2

Bolt Beranek and Newman Inc.

1

BBN

~~77-1105~~

Report No. 3459

The C^3–System User
Vol I: A Review of Research on Human Performance as it Relates to the Design and Operation of Command, Control and Communication Systems

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

February 1977

Submitted to:
Defense Advanced Research Projects Agency

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
3459

2. GOVT ACCESSION NO.
AD-A116633

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
THE C^3 SYSTEM USER
Vol I: A Review of Research on Human Performance as it Relates to the Design and Operation of Command, Control and Communication Systems.
~~Vol II. Workshop Notes~~

5. TYPE OF REPORT & PERIOD COVERED
Final Report
Feb. 1976-Feb. 1977

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
R. S. Nickerson, M. J. Adams, R. W. Pew,
J. A. Swets, S. A. Fidell, C. E. Feehrer,
D. B. Yntema, and D. M. Green

8. CONTRACT OR GRANT NUMBER(s)
MDA903-76-C-0207

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Mass. 02138

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

12. REPORT DATE
February 1977

13. NUMBER OF PAGES
371

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Unlimited Distribution

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Command, Control and Communication; Human Factors; Psychology; Displays; Information Processing; Human Performance; Memory; Attention; Decision Making; Problem Solving; Stress; Interactive Systems; Person-Computer Interaction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This is the final report for Contract No. MDA903-76-C-0207, Human Factors in Command, Control and Communication Systems. Volume I contains the results of several literature surveys and is organized around the following topics: Information Presentation and Processing; Monitoring and Attention Sharing; Memory and Information Storage; Decision Making and Problem Solving; Effects of Stress on Performance; and Person-Computer Interaction. The final chapter lists several recommendations for future research. Volume II contains supplementary material relating to the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

20 (continued)
several workshops that were held during the course of the project.

Report No. 3459 Bolt Beranek and Newman Inc.

THE C^3-SYSTEM USER

VOL. I: A Review of Research on Human Performance as it Relates to the Design and Operation of Command, Control and Communication Systems

February 1977

R. S. Nickerson
M. J. Adams
R. W. Pew
J. A. Swets
S. A. Fidell
C. E. Feehrer
D. B. Yntema
D. M. Green

This research was supported by the Defense Advanced Research Projects Agency under ARPA Order No.: 3167
Contract No.: MDA903-76-C-0207

Submitted to:

Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209
Attention: Dr. Craig Fields

THE C^3-SYSTEM USER

Chapter Authors:

Chapter 1, Information Presentation and Processing:
M. J. Adams, R. S. Nickerson & D. M. Green

Chapter 2, Monitoring and Attention Sharing:
J. A. Swets

Chapter 3, Memory and Information Storage:
M. J. Adams & R. S. Nickerson

Chapter 4, Decision Making and Problem Solving:
C. E. Feehrer & D. B. Yntema

Chapter 5, Effects of Stress on Performance:
S. A. Fidell

Chapter 6, Person-Computer Interaction:
R. S. Nickerson & R. W. Pew

Chapter 7, Problems for Future Research:
R. S. Nickerson

PREFACE

This is the final report for Contract No. MDA903-76-C-0207, Human Factors in Command, Control and Communication Systems. An objective of this project was to summarize experimental findings in the areas of human information processing and man-computer interaction as they pertain to the design and operation of command, control and communication systems, and to identify further researchable problems in these areas. A further objective was to inform a variety of researchers about DARPA's program in the human factors of C^3 systems, and to involve these researchers in working seminars on the topics that represented intersections between their own special areas of interest and C^3 problems. To these ends, we have reviewed the literature that pertains to the several problem areas that we believe to be most germane to human factors in C^3 systems, and we have convened a series of workshops at which experts in these various areas exchanged their views concerning current work in their respective areas of expertise and how it relates to the design and operation of C^3 systems. This report documents the results of these activities.

The report is composed of two volumes. Volume I contains the results of the literature surveys and is organized around the following specific topics within the general domain of human information processing: Information Presentation and Processing; Monitoring and Attention Sharing; Memory and Information Storage; Decision Making and Problem Solving; Effects of Stress on Performance; and Person-Computer Interaction. The final chapter lists several recommendations for future research.

Volume II contains supplementary material relating to the several workshops that were held during the course of the project.

Prior to the convening of each workshop, every prospective attendee received a set of background notes on C^3 systems, and a list of questions that were representative of those that might be discussed at the workshops. These questions were intended only to provide people with points of departure for relating their own work to C^3 problems. Following each workshop, a set of notes was prepared and distributed to the attendees. These notes did not constitute workshop minutes or even an integrated narrative account of what transpired. We simply tried to capture the main ideas that were discussed, more or less in the order in which they were brought up, and to record them with minimal commentary.

The first section of Volume II contains background notes that were sent to the prospective attendees before the convening of each workshop. Each of the subsequent sections contains the pre-workshop questions and the post-workshop notes associated with one of the workshops.

Two additional papers, the preparation of which was supported in part by this contract, have been submitted under separate cover. The first of these papers will appear in the Proceedings of a NATO Advanced Study Institute, Mati, Greece, 15-17 September 1976, and the second in the Proceedings of an ACM/SIGGRAPH Workshop on User-Oriented Design of Interactive Graphics Systems, Pittsburgh, Pennsylvania, 14-15 October 1976.

Table of Contents

CHAPTER 1

INFORMATION PRESENTATION AND PROCESSING

A central problem for many C^3 personnel is the need to absorb large amounts of information. The efficiency and effectiveness of C^3 operations must depend to some degree on how successful the system is in presenting information in a form that makes assimilation and storage easy and convenient.

We begin with the general observation that the information that is used in C^3 operations has many different forms. Some is nominal--the names of places, people, materiel. Some is structural or pictorial--maps, deployments, distinguishing shapes of weaponry. Some is quantitative--troop strengths, resource data. Some is statistical or probabilistic--engagement outcome predictions, weather forecasts. And, of course, the information with which any given individual has to cope is likely to be of more than one type.

In general, the nature of the information that is to be presented should dictate, within limits, the manner of presentation. However, beyond some fairly obvious principles--for example, the fact that describing geographical relationships with words is usually inefficient in comparison with using a map for the same purpose--relatively little is known for certain regarding how to optimize display design. What advantages does one modality for information presentation have over another? When should information be presented redundantly in more than one modality? What factors should determine how information is encoded within a given modality? What principles should govern the organization of information in a display? To what extent do display requirements depend on whether the information, or some part of it, has to be retained in memory over a period of time?

The purpose of this chapter is to review experimental work that relates to these kinds of issues. We did not expect, when beginning this review, to be able to answer these questions definitely; the intent was to extract from the literature basic research findings that are relevant to these issues and to present them in an organized way.

Hearing as an Information Channel

The efficient exchange of information among personnel in modern command and control systems depends upon a considerable amount of auditory information processing. In this section we treat the auditory system as an information channel and review the general properties of this system, emphasizing both its particular strengths and weaknesses. We limit our attention here to a consideration of nonspeech material. But inasmuch as speech is the primary means of auditory communication among people, a separate portion of the chapter will be devoted solely to it.

Spectral Resolution

The ear is capable of making a reasonably accurate spectral analysis of the acoustic waveform in a remarkably short time. The addition or deletion of a small amount of energy in a broadband spectrum is quickly detected. The ability to note the presence or absence of pure tones is also good, and one can listen selectively to the first half-dozen components of a complex, periodic wave (Plomp, 1964). This analytic ability of the auditory system stands in marked contrast to visual color perception. In vision, markedly different spectra are seen as identical; indeed, all of color mixture is based on this fact. In audition, there are no such confusions. Spectra are discriminably different unless the amounts of energy at all wavelengths are nearly identical.

Dynamic Range

Another outstanding characteristic of the auditory sense is its remarkable dynamic range. The ratio of the intensity that corresponds to the threshold of hearing to that which corresponds to the threshold of pain is about 12 orders of magnitude. Even more remarkable, especially when compared to vision, is that little adaptation occurs. The enormous dynamic range is almost instantaneously available in audition. An intense sound of 70 to 90 dB influences the threshold of a very weak sound for only a few milliseconds after the termination of the louder sound. This is to be contrasted with vision, which has a larger dynamic range in terms of intensity but may require 30 minutes of dark adaptation to achieve full sensitivity at the lower intensity levels.

