

DOC-CR-SP-8

Queen

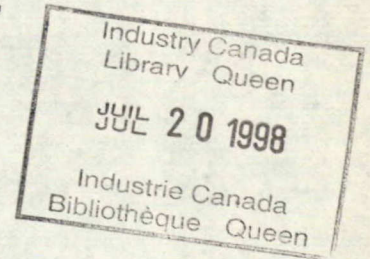
P 91

C655

G7885

1981

2
THEORY AND METHODOLOGY FOR RESEARCH INTO BROWSING
WITH SPECIAL REFERENCE TO VIDEOTEX DATA BASES



1
Author and Contractor: T. Grusec, Ph.D., C.Psych.
Behaviour Research Associates
88 Kenilworth Ave.
Ottawa, Ont., K1Y 3Y7
telephone: (613) 729-1643

Department of Supply and Services Contract No. 5ER.36100-1-0310

Department of Communications Scientific Authority:

Thomas Whalen, Ph.D.
Behavioural Research Group
300 Slater St., room 1662
Ottawa, Ont., K1A 0C8
telephone: (613) 996-8871



March 31, 1981

THEORY AND METHODOLOGY FOR RESEARCH INTO BROWSING
WITH SPECIAL REFERENCE TO VIDEOTEX DATA BASES

| | |
|---|----|
| 1.0 Objectives | 1 |
| 2.0 Introduction, method of approach, and summary of report | 2 |
| 3.0 What is browsing? | 4 |
| 4.0 Research on browsing in libraries | 8 |
| 5.0 A verbal report model | 12 |
| 6.0 A browsing research model with dependent variables | 17 |
| 7.0 Independent variables for studying browsing in videotex data bases | 22 |
| 8.0 Additional dependent variables for studying browsing | 24 |
| 9.0 Bibliography | 25 |
| 10.0 Appendix - Verbal reports as data, Ericsson & Simon | 26 |

1.0 OBJECTIVES

1. To develop a methodology for studying the use of videotex data bases or index systems, such as Telidon, for the purpose of browsing.
2. To describe, in detail, the techniques for measuring the dependent variables in studying videotex browsing using the developed methodology.
3. To describe one or more experiments to determine the effectiveness of videotex data bases for browsing using the methodology and measurement techniques.
4. To justify the methodology and the measurement techniques by providing a bibliography of publications in which this or similar methodologies and measures were used.

2.0 INTRODUCTION, METHOD OF APPROACH, AND SUMMARY OF REPORT

A general search of the psychological literature disclosed nothing of interest that was relevant to the topic of browsing or to subject matters that might be related to this concept.

We then turned to the field of library science where it was suspected that there ought to have been interest and research on the topic. With the very kind help of Professor Ann Schabas of the Faculty of Library Science, University of Toronto, highly relevant literature was found. The most interesting material was in works by Richard J. Hyman.

In section 3.0, we examine the varied connotations of the term "browsing" as outlined in the library science literature. We arrive at a conceptualization in which browsing is viewed as an integral part of all information search and is not clearly distinguishable from directed search. Overt behaviour alone is insufficient for understanding browsing since moment to moment cognitive intentions underlying the ongoing behaviour must be understood.

We examine the research on browsing in libraries in section 4.0. The methodologies used in that research were not found useful for our purposes since they attempt to focus on overt behavior alone, and, at that, not on moment to moment behaviour, or else they deal with cognitions which are too far removed in time and place from the behaviour.

In section 5.0, we outline a verbal report model recently presented by Ericsson and Simon. That model directly provides a technique, "thinking aloud", which, when aligned with ongoing browsing behaviour, provides us with the key dependent variables necessary for research into browsing as we have conceptualized it.

The verbal report model clearly specifies the conditions under which we may obtain veridical reports about ongoing cognitive processes.

A descriptive browsing research model is presented in section 6.0, including an outline of how to record the dependent variables, especially the key ones related to ongoing cognitive Processes. The non-standard nature of the data is discussed.

An unlimited set of independent variables for studying browsing may be plugged into the browsing research model. Some of these are mentioned in section 7.0. An innovative research program on browsing may be instituted.

Additional dependent variables are considered in section 8.0. These are quite different from the major ones considered previously. But they may be tacked on to the end of any browsing experiment at low cost and effort and may yield additional data of heuristic value.

A bibliography of the relevant publications is presented in section 9.0. Because of its central importance in providing the key dependent variables for browsing research, a copy of the Ericsson and Simon report is included as section 10.0.

