers. *Bulletin of the American Society for Information Science*, 25 (3) [Available at:

- <http://www.asis.org/Bulletin/Feb-99/kuhlthau.html>. Visited 24th October 2000].
- Mignerey, J.T., Rubin, R.B., & Gorden, W.I. (1995). Organizational entry—An investigation of newcomer communication behaviour and uncertainty. *Communication Research*, 22(1) 54–85.
- Moore, P., & Thomas, H. (1988). *The anatomy of decisions*. Harmondsworth: Penguin.
- Mote, L.J.B. (1962). Reasons for the variations in the information needs of scientists. *Journal of Documentation*, 18(1), 169–175.
- Palmer, J. (1991). Scientists and information: I. Using cluster analysis to identify information style. *Journal of Documentation*, 47(2), 105–129.
- Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Slevin, M.L., Terry, Y., Hallett, N., Jefferies, S., Launder, S., Plant, R., Wax, H., & McElwain, T. (1988). BACUP—The first two years: Evaluation of a national cancer information service. *British Medical Journal*, 297, 669–672.
- Spink, A. (1996). A multiple search session model of end-user behaviour: An exploratory study. *Journal of the American Society for Information Science*, 46(8), 603–609.
- Spink, A., Bateman, J., & Greisdorf, H. (1999). Successive searching behavior during information seeking: An exploratory study. *Journal of Information Science*, 25(6), 439–449.
- Targett, D. (1996). *Analytical decision making*. London: Pitman.
- Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, 56(5), 540–562.
- Van Rijsbergen, C.J. (1986). A non-classical logic for information retrieval. *Computing Journal*, 29(6), 481–485.
- Van Rijsbergen, C.J. (Ed.). (1998). *Information retrieval, uncertainty and logics*. Dordrecht: Kluwer Academic Publishers.
- Wilson, T.D. (1981). On user studies and information needs. *Journal of Documentation*, 37, 3–15.
- Wilson, T.D. (1999a). Models in information behaviour research. *Journal of Documentation*, 55(3), 249–270.
- Wilson, T.D. (1999b). Exploring models of information behaviour: The “Uncertainty” project. In T.D. Wilson & D.K. Allen (Eds.), *Exploring the contexts of information behaviour: Proceedings of the second international conference on research in information needs, seeking and use in different contexts*, 13/15 August 1998, Sheffield, UK (pp. 55–66). London: Taylor Graham.

## Appendix 1: Question on Problem Stages, with Definitions

What stage are you at in terms of defining or resolving the problem, or in presenting the answer?

Please respond using one of the categories indicated below. If no single category sums up where you are in your work please indicate the approximate position.

*[Interviewer to probe if client doesn't specify one of the categories below to determine problem phase]*

*Problem recognition:* I am trying to determine whether or not the topic I'm interested in is a real problem from the point of view of the discipline or area that interests me. I need a search so that I can discover whether others have identified the same issue as problematical.

*Problem definition:* I have identified a real problem and now need to define it more closely or carefully so that I can determine how to approach the problem and how it relates to other problems in the field. I need a search to help me define my research objectives.

*Problem resolution:* I am in the process of resolving the problem (engaged in laboratory experiments, field work, etc.) and now need a search to enable me to proceed with and complete that work. The question deals with a particular problem that I need to resolve in completing the work or with aspects of methodology, research approach or research methods.

*Solution statement:* I have effectively finished the work I was doing and am either tying up loose ends, or finding out from related work how best to report my research or where best to report it.

## Appendix 2: Questions on Uncertainty

How certain are you:

(a) that you have recognized a real issue to investigate?

Very uncertain |—————| Very certain

*(Prompt for uncertainty question (a) a real problem is—A problem that you think needs to be solved)*

(b) that you have defined the topic appropriately?

Very uncertain |—————| Very certain

(c) that the issue *can* be resolved?

Very uncertain |—————| Very certain

(d) that an effective way of presenting the results can be found?

Very uncertain |—————| Very certain

(e) that relevant information is available and can be found?

Very uncertain |—————| Very certain

## Appendix 3—Example of a Problem Statement Elicited in the Presearch Interview

### *User's Description of the Problem*

The problem we are interested in understanding how the enzymes, DNA methylase transferases, work, and how they methylate DNA and what are the consequences. But the thing we are interested in also is how we can develop ways of testing for genetic diseases. The problem is that you can find information about the methylase transferases enzymes and about methylation, but it is difficult then to bring this together with information about genetic diseases. So you can have a search for these things but pulling them together is difficult. One of the other things that have become apparent to me is that, particularly in America, labs have actually started patenting rather than publishing. It is now possible, as I discovered very recently, to search in the IBM database for American patents. What is difficult is to search for any patents in Europe or the UK free of charge, the second problem is that on-line searching only covers abstracts and, being someone who avoids the library as much as possible, it would be nice to use some more powerful or bigger search engine, something like DIALOG that would give you the option to recover the full text of it, particularly electronically, although sometimes document delivery would be useful.

So the problem is understanding how this particular class of enzymes work. Then it is to understand how we can make use of them in developing a test for genetic diseases. And then it is understanding which genetic diseases are out there that are of global significance—there are new genetic disorders discovered every week, but they will only affect a very small proportion of the population. The most important ones, things like cystic fibrosis is a good case that affects a fairly large number of people, but it is still a fairly small number globally. So we try to think of genetic tests that would cover lots of diseases rather than specific diseases, because there really is not enough money to test for more than five genetic diseases in any one person. So it is kind

of a lot of disparate areas that we are trying to look at. Whereas I might be, let's say, an expert in DNA methyl transfases and know where to look for that information, I don't necessarily know where to look for the other information. I can use my experience of looking for information in one area of biology but I am bound to miss things, and the worry is that I miss really important things, and I know that I miss things because when I scan some articles by chance I might find a reference to something that was highly relevant, and it was just a chance finding. The number of times that you look up an article and find an article next to it says that there must be a lot of information out there that just is not picked up.