Basic Discrimination Capacity

The ear can detect changes of about 10% in the intensity of pure tones, periodic waves, and even noise waveforms of moderate bandwidth. Thus, the movement of a steady source by a distance of about 5% away from, or toward, the observer can readily be detected in a free field on the basis of the change in intensity alone. Changes in the frequency of a pure tone can be heard if the change is more than 0.3% of the base frequency, at least for most frequencies. Since the distance between successive notes on our 12-tone scale is about 6%, the ear's frequency discrimination corresponds to about 1/20 of that interval. Again, the speed of these discriminations is impressive. Samples as short as 1/10 sec. are sufficient to support performance at these discrimination limits.

Temporal discrimination is also extremely good if the subject is given practice at the task (Green, 1971; Ronken, 1970). Practiced subjects can learn to discriminate ordered pairs of stimuli, for example, whose total duration is only 2 msec. and whose power spectra are identical, so the phase spectra is the only possible basis of discrimination (Patterson & Green, 1970). Surprisingly, in light of the temporal acuity of the auditory system demonstrated by such results, unpracticed subjects find it difficult, or, impossible, to report the order of the elements of a recurring series of arbitrary acoustic stimuli whose onsets are separated by less than about .5 sec. (Bregman & Campbell, 1971; Warren, Obusek, Farmer, & Warren, 1969). The limitation in this case appears to be central in origin, however, and people can learn to discriminate among differently ordered sequences with very rapidly presented elements if they are given sufficient practice at the task (Nickerson & Freeman, 1974).

The limits cited above refer to discrimination in controlled laboratory situations in which subjects are able to concentrate fully on the discrimination task. In operational settings, factors such as distraction, boredom or fatigue may make discrimination more difficult. But increasing the size of the stimulus difference by a factor of 3 would probably overcome any degradation caused by these other factors. Later we turn to the problem of recognizing or identifying one stimulus among a set of many. Except for special classes of signals, the ear is remarkably poor at this task. Before taking up the topic of identification or recognition, however, we will turn to some phenomena that reflect our ability to localize sound sources in space.

Binaural Phenomena

Sounds in normal life are well localized in space, and sounds presented by earphones appear to occupy distinctive and well-defined positions within the head. If a click arrives at one ear 20 msec. before it arrives at the other, one can reliably detect which ear was first stimulated. Such acute interaural temporal differences correspond to the motion of a distant source of about 1°; a source located straight ahead, for example, can be heard to move about 2 in. at 10 ft. Movements would have to be somewhat larger to be detectable at the sides of the head. Although some use of this ability has been made, it is largely neglected in most display systems.

In addition to this positional sense, the two ears can be utilized to enhance the detectability of a signal if the signal and its masker have different interaural phase relations. For example, in the presence of a loud masking noise that is in the same phase at the two ears, such as might be encountered in a tank or airplane, reversing the signals at the earphones (to make the speech antiphasic) can improve speech reception by 3 to 6 dB (Licklider, 1948). Even larger gains in detectability can be realized for sinusoidal signals, if the noise and signal possess different interaural phase relations (Hirsh, 1948, Green & Henning, 1969

A closely related ability is that of selectively attending to one sound source in a clutter of many, the so-called "cocktail party effect." This ability to isolate one source and ignore others involves a host of variables beyond the sensory ones, but it should be recognized, and it has implications for practical information-acquisition situations. Capitalizing on this binaural ability could prove to be useful, for example, in conference calls. The auditory signal could be coded so that each speaker would occupy a different point in perceptual space. At the very least, such coding should facilitate a listener's recognition of who was talking. In addition, it could minimize the interference produced when two or more speakers start talking simultaneously--a circumstance that often happens and can be highly disruptive in multiperson telephone conversations.

There are many possibilities that one might explore to implement such an idea. The simplest would be to employ two lines and code a different delay between the two lines for each talker. Each listener using earphones would hear each speaker at a different point in perceptual space. More complex signal-processing techniques might be more effective, but this one would be inexpensive. Loudspeakers rather than earphones could be used if even more elaborate signal processing were utilized.

Finally, we cannot leave the topic of binaural phenomena without mentioning dichotic listening. Dichotic listening involves monitoring two distinct and often different sources presented to the two ears simultaneously. Variants of this task have been studied extensively. In fact, the task has been the standard paradign in enumerable studies of selective attention. Audition is favored over vision for this purpose because the two ears may be treated as independent information channels, and there are no obvious peripheral factors such as direction of fixation to contaminate the results. Interpretations of our ability to perform dichotic listening tasks tend to emphasize central factors. Usually, a stream of material is presented to one ear and the subject is directed to attend to that ear while secondary material is presented to the "unattended" ear.

A basic finding from this research is that material presented to the unattended ear is not completely ignored. Change in the voice of the talker, for example, or the occurrence of one's name, is immediately noted. An obvious application of this fact to C^3 operations would be the utilization of auditory signals for the purpose of catching one's attention for important and urgent messages. For an extensive review of research on dichotic listening, see Kahneman (1973).

Auditory Recognition or Identification

A large and impressive literature on the use of artificial sounds attests to the fact that absolute identification accuracy decreases sharply as the number of sounds increases. If stimulus variation is confined to a single dimension, such as intensity, then accuracy begins to fall once the number of stimuli exceeds about 4 or 5, and asymptotic performance in terms of the information measure (bits/judgment) occurs between 2 and 3 bits (The Magical Number 7 ± 2 of Miller, 1956). Since this limitation holds even when all of the sounds are pair-wise discriminable, the decline in identification accuracy is usually thought to be a failure of "memory," "response encoding," or some other higher order process. It clearly cannot be attributed to a simple sensory failure. One way to lessen the decline in accuracy caused by increasing the number of sounds is to vary the sounds along several stimulus dimensions such as pitch or duration of on-off ratio, rather than simply to divide a single stimulus dimension more finely (Pollack & Ficks, 1954).

Whatever the nature of this puzzling limitation, it is not an absolute one. We can recognize many spoken words, and we do learn to associate specific sounds with a large number of everyday sound sources, such as cars, dishwashers, birds, dogs, etc. The identification of these natural sounds should, a priori, be more difficult to learn than the sounds used in laboratory tests since natural sounds tend to change a great deal from presentation to presentation. Despite this variation, however, many common sounds are identified with an accuracy that would be the envy of any laboratory test. We know from the work of several investigators (see Chapter 3 of this report) that subjects can discriminate among "old" (previously seen) pictures and "new" ones, at very high accuracy levels even when the set of "old" pictures numbers in the hundreds or thousands. Miller and Tanis (1971) found somewhat similar results for common sounds in an auditory recognition task. There were 350 old sounds and the test contained 100 pairs of old and new sounds. The accuracy was somewhat lower (69%) than that obtained in visual recognition studies but this could have been due in part to the poor fidelity of the recording, which was no better than telephone lines.

It would be desirable to understand better the information bottleneck associated with the absolute identification task. It is a major source of difficulty in training sonar operators. Various proposals have been made to improve the efficiency of this task, but they have been based mainly on hunches or intuitions and not on any fundamental understanding of the limitations of the identification process. Testing of all of these different suggestions is both time-consuming and expensive.

A similar observation may be made vis-a-vis the evaluation of various ideas about how auditory information should be presented. At present, the design of auditory-display formats is largely an intuitive process. No general guidelines exist other than the obvious one, namely to maintain the detectability of the signal. Again, the testing of the different design formats is both time-consuming and expensive. Nevertheless, there are many situations in which one might wish to use auditory codes. They could be particularly valuable when the visual channel is in danger of being overburdened.

Sound signals have advantages over visual signals for certain purposes. They do not require constant attention; indeed, they both alert and communicate simultaneously. Despite this fact, our present ignorance about the absolute identification bottleneck makes us reluctant to employ artificial sound signals. We know that accuracy will fail when more than a few alternative sounds are employed. Our ignorance of the nature or cause of the decline in accuracy is almost total. There has been little work on this problem, either theoretically or empirically in the past 20 years. This lack of effort is especially unfortunate in view of the potential practical gains.