3.0 WHAT IS BROWSING?

Hyman (1972) informs us that the etymology of "browsing" is not entirely clear. Many scholars and most dictionaries trace it to Middle High German and Old French terms ("brouz") which, as nouns, refer to "young shoots" of trees or shrubs. The act of feeding on these (as in "grazing") is browsing. The connotation was of casual as opposed to serious feeding. Metaphorical extension to casual perusal of books first occurred in the early to middle 19th century. Common useage has extended the term beyond the context of books and reading matter to almost any activity in which searching may be involved.

In the library context, Hyman begins with a tentative definition of browsing as "a patron's random examination of library materials as arranged for use." He then goes on to examine the great ambivalence about the nature and value of browsing which has dogged the term from the beginning and persists even today among some librarians. We may discern at least three dimensions.

The first of these is an evaluative one - "good - bad". At the negative end, browsing was seen by some in the 19th century as "self indulgence by the untutored in objectionable works", the surreptitious glancing through dirty books and the like. At the positive end, browsing was seen as beneficial self-education for the general reader. Many feel that it is part of the art of reading, part of a life adventure in learning.

Another dimension is the one of seriousness. Here, at one end we have superficiality, capriciousness and aimlessness, and at the other, the view that browsing may be a valuable tool for the serious scholar with serendipidy as one of the fallouts.

A third dimension we discern is the one of degree or kind of certainty which underlies browsing. We may have a relatively

high degree of unpredictability of browsing outcome as in some kinds of shopping activities. At the other end is highly purposeful search where something relevant is expected to be found with high probability although the exact object is not known in advance.

Hyman finally opts for a functional definition: "browsing is that activity ... whereby materials ... are examined in the reasonable expectation that desired or valuable items or information might be found among those materials as arranged ..." Browsability then becomes that characteristic of the arrangement of the materials which supports the expectation. (Hyman, 1972, p.131).

Unlike librarians, whose views vary widely, information scientists have more consistently treated browsing as a serious phenomenon. One view which we found most congenial is that of Malcolm Rigby.

In every human activity involving the quest for objects, information, or knowledge, two approaches may be used, one the direct or specific approach and the other the indirect or exploratory approach.

Either approach may be equally rewarding, although not necessarily so in any given instance, or for any given person. But neither can be ignored in overall planning for education or research, or in designing library or information-retrieval systems, without drastically inhibiting or even frustrating progress or achievements. Most activities involve, however, a combination of the two approaches in widely varying proportions. (quoted in Hyman, 1972, p.116-117)

Similar views of many others might have been cited but Rigby expressed matters very well and in a way which pertains directly to videotex data bases, such as Telidon. To date, most research on Telidon as a data base has focussed on the directed search aspect only, and has looked exclusively at retrieval efficiency when a particular item or page, specified in advance, was sought.

The last sentence in the quotation from Rigby is especially ger-

mane, stating that directed search and browsing co-exist in temporal simultaneity, although in widely varying proportions from time to time. This means that while browsing may, at times, be a predominant activity, elements of browsing are probably present to some degree even under directed search conditions. Some of the outcome of such elements, of course, would be scored as "errors" in the usual research strategy for directed search which indirectly instructs an experimental observer to inhibit browsing. In other words, realistic research on the way people use data bases has not been done. This is not to denigrate the experiments on directed search which have a very sound rationale of their own, but only to reinforce the importance of examining browsing in the videotex context.

So, what is browsing? The conceptual view we take is that browsing may be an integral part of all information search, in any context, but it may vary not only in general predominance in a particular situation but also in its role from moment to moment in that situation as a person examines informational stimuli. From this it follows that observing and recording overt behaviour alone does not enable us to distinguish between browsing and directed search in the particular mix of the two that is guiding overt behaviour at any moment in time. Some experimental control over browsing can be exerted by instructions to observers. But the overt behaviour alone may tell us little about what the observer is actually doing at a given moment.

This means that there is no such thing as "browsing behaviour" which is any any way different from "search behaviour" or "non-browsing behaviour". In essence, we contend, browsing is cognitive intention aligned with overt behaviour. I may be looking for a very specific item or just looking for anything interesting. The looking is the same. Only my intentions differ. And cognitive intentions may vary from moment to moment as the person interacts with informational stimuli. We are thus in muddy waters with no clear distinction between browsing and directed search.