Speech

Speech clearly is a preferred means of acquiring and transmitting information for most purposes. In part, this is probably due to the fact that most people can transmit information easier and faster by speaking than by any other means. Table 1, which is taken from Newell et al. (1971), compares the rates at which a person can emit words for each of several output modes.

Some Studies of the Utility of Speech

In spite of the assumed advantages of speech--perhaps because of it--very few formal studies have been done to evaluate its effectiveness relative to other means of communication. An exception is a series of investigations by Chapanis and his colleagues. These studies will be reviewed here in some detail because they were motivated by the question of how to design computer systems that can be used effectively by people in an interactive fashion--a question that has considerable significance for the design and operation of C^3 systems. Chapanis (1971) has argued that before truly symbiotic person-computer systems can be developed, much more must be learned about how human beings communicate with each other, especially when collaborating on problem-solving tasks. In keeping with this view, he and his

Table 1. Speeds of Man's Output Channels

1.	READING OUT LOUD	~4 WORDS/SEC.
2.	SPEAKING (SPONTANEOUSLY)	~ 2.5 WORDS/SEC.
3.	TYPING (RECORD)	~ 2.5 WORDS/SEC.
4.	TYPING (SKILLED)	~ 1 WORD/SEC. (~5 STROKES/SEC.)
5.	HANDWRITING	~.4 WORD/SEC.
6.	HAND PRINTING	~.4 WORD/SEC.
7.	TELEPHONE DIALING (TOUCH TONE)	~.3 WORD/SEC. (~1.5 DIGITS/SEC.)
8.	MARK SENSE CARDS	~.1 WORD/SEC (~.5 DIGIT/SEC.)

From Newell et al., 1971.

colleagues have conducted a series of experiments to address the question of how people naturally communicate with each other in certain problem-solving situations, and, more particularly, how their communication is affected by the devices and techniques they are permitted to use for that purpose (Chapanis, Ochsman, Parrish, & Weeks, 1972; Chapanis & Overbey, 1974; Chapanis, Parrish, Ochsman, & Weeks, in press; Ochsman & Chapanis, 1974). For summaries of this work, see Chapanis (1973, 1975).

The general paradigm has been to vary the means by which the members of a two-person team can communicate with each other while collaborating on a problem-solving task. Each problem is structured in such a way that neither of the team members can solve it by himself, because each has only a portion of the information that is necessary to reach a solution.

Problems are selected to meet certain criteria, among which are the following:

a. Representativeness of tasks for which interactive computer systems might be used either now or in the future.

b. Face validity (the tasks were to be of the type encountered in real-life problem situations).

c. Definite recognizable solutions, typically attainable within an hour.

d. Not demanding of special skills or knowledge.

Among the problems that were selected as satisfying these criteria were the following:

a. Geographic orientation problem. Subjects were to determine the location of the physician's office that was closest to a given home address. One of the team members was provided with a gridded street map and a street index, and the other with a page of alphabetized physicians' listings, taken from the classified section of a telephone directory.

b. Equipment assembly problem. A trash-can carrier was to be assembled. One team member was given the parts and the other the assembly instructions.

c. Information-retrieval problem. The task was to find in The New York Times Index the citation of every newspaper article relevant to a specified topic that appeared in The New York Times during a year.

d. Class-scheduling problem. Four college class courses were to be scheduled for one semester. One member was provided with the names of the required courses and a set of scheduling constraints. The other had a booklet with a complete schedule of classes for the university.

e. Fault-finding problem. A mock-up of an automobile ignition system was to be tested with a commercially available auto analyzer. One of the team members was given the mock-up and the auto analyzer, the other had a set of instructions on how to use the latter.

f. Part-identification problem. The task was to identify, from a stock of parts, a miniature light socket that was identical to one provided by the experimenter. One of the team members was given the socket that was to be matched, and the other a stock of 63 sockets, only one of which satisfied the requirement.

The main independent variable of interest in these studies was the method by which the collaborating problem solvers were permitted to communicate with each other. Several communication modes were investigated, ranging from the highly constrained mode, in which collaborators could communicate only by means of typewritten messages, to a relatively unconstrained, or "communication-rich" mode, in which they communicated by voice, and perhaps gesture, in a face-to-face fashion. Intermediate modes included voice combined with television, voice combined with handwriting, voice combined with typewriting, voice only, handwriting combined with television, typewriting combined with television, handwriting combined with typewriting, and handwriting only.

Several performance measures were used, including total time to problem solution, time spent in various types of activities (sending messages, receiving messages, searching for information, making notes, waiting, etc.), and a set of linguistic measures (number of messages generated, number of words per message, percentage of sentences that were questions, total number of words used by a subject, total number of different words used by a subject, type-token word ratio, word-production rate, etc.).

Perhaps the major finding that has come out of these studies is that of the overriding importance of voice for effective interperson communication in such problem situations. By and large, modes of communication that included a voice channel were associated with markedly better performance than were those modes which did not include a voice channel. On the average, the time required to reach a solution tended to be about half as long when voice communication was possible than when it was not. Moreover, given the presence of a voice channel, the presence or absence of other means of communication (facial expressions, gestures, handwriting, typewriting) had relatively little additional effect. Ochsman and Chapanis (1974) summarize their results in the following way. "The practical implications of these data can be simply stated. The single most important decision in the design of a telecommunications link should center around the inclusion of a voice channel in the solution of factual real-world problems. Little else seems to make a demonstrable difference" (p. 618).

The finding that the addition of video to voice had little effect on performance is a particularly noteworthy one in view of the interest that has been shown in the possible use of video terminals for teleconferencing purposes. A question that deserves further research is whether the finding of Chapanis and his colleagues that the addition of video to voice adds a negligible increment to performance efficiency is true only of certain types of tasks. Perhaps the challenge is to identify those tasks for which video would make a nontrivial difference.

Among the other findings that Chapanis and his colleagues have reported are the following:

a. While more time was spent in sending and receiving messages with nonvoice communication modes than with voice modes, not all of the difference between voice and nonvoice modes is accounted for by this fact. Noncommunicative activity also consumed more time with the nonvoice than with the voice modes (Chapanis, Ochsman, Parrish, & Weeks, 1972; Ochsman & Chapanis, 1974).

b. Subjects who communicated by voice tended to spend more time simultaneously doing other things while communicating than did subjects who did not use a voice channel (Chapanis, Ochsman, Parrish, & Weeks, 1972). In the words of Ochsman and Chapanis (1974), "Subjects in the hard-copy modes tend to concentrate more on the communication task, i.e., on sending or receiving. Subjects who communicate by speaking find it easier to conduct other activities simultaneously regardless of whether they are speaking or listening to spoken messages" (p. 614).

c. Although subjects solved problems much faster when using voice mode than when using a nonvoice (typing or handwriting) mode, the number of messages exchanged was several times greater in the former case than in the latter (Chapanis, Parrish, Ochsman, & Weeks, in press).

d. It follows from the preceding point that the rate of word production varied greatly as a function of communication mode. In one case, the rate of word emission with voice communication (3 words per sec.) differed from that of inexperienced typists communicating by using the typewriting mode (at 6 sec. per word) by a factor of 18 (Chapanis, 1973). In general, voice communication proved to be much faster, and more wordy, than other modes.

e. The type-token ratio (ratio of different words to total words) was higher for the nonvoice than for the voice communication modes, suggesting that subjects were more selective or less redundant in their use of words in the former case (Chapanis, Parrish, Ochsman, & Weeks, in press).

f. Average message length or average number of words per message did not differ significantly as a function of communication mode (Chapanis, Parrish, Ochsman, & Weeks, in press).

g. When communication was limited to typewritten messages, total time to problem solution was affected very little by whether the subjects were experienced typists (Ochsman & Chapanis, 1974). Chapanis (1975) attributed this result to the facts that subjects spent, on the average, less than a third of their time communicating and that the process of communicating via typewriter is very different from that of typing copy; consequently, the skilled typist may have less of an advantage over the unskilled in the former case.

h. Solution time for subjects who communicated via handwriting was approximately the same as for those who communicated via typewriting (Ochsman & Chapanis, 1974).

i. Providing the subjects with the ability to interrupt each other had little or no effect on the time required to solve problems, or on the rate at which words were produced (Chapanis & Overbey, 1974).

j. Ability to interrupt did, however, have certain noticeable effects: Shorter messages were produced, more of them were produced, and they were exchanged at a greater rate (Chapanis & Overbey, 1974).

k. Error rates tended to be remarkably high in general: 1 per 4 words, and 1 per 3.1 words, for experienced and inexperienced typists, respectively; 1 per 14 words in handwriting mode (Chapanis, 1973).

l. The typed and handwritten protocols, as well as the transcriptions of voice messsages, were characterized by a general unruliness. Misspellings, conventional and idiosyncratic abbreviations, violations of rules of punctuation and syntax, as well as errors of fact were so common as to make errorless sentences rare (Chapanis, Parrish, Ochsman, & Weeks, in press).