The key concept is cognitive intentions. Since these may vary from moment to moment we will refer to them as momentary intentions which guide or motivate behaviour. Intentions can only be studied by verbalizations of the observer. But verbal report has a mixed history in psychology with much residual ambiguity about its status. Fortunately, recent theoretical and empirical work has clarified the nature of verbal report considerably and in section 5.0 below we will present a verbal report model which provides the very tools we need for studying momentary intentions in browsing.

Before we get to that section, however, we will briefly examine the methodologies used by librarians in studying browsing. The reader may anticipate that these methodologies are of little use for our purposes. However, these attempts are of interest in further clarifying our conceptual approach to browsing and for the general light they shed on methodological issues.

4.0 RESEARCH ON BROWSING IN LIBRARIES

The term "browsing" is not generally found in topical indexes. This is true not only in psychology, but also in library science literature. Hyman thinks this is both because of the relatively recent metaphorical adoption of the term and because of some of the negative connotations of the activity as a wasteful use of resources. Even today some librarians express such negative views at least about some aspects of browsing.

However, it is an active topic although an invisible subject, and is implicit in the concerns and debates of librarians over classification schemes, open shelf access, the nature of catalogs, library use aids and other matters. Index entries such as "direct shelf approach", "books and reading", "library useage", "stack use", and others are often guises for the browsing concept.

Browsing is a central and explicit theme in Hyman's work within the larger context of shelf classification. The costs of creating and maintaining open shelf library systems is enormous. Were the library user not permitted stack access, much more efficient use of resources would be possible. For example, books could be grouped by size. Storage space would thus be minimized with greatly increased density of books on shelves and much saving in real estate costs. The New York Public Library uses a size scheme of this type for part of its reference collection. Only the possibility of meaningful browsing, with materials spatially arranged for their interrelatedness justifies open access and the attendant classification systems.

Thus a continuing concern of librarians is whether actual user behaviour justifies the costs and inefficiency of open shelf access. Library research has been undertaken to provide answers to such questions.

Among the methodologies used in this research were several which

are not directly related to individual user behaviour but only to the collective outcome of such behaviour in the statistical sense. These include citation analyses, circulation studies and mathematical models. A second category of studies are ones in which responses are obtained from individual library users, but, rather than being accessed in the library situation, the users were simply asked to reflect upon their past library use in a structured way. Methodologies here were primarily survey questionnaires.

Potentially more interesting to us is a third category in which library users were accessed in the library situation and were asked about some aspect of current library behaviour ie. about actual behaviour in which they were engaging or had very recently performed. In this third category, face to face interviews were used, or questionnaires were attached to books, or diaries and questionnaires were designed to get information about the next several books which the user accessed.

Even in the third category, however, information gained about browsing was not very revealing. To be fair, many of the researchers were aware of flaws in their methods and set forth their conclusions in an appropriately tentative way. The nature and importance of browsing as revealed in the various studies, with few exceptions, tended to ascribe some but minimal importance to the activity. Hyman (1972, 1980) presents thorough and cogent critiques of these studies in his reviews. Generally, none of the studies were sufficiently revealing or convincing to change any librarians' preconceptions, whatever they might have been, about classification, open access or other library practises.

Some of the researchers were quite aware of basic methodological problems in attempting to study browsing. For example, we cite Fussler and Simon (quoted in Hyman, 1980, p.11-12)

In studying non-recorded use, it is difficult to define the unit of behaviour that will be counted as "use" in a manner that the unit is unambiguous and practicable to count ... an observer might follow a reader and observe his behaviour. But the behaviour will almost surely be affected by the presence of the observer. Furthermore

this technique runs into either exorbitant cost or difficult sampling problems ... Nor does there seem to be a satisfactory mechanical or electronic method of describing browsing ... motion-picture or closed circuit TV (is) impractical since cameras would have to be placed either in many different areas of the library or on the reader's head. Furthermore such techniques are unlikely to tell us which books were used and which were not.

These authors go on to suggest that "touching the book" may be the best countable unit but even then, the writers realized that one would need to know something about the value or importance of a touch and this may not be revealed in the behaviour.

This citation from Fussler and Simon reveals a number of conceptual and empirical problems which are shared by most of the library researchers. But these problems are clarified by our conception of browsing presented in section 3.0 above as well as by general principles of psychological research.

1. There is a fear of disturbing the library user in a natural setting. Instead, a full fledged experiment is required with the experimental observer fully cognizant of his or her participation. Generalizability is an empirical matter.
2. There is an implication that a large number of observations are required on a large number of users. Operant behaviour and other approaches have clearly paved the way for intensive and meaningful studies on small numbers of experimental observers.
3. There is conceptual confusion between browsing and non-browsing. We contend that the distinction is meaningless. The focus should simply be on an experimental users interaction with library materials. The users motives at each moment in time may be determined quite clearly as the verbal report model of the next section will show.
4. Finally, there is the insistence that overt behaviour must

must reveal everything. We have already presented the case that this is fallacious.

In talking about the necessity of taking "value" into account, Fussler and Simon approach, in a somewhat distant way, our notion of taking cognition into consideration but they do not see how to do it.

How to do it is the subject of section 5.0 which we present next. It is of interest that Hyman concurred with the views of another author that fundamental research on the mental associations of library users while using the stacks might be pertinent. He wrote, "such research would be more likely pursued by specialists in nonlibrary disciplines, eg. neuro-physiology and the psychology of learning. Results might be of considerable importance for librarians." (Hyman, 1972, p.393)

5.0 A VERBAL REPORT MODEL

We introduced the idea of momentary intentions in section 3.0 above where we presented our conception of what is browsing. In the present section we face the problem of verbal report since momentary intentions can only be known by verbalization. We draw directly on a recent paper by Ericsson and Simon (1980). This paper is of central importance in providing the needed techniques for measuring the dependent variables in browsing research as we conceptualize it. Accordingly, we have included a copy of the original article as section 10.0 below. Our presentation here selects out only those aspects of the article which apply directly to the browsing problem.

In their paper, Ericsson and Simon address some problems of verbal reports which have not been properly treated in psychology. For good historical reasons, one of the major and long term impacts of behaviourism has been to relegate virtually all verbal reports to a questionable status. Despite the fact that verbal reports were always admissible as data, and were essential in studies of verbal learning and memory, no distinctions were generally made among categories of responses such as thinking aloud, after-the-fact reports of cognitive processes, and introspections of trained observers. All of these and other kinds of verbalizations tended to be dismissed as, at best, possibly useful sources of hypotheses, and, at worst, totally useless and especially so as verification tools. The authors' model attempts to clarify the mechanisms which generate verbal reports so that sensible distinctions can be made among the various kinds of verbal reports as to their probable validity as reports of cognitive processes. Indeed, after a critical examination of a large volume of experimental literature, a subset of which is presented in their article, carefully selected "...not to screen out studies that are troublesome for our theoretical framework", the authors find impressive support for their model.

Ericsson and Simon postulate three conceptual memory systems; a sensory system, a short term memory (STM) and a long term memory (LTM) system. No contentious theoretical issues are involved in this memory systems postulate so that it doesn't matter whether STM is or is not a subset of LTM rather than separate, or whether the contents of STM decay in time unless rehearsal occurs, or only when replaced by new incoming items.

The sensory system has a very short temporal duration, so that the input decays very rapidly. With the aid of information in LTM, sensory stimuli are directly recognized with great speed so that intermediate stages of processing between sensory input and LTM information retrieval are not stored in STM. However, the end product, "recognition", is stored in STM. The LTM system, of very large capacity, is accessed by recognition processes (as above) or by association processes. The latter are much slower temporally than is direct recognition. But the end products of the association processes are also, like recognition, stored in STM.

Intermediate stages of processing within association processes may or may not be stored in STM also (eg. the steps in searching for a name when its retrieval requires active search). The STM system is of limited capacity, holding only a few chunks (familiar patterns) of information at any one time. As new information enters STM, information previously there may either be lost or else, by a relatively slow re-assembly and fixation process, may be transferred into LTM as new (to LTM) chunks of information which are stored relatively permanently.

Information currently or very recently attended to is in STM and only the most recently attended to information is directly accessible there by verbal report. A portion of STM always becomes fixated in LTM before it is lost, although usually in some reprocessed form, and this portion can later be retrieved by bringing it back into STM where it can be attended to and reported verbally (in its reprocessed form). Thus, all verbal reports about ongoing cognitive processes are about the contents of STM which are most recently attended to.

The above leads to a taxonomy of verbalization about cognitions in which the critical variables are (1) the elapsed time between the original presence of contents in STM about which verbal report is requested and the time of verbalization and (2) the nature of the mapping between attended to information units and the size of the unit about which verbal report is requested, varying from one-to-one, to many-to-one or other. It is one-to-one where the verbalized unit corresponds directly to the attended to unit. Where the verbalization summarizes, or is about, several attended to units, the mapping is many-to-one. The following table sets out the taxonomy.