Chapanis and his colleagues drew a number of general conclusions from their results. First and foremost, they concluded that a voice channel is of overriding importance for any system that is to be truly interactive. They concluded also that for a computer system to be truly conversational, it must be forgiving of the types of errors and irregularities that tend to characterize people's efforts to communicate with each other. The provision for the receiver of a message to have the ability to interrupt the sender was seen as probably not worth its cost.

Perhaps as important as the conclusions that were drawn are the questions that the results prompted the investigators to pose for future research. Foremost among these questions are several that relate to the determinants of communication patterns. How, for example, do such patterns depend upon such factors as the nationalities of the communicators, the purpose of the communication, the number of communicators? More generally, Chapanis asks, "What are the rules that govern normal human communication?" A complete answer to the last question may be a long time in coming; Chapanis and his colleagues are undoubtedly right, however, in their assumption that computer systems that are able to engage in truly human-like conversational interactions with their users are unlikely to be developed until considerably more attention is given to it.

Speeded Speech as a Means of Increasing the Rate of Information Assimilation

A problem facing modern commanders, and other C^3-system personnel, is the need to assimilate large amounts of information over relatively short periods of time. It is important, therefore, that information be presented in the most effective ways. How this requirement should be translated into report preparation guidelines, briefing procedures and display designs is not entirely clear, because in spite of considerable research on the topic, relatively little is yet known regarding the variables that determine the rate at which information can be assimilated by human beings.

Among the approaches that have been explored for the purpose of increasing the rate of intake of verbal information is that of speeded, or time-compressed, speech. The purpose of this section is to review this work briefly and to consider whether the possible application of speeded speech in C^3 systems warrants further investigation.

People vary considerably in the rate at which they emit words when speaking "normally." Rates varying from about 100 to about 250 words per minute have been obtained from samples of radio announcements, lectures, and spontaneous speech (Abrams, Goffard, Kryter, Miller, Sanford, & Sanford, 1944; Gregory, 1969; Maclay & Osgood, 1959; Voelker, 1938); however, the mean rates that have been reported tend to be within the 125 to 175 words-per-minute range.

From a theoretical point of view, it is interesting to consider what determines the rate at which we typically speak. In the absence of any data, there are several plausible answers. The limiting factor might be the speed with which one can manipulate the articulatory apparatus. Or, it might be the more cognitive demands of speech production, namely, the process by which utterances are conceived before being realized as articulatory gestures. It might be the limited temporal resolving power that the auditory system can apply to the decoding of the speech signal. Or, it might be more cognitive factors on the input side, namely, limitations in the speed with which the brain can extract meaning from the coded signals it receives from the peripheral auditory system.

In fact, the speed with which the articulatory apparatus can be manipulated can be ruled out as the limiting factor, because people are demonstrably able to talk faster than they typically do, at least for short periods of time. On the basis of data presented by French, Carter, and Koenig (1930), and by Hudgins and Stetson (1937), Miller (1950) estimated that the speech musculature is capable of sustaining a syllable production rate of about 5 per second (higher rates are obtainable during a single exhalation). Assuming an average of about one-and-one-quarter syllables per word, this translates to about 240 words per minute. And at least one investigator has reported obtaining intelligible speech with a word-emission rate of 285 words per minute from a speaker attempting to talk as rapidly as possible (Goldstein, 1940).

The temporal resolving power of the ear is also an unlikely place to look for an explanation of why people typically do not talk faster. While people are surprisingly poor at reporting the order of occurrence of auditory stimuli that are presented at moderate rates of several hundred milliseconds per stimulus (Bregman & Campbell, 1971; Warren, Obusek, Farmer, & Warren, 1969), the evidence suggests that this is not the limiting factor in determining speech rates. Indeed, performance is greatly improved when the stimuli are words or speech-like sounds, such as vowel segments, and are joined by the sorts of transitional sounds found in normal speech (Thomas, Hill, Carroll, & Garcia, 1970; Warren, Obusek, Farmer, & Warren, 1969). Furthermore, the difficulty in reporting order appears to be, in large measure, a response-encoding problem, inasmuch as people can learn to discriminate one order from another even when the individual stimuli are a few milliseconds in duration (Nickerson & Freeman, 1974). More generally, the temporal resolving power of the ear has been shown to be adequate to distinguish differences as small as 2 msec. even when spectral and phase cues to order are eliminated by experimental control (Green, 1971; Patterson & Green, 1970).

In short, the rate at which speech is typically produced appears not to be determined either by sensory or motor limitations. Is it the case then that we speak as slowly as we do because of limitations in the rate at which higher level processes can conceive utterances, or extract the meaning from incoming speech? Nichols (1955) has estimated that silent thought proceeds at about 400 words per minute, and Foulke, Amster, Nolan, and Bixler (1962) report 250 words per minute as an average silent reading rate for moderately difficult material. Although some investigators have found that the intelligibility of speech begins to fall off as speech rate is increased beyond about 150 words per minute (Abrams et al., 1944), the results of several studies indicate that speech may remain highly understandable at rates well over 200, and sometimes over 300, words per minute (Fairbanks, Guttman, & Miron, 1957; Fairbanks & Kodman, 1957; Foulke, 1971; Foulke et al., 1962; Friedman, Orr, Freedle, & Norris, 1966; Goldstein, 1940; Harwood, 1955; Langford, 1968; Nelson, 1948). Thus, it would appear that cognitive limitations also cannot account for the relatively slow rates at which speech usually is produced. It may be the case, however, that utilization of maximum rates on a sustained basis would require an expenditure of energy that the individual is unwilling or unable to make. Thus, typical speech rates may represent a compromise between the desire to transmit and receive information and a reluctance to make the effort necessary to function at peak capacity for very long periods of time.

What seems to be clear is that people are capable of producing and understanding speech at much higher rates than those usually encountered in normal conversations. What the limits are remains to be determined. Moreover, in the case of speech reception at least, there may be several different limits, depending on the nature of the listener's task. For example, the maximum rate to which one can listen effectively for the purpose of understanding every detail of what is being said may be quite different from the maximum rate to which one can listen for the purpose of detecting something specific (e.g., specific word(s), a speaker's voice, a particular topic).

Research interest in rapid speech has been focused primarily on questions relating to listening and understanding; very little attention has been paid to the question of how to facilitate talking at faster rates. This focusing of attention on reception, as opposed to production, has been due in large part to the fact that the speed of recorded speech can be instrumentally increased, so that it may be presented to the listener at rates in excess of those at which people do or can produce it vocally.

Investigations of the intelligibility or comprehensibility of instrumentally speeded speech have been motivated by both theoretical and practical interests. For the theorist, speeded speech provides a means for studying the human being's limitations as an information processor. To the educator, or training specialist, and to others who are concerned with the problem of conveying information to people aurally, speeded speech appears to represent a potential way to save large amounts of lecturing time, and to increase the efficiency of information transfer. The potential utility of this approach for presenting recorded educational material for blind persons is obvious, especially when compared with the 110-words-per-minute reading rate typically achieved by an experienced adult user of Braille (Gregory, 1969). The military also has shown an interest in instrumentally speeded speech as a means of increasing the efficiency of lectures, briefings, and other oral presentations (Briscoe, 1973; Dailey, 1974).