Types of verbalization processes as a function of time of verbalization and of the mapping between attended to units and verbalized units of information

| | <u>attended to/verbalized mapping units ratio</u> | |
|-----------------------------------|---|---|
| <u>time</u> | <u>one-to-one</u> | <u>many-to-one or other</u> |
| while information is still in STM | concurrent talking or thinking aloud | intermediate inference and generative processes |
| after information has left STM | retrospective probing | generalized and interpretive reports |

There is also the phenomenon of automation. This occurs with continued practise of particular processes. Intermediate steps which might be carried out in STM at early stages of practise and thus be available then for verbal report, might, later on, when well practised, occur without inputs and outputs to STM. This automation greatly speeds up cognitive processing and makes the intermediate steps unavailable for verbal report since they then do not have representation in STM.

Another important consideration concerns the nature of the information chunks in STM. Where these chunks are already in an aural mode, and where a task being processed in STM is not so difficult as to produce overload, then a direct and easily accomplished externalization of STM contents via verbal report is possible without making burdensome demands on the STM system - or, at worst, verbalization may slightly slow down the speed of performance of the primary cognitive task being carried out in STM. But, even in the latter case, verbalization of the "thinking aloud" variety will not change the information attended to (ie..no interaction) so long as that information is in an aural mode. (The writer should note here that he is not entirely clear how one distinguishes between aural and non-aural.)

However, when that information is in a non-aural mode (eg. pictorial information, or information about perceptual-motor processes) then a recoding of STM contents into an aural mode must first occur before a verbal report can be made. Such recoding will make some demands on processing capacity and time. In this case then, the primary cognitive task being accomplished in STM may take priority and some attended to information may not be verbalized either readily or at all.

This model thus makes explicit the conditions under which a verbal report about cognitive processes can be expected to yield a veridical and valid dependent variable statement of those processes as moment to moment events. These conditions are that:

1. the cognitive processes of interest are in an aural mode in STM and not in a perceptual-motor or other non-verbal mode.
2. the verbalization does not require an additional kind of processing which would not occur in the absence of verbalization requirements. Violation of this rule would

occur if scanning or selective filtering of STM contents were required or if the verbal report required comment on unattended to aspects of the primary cognitive task performance. These cases would require additional inputs from sensory and LTM systems before verbalization could occur.

3. the verbal report of processes should occur as close to the time the processes are attended to in STM (concurrently) and not at some later time, after the fact (retrospectively).
4. no constraints should be placed on the verbalization, eg. by providing forced choice alternatives or probes. Verbalization should not summarize STM contents or put them into some form that is different from that in which they reside in STM. It should be entirely free, of the "talk or think aloud" sort.

Violation of these rules may yield reports which are the subject's inferences, theories, reconstructions, interpretations or the like rather than being direct reports of ongoing cognitive processes. Ericsson and Simon report that "in no study known to us using general instructions has the investigator complained that subjects have reported too much information from actual memory" (p. 222). Excess verbosity, then, does not appear to be a problem.

After outlining the model, the authors reviewed evidence relevant to it, focusing specifically on some classical issues. Here, we will simply state their conclusions about these issues, leaving it to the reader to pursue the details in their paper which is appended to this report as section 10.0. First, regarding effects of instructions to verbalize on the processes constituting the primary cognitive tasks which are studied (a question of interaction), they concluded that the course and structure of these cognitive processes is not changed. Secondly, regarding the completeness of verbal reports, they acknowledge that there may be some incompleteness; for example under high cognitive task load condit-

ions, or where automation has occurred or where intervening tasks are required concurrent with verbalization - in general, where reports are expected about matters not normally attended to. But incompleteness "... does not invalidate the information that is present " (p. 243). Finally, about the consistency of verbal reports with other behaviour, this is an empirical matter requiring the collecting of other records of behaviour. They found such inconsistency in the evidence only when their model would actually predict it, eg. in retrospective probes.

The direct and simple implication of this verbal report model for us is that experimental observers should be simply asked to "think aloud" in a completely free and uncensored mode while they are performing the tasks at the video terminal and keypad in browsing research, or for that matter, in directed search investigations too, if cognitive events are what we wish to access as experimenters. And, we contend, cognitive events contained in STM are precisely what our browsing research model which we elucidate further in the next section, calls for.