The easiest way to increase speech rate is to record it at one speed and play it back at another. Fletcher (1929) was perhaps the first to experiment with the playback speed of recorded speech in this way; others who have also done so include Steinberg (1936), Goldstein (1940), and Klump and Webster (1961). This method has the advantage of simplicity, but it changes the fundamental frequency and the frequencies of the formants, and thereby drastically alters the perceptual characteristics of the speech. The overall effect is to give the voice a "Donald Duck" quality that listeners find somewhat annoying even when understandable. Moreover, comprehensibility drops off fairly sharply even for speed-up factors of considerably less than two.

A qualitatively different method of time compression, which does not have the same disadvantages, grew out of work by Miller (1946) and by Miller and Licklider (1950), who found that alternating brief segments of speech could be blanked out of a recording without decreasing intelligibility greatly. Garvey (1953) and Garvey and Henneman (1950) eliminated the gaps left by the blanked-out segments, by manually discarding the segments and splicing the tapes so as to abut the speech on either side of each segment that was discarded. Fairbanks, Everitt, and Jaeger (1953) developed a technique for accomplishing the discarding and abutting electronically; subsequently, several other procedures that accomplish the same thing have been developed. The distinguishing characteristic of this method of speech compression is the elimination, from the speech signal, of segments of fixed duration at regularly spaced intervals, without regard for the contents of those segments. We will refer to it, therefore, as the fixed-interval sampling method. The method often is referred to simply as the "sampling method," but the fact that the sampling occurs at regularly spaced intervals should be borne in mind. It seems to be generally agreed that the duration of the discard interval should be no longer than about 30 msec.; otherwise, the likelihood of discarding critical feature information becomes unacceptably high.

Intelligibility, or comprehensibility, typically has been found to fall off gradually as the compression ratio (percentage of original signal that is discarded) is increased to some point, and then rapidly as it is increased still further. Where the break point occurs may depend on such factors as the nature of the speech signal (e.g., isolated syllables or words versus connected speech) and the duration of the discard segment. Using CNC stimuli and a 20-msec. discard segment, Beasley, Schwimmer, and Rintelmann (1972) got a sharp decline when the compression ratio was increased from 60 to 70%, but relatively little change between 0 and 60% (none between 0 and 30%). Daniloff, Shriner, and Zemlin (1968) obtained a similar result with 11 vowels in an /h-d/ context. Fairbanks and Kodman (1957), using phonetically balanced monosyllables as stimuli, did not get large intelligibility decrements until the compression ratio got to 80%, a fact that Beasley et al. (1972) attributed to their use of a 10-msec. discard segment, highly trained listeners, and relatively loud signals. Gerber and Scott (1971) got intelligibility of 84% with a compression ratio of about 67%. They obtained 90% intelligibility with a dichotic listening procedure in which one-third of the original speech was presented to one ear, and a different one-third simultaneously to the other.

In contrast to these results, Wingfield (1975) got a steep decline in the intelligibility of sentences as the compression ratio was increased beyond 40%. The duration of his discard segment was 20 msec. Given the generally beneficial effect of context on speech intelligibility, Wingfield's stimuli might have been expected to be understandable at higher compression ratios than were those of studies using isolated words or syllables. Several factors may have tended to work in the other direction, however. First, the (uncompressed) speech was recorded at the relatively high rate of 207 words per minute. Consequently, the rates of 414 and 518 words per minute that were associated with Wingfield's two largest compression ratios (50% and 60%) were equivalent in a sense to compression ratios of 66% and 73% had the original speech been produced at a rate of 140 words per minute that is close to what several investigators report as average (Abrams et al., 1944; Gregory, 1969; Maclay & Osgood, 1967). Second, the fact that the sentences were relatively long (10 words each) raises the possibility that the effects of signal degradation may have been amplified to some degree by short-term memory loss. More generally, while perception of words in sentences should benefit from contextual cues, the understanding of running speech requires a real-time processing of the input that the understanding of isolated words does not. Part of the difficulty with running speech, in other words, may be due to an inability to keep up with the input. That intelligibility and comprehensibility are not the same things has been stressed by Foulke and Sticht (1966) who made both types of measurements on compressed speech varying in rates from 225 to 425 words per minute. Both measures fell off as speech rate increased, but comprehensibility declined more rapidly than intelligibility at the higher rates.

On balance, experimental results seem to indicate that speech rates of up to at least 275 to 300 words per minute can be understood when produced with the fixed-interval sampling technique (Downing, 1971; Fairbanks, Guttman, & Miron, 1957; Fairbanks & Kodman, 1958; Foulke, 1968, 1971; Foulke, Amster, Nolan, & Bixler, 1962; Foulke & Sticht, 1966; Friedman, Orr, Freedle, & Norris, 1966; Garvery, 1953; Gregory, 1969; Langford, 1968, Reid, 1968). Whether much higher rates could be understood, given either better compression techniques or more effective training procedures for listeners, remains to be seen.

Several techniques for speeding speech have been investigated in addition to acceleration of playback speed and fixed-interval sampling. Diehl, White, and Burk (1959) tried discarding portions (from 50% to 75%) of between-word pauses. The resulting speech appeared to suffer no decline in comprehensibility; however, the method is limited in the degree to which it can increase the speech rate by the relatively small percentage of overall speech time for which interword pauses account.

Scott and Gerber (1972) have tested a pitch-synchronous method of time compression in which every other pitch period of voiced sounds is discarded. Applying the method to monosyllabic words containing no unvoiced consonants, they found the resulting speech to be more intelligible (92% versus 88%) than when the same words were compressed (also by 50%) with the fixed-interval sampling method.

We return to our original question of whether the possibility of using speeded speech in C^3 systems warrants further investigation. We believe the answer is yes. The evidence cited above is quite clear that people can comprehend speech that is presented at rates at least a factor of two greater than those typically encountered in normal conversation. Moreover, it is possible that even greater gains than those that have been reported could be realized as a result of the development of more effective compression techniques and listener-training programs. Some promising directions for future research, we believe are the following:

a. Search for better compression algorithms for use in computer-based systems. Several investigators have recently developed computer programs for compressing speech (Cramer & Talambiras, 1971; Scott, 1965). The advantage of this approach over the fixed-interval sampling method is that it makes possible the selective discarding of speech segments depending upon their presumed expendability. Thus, for example, silent intervals or large portions of sustained sounds might be discarded, whereas very brief sounds (e.g., stop consonants) or transitional sounds might be retained. What can be accomplished by means of selective sampling algorithms has not yet been thoroughly explored, and it remains to be seen how successful in producing intelligible speech such procedures will be. Experimentation along these lines should add considerably to our knowledge of speech perception, in addition to other consequences it may have. In the past, computer-based time-compression techniques have been viewed as of primarily theoretical interest, because of the relatively high cost of the computing capability they require (Dailey, 1974; Gregory, 1969); C^3 systems of the future, however, are likely to have more than sufficient computing power to do whatever speech compressing is desired. Certainly, failure to explore a potentially effective approach to information assimilation on the assumption that sufficient computing power will not be available is probably not justified.

One goal for research, therefore, should be the identification of more effective computer-based compression algorithms. This means, among other things, experiments addressed to the question of what portions of the speech signal carry critical information, that is to say, information that must be retained in order to preserve the comprehensibility of the compressed speech.

b. <u>Investigations of the utility of speeded speech for applications to various listener tasks</u>. The point has been made by others that in educational contexts speeded speech has proven to be especially useful for purposes of review and organization, in contrast to listening to unfamiliar material with the objective of assimilating and retaining as much new information as possible. A potential utilization of speeded speech in a C^3 system that would seem to warrant careful investigation is that of providing a commander with review-type briefings designed to refresh his memory concerning some aspect(s) of a tactical or strategic situation that he faces. A briefing concerning order of battle of an enemy force is one example of such an application. A review of recorded minutes of recent critical staff meetings would be another.