Reflecting back on the library research reviewed previously above, it is apparent from the viewpoint of this verbal report model that all reports about their library useage which were asked of library users were of the retrospective, interpretive or inferential sort. We add to this the fact that none of those studies examined browsing, or library material useage directly, in the moment to moment ongoing sense. Thus all those studies suffer the major criticisms that (1) ongoing browsing behaviour was not studied and (2) ongoing cognitions directing the behaviour were not studied.

6.0 A BROWSING RESEARCH MODEL WITH DEPENDENT VARIABLES

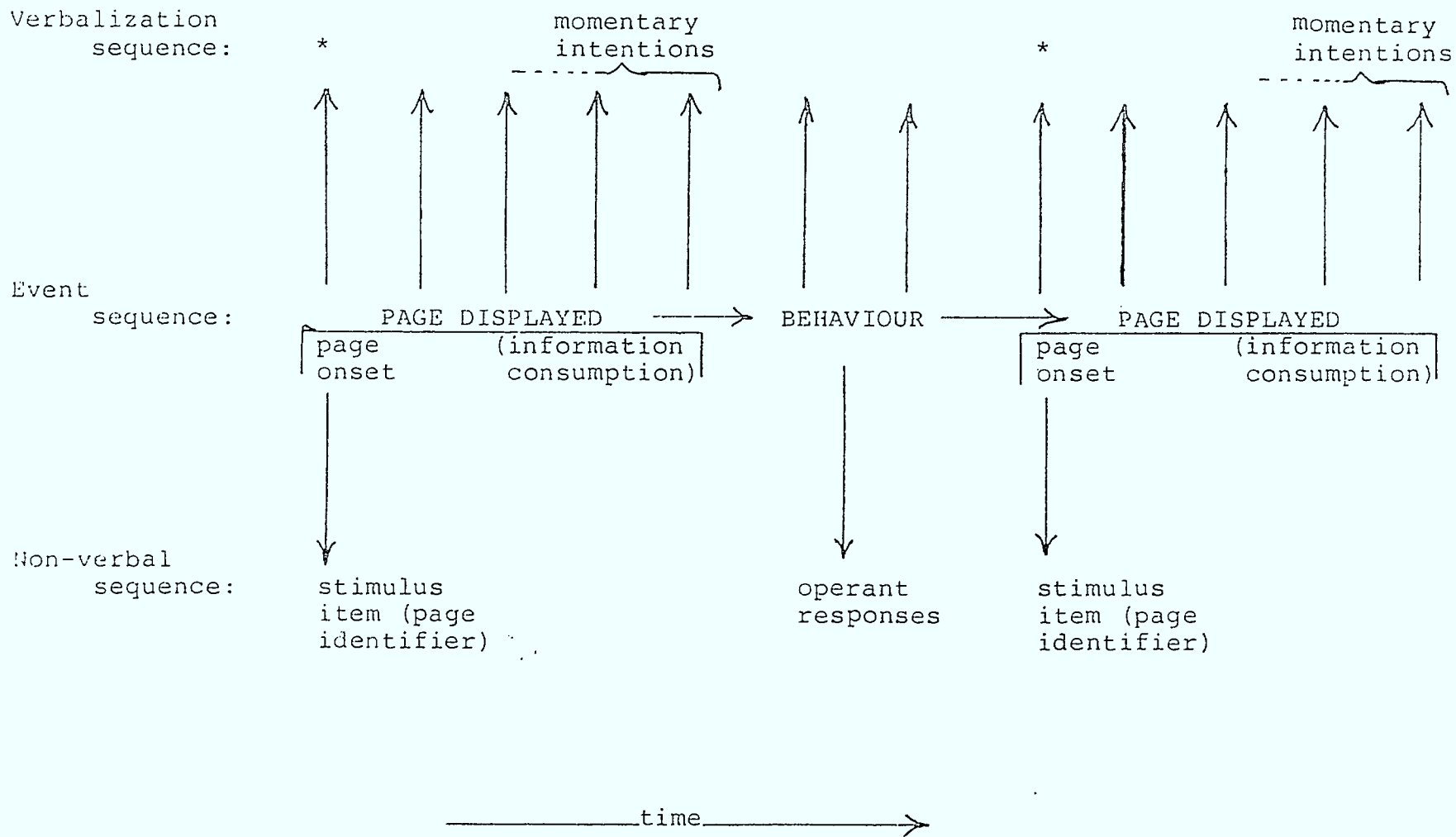
On the following page, we have presented a schematization of the essential features of a descriptive browsing research model as applied to the videotex situation, with indications for experimental recording of dependent variables. The general model is intended to apply to any information and not just to videotex or other verbal information.