Another potential application of speeded speech that deserves some study is that of monitoring recorded speech for indications of material that is worth listening to carefully. Indications for which one might listen include specific words (e.g., proper names), the voices of specific speakers, or particular subject matter. Most of the experimentation to date has been addressed to the question of comprehensibility; consequently, little is known about how effectively people can monitor speeded speech for the purpose of detecting specific characteristics or content. The question of monitoring is of considerable practical interest as it relates to military intelligence operations. A common problem in this context is the need to listen to many hours of recorded speech in order not to miss a few minutes, or possibly a few seconds, of critical information.

c. <u>Design and development of prototype interactive speech-rate modification systems</u>. One can imagine a system that would permit a user to adjust the rate of the speech to which he was listening and even, perhaps, to select from among a variety of compression and expansion algorithms. Such a system would permit a commander or an intelligence analyst to modify the rate of the speech to which he was listeneing as a function of its content and his purpose for listening. Implementation and evaluation of such systems would require the collaboration of speech scientists, computer scientists, and human factors engineers.

Vision as an Information Channel

The eye is very richly innervated in comparison with the ear. There are more than 100 million receptor cells in the human eye as compared to approximately 25,000 receptor cells in the ear. This difference may be somewhat misleading, however, because the degree of funneling is much greater from receptor cells to ascending nerve fibers in the visual system than it is in the auditory system: The 100 million visual receptors map onto about 1 million fibers in the optic nerve; whereas the 25,000 auditory receptors map onto roughly 25,000 fibers in the auditory nerve in a more or less one-to-one fashion. Also, the amount and nature of processing that is done peripherally by the two systems is quite different.

Dynamic Range

The dynamic range of the visual system is also somewhat greater than that of the auditory system, as has been noted. The ratio of the intensity of a visual stimulus that is just barely detectable to one that can cause pain or damage, is approximately 13 or 14 orders of magnitude. This range really represents two systems, however, the scotopic system and the photopic system. The scotopic system is particularly well suited to night vision as it is sensitive to luminances between 10^{-6} and 1 millilamberts. The photopic system is better suited to daylight conditions as its sensitivity ranges from approximately 1 to 10^7 millilamberts. Notably, the full dynamic range of the combined visual systems is never available. The scotopic system requires nearly 30 minutes of dark adaptation to become maximally sensitive, and the detection threshold rises precipitously when the eye is exposed to bright light. Full light adaptation may take only 1 or 2 minutes.

Photopic vs. Scotopic Vision

There are important differences between the photopic and scotopic systems in addition to the difference in their sensitivity ranges. Whereas the photopic system discriminatively encodes differences in wavelength, the scotopic system does not; thus, we are effectively colorblind at night. Further, most of the photopic receptors, or cones, are densely concentrated in the fovea or the center of the visual field, and each cone in the fovea is more or less directly connected with one fiber in the optic nerve. This gives us a high degree of spatial acuity and sensitivity to colors at the center of the field under daylight conditions. By contrast, the photopic receptors, or rods, are distributed more evenly about the retina. The important exception is that the rods are totally absent from the fovea; thus, we are literally blind at the center of the visual field under low levels of illumination. Although the ratio of rods to cones is about 100 to 1, the peripheral receptors are quite poorly connected to the optic fiber. In contrast to the one-to-one pairings between fibers in the optic nerve and foveal receptors, a single fiber in the optic nerve may collectively represent several hundred peripheral receptors. This means that we are very poorly attuned to differences in color or form under night lighting, or, more generally, under peripheral vision. On the other hand, the pooling of peripheral receptors makes us highly sensitive to movement of objects in the peripheral field, even under conditions of dim illumination.

Visual Field

The "visual field," or the total area that is registered by the stationary eye, extends approximately 180° laterally by 150° vertically (Gibson, 1950), so in some sense we can see virtually everything before us. However, spatial sensitivity is not uniform across the field. Maximal visual acuity is circumscribed within the fovea, an area only 2° in diameter at the center of the field; the projection of this area is about the size of a dime held at arm's length. The acuity function drops off very rapidly towards the periphery. Within 8° from the center of the field, spatial sensitivity declines by 85% (Alpern, 1971). Beyond 25°, only the presence or absence of stimulation can be reliably discerned (Edwards & Goolkasian, 1974). Notably, these parameters are approximate. When an observer actively attends to a peripheral area of the field, sensitivity in that area is increased (Engel, 1971; Grindley & Townsend, 1968). Conversely, peripheral sensitivity declines when the foveal load is increased (Ikeda & Takeuchi, 1975), or when the observer is stressed (Easterbrook, 1959; Kahneman, 1973) or concurrently engaged in a relatively demanding task (Liebowitz & Appelle, 1969).

Saccadic Movements

It follows from the forementioned characteristics of the visual system that, beyond a few degrees of visual angle, the amount of information effectively conveyed by a visual display will depend primarily on the amount that is brought under direct fixation by the viewer. Unfortunately, people do not, and except under unusual circumstances apparently cannot, smoothly sweep their gaze about a stationary array (Alpern, 1971). Instead, stationary arrays are examined through series of discrete fixations. Moreover, visualization is suppressed during the eye movements or saccades between fixations (Volkman, Schick, & Riggs, 1968). Consequently, stationary arrays can only be assimilated in small, successive fragments.

For a wide variety of visual tasks, including reading, saccadic movements tend to occur at rates of approximately 3 to 5 per second. The rate can be affected somewhat by training on a specific scanning task (Gould, 1969), but only within fairly narrow limits. The musculature of the eye is capable of sustaining higher movement rates than usually attained (Alpern, 1971), which suggests that the rates that are typically observed reflect limitations on the rate at which incoming information can be processed by higher centers.

Visual Scanning

The Perception of Wide-Screen Displays

Perhaps the most obvious way to maximize visual input is to pack the visual field through the use of wide-screen displays. However, while the intuitive appeal of wide-screen displays is that they provide a means of presenting a lot of information simultaneously, we have seen that stationary arrays can only be assimilated in small, successive fragments. Beyond a few degrees of visual angle, the amount of information effectively conveyed by a display will depend less on its size than on the number and distribution of the viewer's visual fixations.

The use of wide-screen displays may actually be a disadvantage from this perspective. As the display size increases, so, too, may the probability that the viewer will fail to fixate upon, and, therefore, fail to assimilate, any given part of it. Unless an appropriate pattern of fixations is elicited, wide-screen displays may result in some loss in the amount of information effectively transmitted. In order to evaluate the potential utility of wide-screen displays, we need to understand better the factors governing the way in which they are scanned.

Determinants of Visual Scanning Patterns

In view of the above, it is important to understand the factors governing the distribution of the viewer's visual fixations. Several investigators have studied the scanning patterns elicited by complex, but coherent, stationary arrays such as photographs of real-world scenes and art reproductions (Antes, 1974; Buswell, 1935; Mackworth & Bruner, 1970; Mackworth & Morandi, 1967; Yarbus, 1967). In these studies, the sequence and durations of visual fixations were directly recorded from observers as they scanned the arrays. If an observer is asked to examine freely such a display for several seconds or minutes, a characteristic pattern emerges: much of the array may never be fixated at all; instead, a few selected and noncontiguous elements of the scene are repetitively fixated (see Fig. 107 in Yarbus, 1967, p. 172). Moreover, there is generally strong agreement among observers with respect to which elements of a scene receive the most and the least attention. These findings suggest that the distribution of visual fixations is guided by cues extracted from the peripheral field, and, further, that these cues are reliable in the sense that they are of general, rather than idiosyncratic, value.

Efforts to identify the critical properties of these cues have yielded few definitive guidelines for display design. The aspects of a complex meaningful visual scene that elicit observers' attention generally cannot be defined in terms of physical characteristics. That is, the salience of a pictorial element is determined by neither color, relative brightness, outline, contour, nor intricacy of detail (Mackworth & Bruner, 1970; Yarbus, 1967). The exception to this rule is movement. To quote Neisser (1967), "When something moves in a portion of the field to which we are not attending, it usually captures our attention at once" (p. 92). In fact, the peripheral visual field is exceptionally sensitive to movement (Sweet, 1953). It is known that for several species of lower animals, movement in the peripheral field triggers a reflex to bring the moving object into direct fixation (Bizzi, 1974; Schiller & Streyker, 1973), and this may be true for human beings as well (Gibson, 1966; Moray, 1970).

Those elements of stationary displays that attract the observer's attention seem to have only one common characteristic: they seem intuitively to be the most important or meaningful elements of the array. This observation has been borne out by studies using simpler, but methodologically tractable, visual stimuli. For example, Berlyne (1958) measured distributions of fixations across pairs of visual stimuli that differed in relative "informativeness" or redundancy. He found that subjects reliably spent more time looking at the more informative of the two stimuli.

Berlyne (1960, 1966) later distinguished two classes of stimulus properties that attract attention. The first class consists of physical properties, like contours, brightness, and colors. The second, more powerful, class consists of collative properties such as novelty, complexity, incongruity, or significance. Inasmuch as both classes of properties are effective for infants as well as adults, their attractiveness seems to reflect some basic human characteristics. In particular, the salience of the collative properties suggests that observers prefer perceptual events that demand some cognitive work to those that do not. Notably, this relationship is bounded: when stimuli become excessively complex or incongruous, they are perceived as incoherent noise; in this case, they are no longer attractive and may even be aversive.

Senders (1965, 1967) has provided a very different, and unusually convincing, demonstration of the relationship between eye movements and information. In his experiments, subjects were required to monitor the needle positions on a set of six dials, like those on an airplane instrument panel. In this situation, the probability that a dial will contain new information for the subject at any given moment depends jointly on the rate at which its needle is moving and the time elapsed since its last fixation. Senders found that experienced observers distributed their visual fixations across dials in remarkably close fit with an optimal allocation policy based on the mathematics of information theory. Moreover, when the statistical redundancy of the dials was manipulated, the observers adjusted their attention accordingly. Clearly, Senders' subjects were not responding to superficial aspects of the visual situation. Rather, their behavior reflects continuously adjusted estimations of the expected informativeness of each of the dials. Interestingly, this allocation policy was apparently not consciously determined by the observer. Several of Senders' subjects were quite confident that they could describe their fixation strategies, only to be surprised by the actual records.

The difficulty in predicting fixation patterns for complex, real-world displays is primarily one of estimating the informativeness of the elements of those displays. Several investigators have tried to quantify the importance of such elements. To this end, the standard technique has been to partition the array into glance-sized sections and to obtain subjective ratings of the recognizability or information contained in each of these sections (Antes, 1974; Mackworth & Morandi, 1967; Pollack & Spence, 1968). These ratings are then used to predict either the actual recognizeability of the fragments or the frequency with which they are fixated when the picture is scanned. The fragments of the picture may be presented piecewise, or they may be arrayed at once, put together like a jigsaw puzzle; in the latter case, they may be pieced together coherently, as in the original array, or they may be incoherently scrambled. As it turns out, the goodness of the predictions depends largely on whether the same method of stimulus presentation is used for both rating and testing phases of the experiment (Pollack & Spence, 1968). The implication is that the salience of a pictorial element cannot be completely ascribed to intrinsic properties of that element, but, rather, derives in part from its relation to the scene as a whole.

Some of these relational variables may be defined in terms of spatial attributes of the display; for example, the compositional factors long recognized by artists are surely influential. Others, however, are subjectively determined; for example, an observer's scanning pattern is strongly affected by the familiarity or clarity of the display as well as by his particular reasons for examining the display (Gould, 1967; Loftus, 1972; Simon & Barenfield, 1969; Yarbus, 1967).

Given the goal of reliably predicting or controlling what an observer will pick up from a complex array, both classes of relational variables are problematic. When possible, displays should be organized according to appropriate design principles. But the spatial relations for many kinds of visual arrays, for example, facsimile representations such as maps and photographs, cannot be contrived or arbitrarily rearranged. These cases are of particular concern since strategically critical elements may be relatively subtle in terms of the whole composition. Subjective dimensions of familiarity and bias may be even less manageable than they are identifiable.

In addition to the variables that are inherent to the display itself, there are several extrinsic factors that play some role in determining scan patterns (Gould, 1969). Information that the observer has about the target is one such factor. Searching for a target of known color, for example, in a field of multicolored targets will produce a scan pattern with different parameters than a search through the same field when the color of the target is not known (Williams, 1966). The observer's intention or purpose is another factor. Where an observer fixates when he is examining a picture for no specific purpose, for example, may be very different from when he is examining it in order to test some hypothesis about its contents. Still another factor is the observer's previous experience. Gould cites several studies that have shown that people who are trained in specific types of search tasks (e.g., inspecting of X-rays or integrated circuit chips) tend to fixate more frequently, but for shorter durations, than do untrained viewers. This finding is particularly interesting in the light of some results of Shaffer and Shiffrin (1972), and Loftus (1972) who have found that the accuracy of recall or recognition of visually presented stimuli depended strongly on the number of fixations that subjects made on the stimuli, but was relatively independent of the durations of the fixation periods.

Extrinsic Control of Focal Information

In view of the foregoing, for purposes of information presentation, it might be better to establish techniques of extrinsically controlling the focal areas of a display rather than to let the observer pick and choose. In this way, there would be less uncertainty about what he had seen. We are not aware of any studies that specifically examine such techniques, but there are some indications that the usefulness of such techniques may be very limited.

Perhaps the most promising possibilities in this domain are methods by which the subject is required to fixate on a pointer that moves in some programmed way relative to the visual array. The first method by which this could be accomplished is one of continuously moving the pointer about a stationary array. This method is feasible because even though people find it nearly impossible to voluntarily move their gaze smoothly across a stationary array, they are able, and in fact, almost compelled, to track a moving target. Alternatively, one could require the observer to maintain a stationary fixation point and move the array. In this case, a visible pointer is still needed to mark the fixation point because without some stationary reference, the observer's eyes will compulsively pursue the moving array. One of these methods may have some advantages over the other, but we suspect that that will be most strongly determined by characteristics of the array and the display facilities rather than by psychological factors. A more pressing question, and one for which we have no answers, concerns the depth and continuity with which people could extract, interpret, and integrate information from the visual array under either of these techniques.

The second class of possible techniques for extrinsically determining the focal areas of a visual display would consist in dividing the scene into glance-sized sections, and presenting these sections successively. The general reservation with respect to these sorts of procedures is that the exclusion of peripheral visual information may, on balance, prove to be a disadvantage rather than an advantage. Despite the narrow scope of the viewer's acute vision, whatever he gains from the peripheral field seems to contribute importantly to his comprehension of the scene as a whole.

This relationship has already been suggested by the patterns of eye movements that are typically generated when people scan complex arrays. In order for people to fixate selectively upon the most important elements of a display, they must be using peripheral information to generate hypotheses about which areas of the display should be of most interest.

Recently, Biederman (1972) has reported data that extend this notion. He briefly presented subjects with photographs of real-world scenes and then gave them a multiple choice, recognition test for some salient aspect of the scene. In one condition, the photographs had been cut into glance-sized pieces (leaving the test items intact), which were rearranged or jumbled. Biederman found that choice recognition accuracy was consistently worse for the jumbled than for the coherent arrays. This was true even when the subjects were told where to look or what to look for prior to display presentation. The implication is that peripheral information plays an important role in determining the perceptual processing or interpretation of coincident foveal information as well as the locus of subsequent fixations.

Biederman and his colleagues (Biederman, Rabinowitz, Glass, & Stacy, 1974) have argued that during a single fixation the observer is engaged in two interactive modes of visual processing:

(1) the identification of items under direct fixation, and (2) the construction of a holistic conceptualization of the scene. In keeping with this idea, they have shown that the identification of both single, cued items and the gist of an array are adversely affected by jumbling, even at exposure durations that are sufficiently brief to preclude eye movements. The importance of the subject's model of the scene is evidenced by the fact that jumbling reduces both the speed and the accuracy with which he can recognize or match some element from the display (Biederman, Glass, & Stacy, 1973; Pollack & Spence, 1968).