Indeed, the model is even more general than that, since it can encompass any kind of search behaviour. Thus, in accordance with the rationale of section 3.0 above, we regard browsing to be a subset of search behaviour, distinguished, if at all, only by the motives of the searcher in the moment to moment sense. These motives may be predominately of the directed search type, or of the browsing type, or they may fluctuate between the two types slowly or rapidly from moment to moment within an experimental session. General control over these motives is exerted by specific instructions to experimental observers. Recognizing this, however, we will continue to use the term "browsing" model for our present purposes.

Moving in time from left to right in the schema, the event sequence simply shows that the observer goes from one page to another by some behaviour, usually a keypad operant response. We show page onset and information consumption as subsets of each page displayed with information consumption in brackets since consumption may not occur or occur only in minimal form. Except for information consumption, this event sequence is the raw overt stimulus-response sequence which simple observation by the experimenter would reveal.

The non-verbal sequence shows two recordable events, operant (keypad) responses, and the actual stimulus item or page accessed.

SCHEMATIZATION OF THE BROWSING SEQUENCE WITH INDICATIONS FOR EXPERIMENTAL RECORDING



Depending on the purposes of a given experiment, the operant response may or may not be automatically recorded. But it would be advisable to so record it so that detailed "errors" and the like would be known as well as "correct" responses. The page identifier should, of course, always be recorded, if not automatically then by the experimenter so that it could be aligned with the subsequent verbal protocol.

The verbalization sequence is simply the "think aloud" continuous ongoing verbal output which should be captured via tape recording. The upward pointing arrows indicate that the content of the verbalizations is related to the displayed pages and to the operant behaviours. These are what the verbalizations are about. Some of this output, one or a few (perhaps many) verbalizations, will be about the momentary intentions which explain the next occurring keypad operant responses. These are labelled accordingly, "momentary intentions" and it is especially important that these verbalizations be captured. Similarly the first verbalization after page onset is especially important since it will confirm or disconfirm whether an accessed page is in accordance with the momentary intentions preceding the keypad operant response. Of course silence on the matter might indicate confirmation since "confirmation" might not be normally attended to in STM. But, in that case, the experimenter's record of page identifier would show that the "correct" page was reached. But where the correspondence is false, that fact should appear in a properly instructed observer's verbalizations eg. "that's not the page I wanted", perhaps sometimes followed by "but I will accept it anyway since I'm browsing".

Intentions are "momentary" only in the sense of occurring at that point or moment in time, just before relevant behaviour. They may indeed be long-standing intentions, persisting over some extended period of time. Or at the other extreme, the browser may undergo a rapid sequence of changes of intentions before acting on one of them. The verbalization sequence should reveal just what is happening regarding intentions. In browsing, of course, these intent-

ions need not necessarily be specific such as wanting to access a particular page. They may be of a form that might be expressed "I'll take whatever page happens to turn up next."

Previously we mentioned the possibility of not recording the operant response. A properly instructed browser will still be verbalizing so that a report of the keys he or she intends to operate or thinks he or she is operating will be available even if the overt behavioural response is not directly recorded. Other verbalizations might also occur during the behaviour phase, such as "oops, I pushed the wrong key" or "I forget which key I need to press", etc. Again, all such reports will be of interest to the browsing research.

Information consumption refers to the act of reading page contents. Verbalization cannot occur during the act of reading since that act fully loads STM. Depending on the purposes of the investigation being pursued at the time, the browser may or may not be asked to read aloud, although strictly speaking, a "think aloud" instruction may be expected to produce reading aloud. Following the reading, silent or aloud, the entire event and recording sequence will recur until the experimenter ends the session.

The schema does not presume an orderly general sequence of events from page to page. Observers may do much floundering around, getting themselves lost, experiencing various frustrations such as not knowing how to get back to some previous page, etc. But the general schema, calling for appropriate verbalizations by browsing participants and other event recording, can capture all possible events of interest. Everything is a matter of momentary intentions and other cognitive events, operant responses and stimulus sequences.

We believe that the descriptive browsing model in the schema fully addresses the conceptual problems in the literature discussed in sections 3.0 and 4.0 above. The model is simply a conceptual description of browsing, whatever its variants. It focuses on the behaviour and cognitions of individual persons at all their moment to moment idiosyncratic flexibilities.

As mentioned previously, it is a general search model, applicable to libraries or china shops. Only the nature of the operant behaviours and of the stimuli would change. Used in directed search investigations as well as for browsing, the schema and recording may yield a great wealth of data beyond the sparse efficiency measures which now predominate in investigations of directed search.

The recorded events, verbal and other, comprise the set of dependent variables in a program of browsing research. Any of a potentially unlimited set of independent variables can be plugged into the general model and schema after appropriate modifications on the dependent variable side if and as necessary to the pursuit of a particular research question. But such dependent variable modifications will probably be minimal once some basic pilot work has been done.

Such pilot work would be directed toward working out and testing instructions to observers, although Ericsson and Simon's report suggests there should be no problems.

The nature of the data is quite different from that in research which yields efficiency measures. Transcription and examination of verbal protocols and aligning them with recordings of stimulus sequences and operant responses may take many hours for each single hour of experimental sessions. Within the constraints of independent variable manipulations, each experimental observer sets his or her own browsing method so that there will necessarily be idiosyncratic variation from observer to observer. But that, after all, is the nature of browsing.

Certainly the stimulus sequence from page to page will differ from observer to observer. Thus data "summarization" may be either difficult or impossible in the traditional sense. However, behaviour is orderly so that generalities about browsing are expected to emerge at some level although the process of arriving at these generalizations may be very slow.

7.0 INDEPENDENT VARIABLES FOR STUDYING BROWSING IN VIDEOTEX DATA BASES

The previous section has provided the dependent variables for studying browsing. The set of independent variables which may be plugged into the browsing research model is unlimited. This set of independent variables might include:

broadly versus narrowly defined browsing objectives (the whole data base versus French restaurants in the east end of Ottawa)

the effects of different time pressures on browsing

browsing strategy differences between highly experienced versus naive videotex users

idle curiosity browsing versus more highly structured goals

different types of access systems (menu or tree, versus keyword or Boolean)

the effects of different types of visual information presentations

alternative menu arrangements for the same data bases

A research program on browsing should provide much general information and would break entirely new conceptual ground in the light of our critique of library research. As far as the writer knows, the general topic has not been properly researched in any context. One can make a case for the proposition that browsing is a fundamental, important and all-pervasive aspect of human behaviour and cognition. Looking beyond the verbal information context, browsing as we conceptualize it, is highly recurrent, frequent (a part of daily life), inescapable and is engaged in by everyone from infancy to senescence. Yet it has been mostly ignored for study, possibly because of a lack of appropriate tools with which to research it. The present theoretical and methodological suggestions may provide the necessary initial tools.

On a practical level, a research program as suggested here may lead to a codification of rules for information providers that would make for much more effective videotex presentations. Frustrations engendered in browsing by poor presentations may reveal the nature of deficiencies more clearly than investigations into directed search only. The access system question may be one of the fundamental ones to be addressed early (trees, keywords, etc.). A combination of browsing research with investigations of directed search, perhaps sharing the same methodology, has the potential of suggesting hybrid or entirely new access systems which might more effectively subserve the various uses to which videotex will be put.

8.0 ADDITIONAL DEPENDENT VARIABLES FOR STUDYING BROWSING

After the end of reseach sessions, observers could be asked to do two things.

1. draw or construct a map of the information structure of the videotex data base they had been working with during browsing sessions
2. draw or construct a map of an alternative information structure which they think might be better than the one they had been working with.

These activities might produce something of heuristic value and might even prove sensitive to independent variables. Naturally, these are quite different kinds of data than the verbalizations during browsing research sessions. They specifically ask for the observers constructions and theorizing.

Since these dependent variables could simply be added on to the end of each browsing session, they would constitute low cost, low effort additional information.

9.0 BIBLIOGRAPHY

Ericsson, K.A. & Simon, H.A. Verbal reports as data Psychological Review, 1980, 87, 215-251

Hyman, R.J. Shelf classification research: past, present - future? University of Illinois, Graduate School of Library Science, Occasional paper number 146, November, 1980

Hyman, R.J. Access to library collections, Scarecrow Press, Metuchen, N.J., 1972

10.0 APPENDIX - VERBAL REPORTS AS DATA
ERICSSON & SIMON