These results suggest that in order for methods of presenting selected foveal arrays to have the desired effect on visual information assimilation, they must be implemented with care. Unless the subject can appropriately interpret the areas of interest, it matters little whether or not he fixates upon them. Therefore, unless the coherence of the array as a whole can be preserved as the focal areas are switched in and out, such methods may actually result in a loss of control over information assimilation.

Peripheral visual information may also play an important role in the observer's ability to process spatial aspects of the array. One documented example of this has to with the orientation constancy. Normally, the visual environment is perceived as immobile when the viewer tilts his head from side to side. If, however, the field is confined to the foveal of the retina, this constancy breaks down (Wallach & Bacon, 1976). Thus, whenever rotations of the visual field are possible and should be taken into account, as under rotational sluing, the expense of the array should not be narrowly delimited.

The observer's ability to maintain his spatial orientation with respect to stationary displays may also be impaired by narrowly delimiting the field. This will be clear to anyone who has tried to wend his way through unfamiliar territory on the basis of a piecewise map, as in the map books that are often published for big cities. In this situation, the map reader may find himself incessantly flipping back and forth, not only between the maps for neighboring sections of the area, but also between those sections and the reduced map of the whole area, which itself does not depict any of the roads on which he will travel. The difficulty seems less attributable to visual factors than to the difficulty of maintaining a good idea of where one is, relative to where one is trying to go.

Analogous situations in which it is impractical to use a display size that would simultaneously support the optimal expanse and resolution of information must also arise in command and control operations. When available, interactive graphic display systems may provide the solution to these dilemmas, especially when an adequate solution could consist in realizing a compromise scale for the array. As this cannot always be possible, such systems should also have the capacity to zoom in on particular regions of a display, to back off for a more comprehensive but less detailed view, and to move a window continuously around over various areas of a display. In order to implement these techniques to the user's best advantage, we need a better understanding of the ways in which spatially distributed information is most easily and most usefully acquired for different purposes.

One important consideration may be that of whether or not the observer may be asked to remember the array. Noton (1970) has hypothesized that the memory for a visual array is composed of the ordered sequence of motor and sensory events corresponding to the scanpath executed during its examination. Memory for visual detail is embedded in the sensory component in the form of discrete packages of visual information extracted during successive fixations. The spatial organization of the array is captured in the eye movement commands.

In order to recognize a scene, the stored scanpath (or some ecologically permissible transformation, such as enlargement or rotation) is run off; if the visual information encountered on each programmed fixation adequately matches the stored information, the scene is judged to be familiar. If Noton's theory is correct, then techniques that eliminate eye movements may effectively eliminate the possibility of constructing useful cognitive representations of the information presented.

Support for Noton's theory of pattern perception exists only in observations that individuals do adopt fixed scanpaths in examining (Noton & Stark, 1971), reexamining (Noton & Stark, 1973), and even imagining (Hebb, 1968; Kahneman, 1973) visual arrays. While these data are consistent with the theory, they are by no means compelling. A critical test of the theory would be valuable from both theoretical and practical perspectives.

In any case, when cast more generally, the suggestion that eye movements, in and of themselves, may play a functionally significant role in visual information processing activities seems to be valid. Although techniques that obviate eye movements may be advantageous for some applications, they clearly are not for others. We will return to these issues in subsequent sections. For now, suffice it to say that the value of such techniques can only be assessed in the context of the relevant task requirements.

Eye-Fixation Measurement as a Computer-Input Method

The development of computer graphics and computer-controlled displays makes possible the storage and immediate accessibility of far more data or information than one would want to show on a single situation display. Consider, for example, a display that showed the deployment of various types of weapons systems, supplies, and troops in some geographical area. To represent everything that is known about the situation on a given display could make the display so crowded that it would be difficult or impossible for the observer to make sense of it. What one wants is the flexibility of inspecting various subsets of the data on demand. One might like, for example, to be able to have shown, on request, the deployment of armor, the location of ammunition dumps, roads that are passable by specified types of vehicles, the locations of bridges, and so forth. Also, one might wish to be able to call up coded information that might be stored in the computer regarding specific items on the display.

With respect to the latter point, systems that permit such call-up of information at the moment require that the user identify explicitly the item about which he wants the information to be displayed. He may do this by describing the item verbally (via a typed input to the computer), or by locating a cursor on the display so that it points to the item of interest, or by specifying its spatial coordinates. An alternative method for identifying items of interest would be simply to look at them. The rule of thumb might be used that any specified action is to be taken with respect to whatever it is that the user is looking at at the moment that the action is requested. This is well within the current state of the art in measuring eye movements or eye fixations.

Cognitive Correlates of Eye Movements

While the visual information effectively conveyed by a display depends largely on the observer's eye movements, the observer's eye movements depend largely on higher-order, cognitive processes. It is because of this latter relationship that the visual assimilation process seems so recalcitrant to prediction or control. However, there is the possibility of turning this relationship to our advantage. Perhaps by monitoring the observer's ocular activities, we can gain useful information with respect to his state of mind.

A strong correspondence is often found between the locus of visual fixation and the topic of thought. To borrow an example from Kahneman (1973):

> "Let the reader attempt to think of an object in the room, and he will soon become aware of a tendency to look at that object. When one person in a group conversation mentions the name of one of the people present, the collective gaze of the group is immediately drawn to the person mentioned" (p. 60).

Kahneman and Lass (1971, cited in Kahneman, 1973) analyzed this correspondence through a series of experiments. In one case subjects were shown an array of drawings of common objects like a car, a person, a tree, and an airplane, and asked questions like, "What makes of automobiles can you remember?" They found that subjects reliably fixated the relevant drawing while answering the question. Moreover, if the pictures were no longer there, the subjects tended to fixate the location where the relevant item had been.

These sorts of visual fixations have been found to correlate with fairly subtle aspects of cognitive processing. For example, Carpenter and Just (in press) found that when subjects encounter anaphoric references (e.g., pronouns) in written text, they tend to switch their gaze immediately and directly to the last explicit reference to the object in question. Cooper (1974) presented a matrix of pictures to subjects as he read them a story. He found that subjects systematically switched their eyes from picture to picture in keeping with, and even in anticipation of, the semantic thread of the story. Not only did they reliably fixate pictures as they were explicitly named in the story, but also pictures that were only relevant through complex associations.

The strength of this tendency is especially interesting, given that there is no evidence that subjects incorporate the information under fixation into their ongoing thought processes. Although subjects consistently prefer the most relevant object before them, Kahneman and Lass (1971) found that when only a single object was present during questioning, subjects focused upon it just as intently, regardless of whether it was relevant or irrelevant. Further, the task-relevance of the fixated item had no detectable effect on performance.

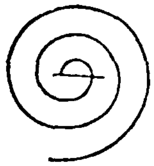
In short, these sorts of fixations seem to be direct reflections of the observer's concurrent semantic set. This fact, in conjunction with the indications that the fixated information does not interfere with or alter the observer's thought processes, suggests the possibility of exploiting these sorts of fixations to extend or corroborate subjects' protocols, or to detect sources of conflict or uncertainty in reports which they are either generating or listening to. The potential utility of such a technique depends on some as-yet unanswered questions such as how compulsive these fixation tendencies are.

The functional significance of these fixations is related to the regulation of attention. Several researchers have suggested that the onset of a new visual stimulus, as after a saccade or a blink, generally triggers a shift in thought (Kahneman, 1973; Loftus, 1976; Potter & Levy, 1969). If this is true, then steady fixation is a means of promoting concentration on the task at hand. Kahneman and Lass' data, discussed above, indicate that the tendency to fixate task-relevant information may be a resulting, as opposed to a contributing, factor of focused attention. However, there are situations in which the opposite seems to be true. Gopher (1971, cited in Kahneman, 1973) has shown that sensitivity for specified auditory stimuli in a noisy background may be augmented by visually fixating in the direction from which they emanate. This spatial sensitization may contribute to people's preference for face-to-face contact in oral conversation;

conversely, it may help to explain the surprising lack of enthusiasm for videophones. In any case, the moral is that, given a task that requires considerable concentration or selective attention, it is inadvisable to concurrently require incompatible visual activity, no matter how trivial.

If ocular activity systematically reflects mental activity, then dynamic mental activity should elicit high levels of ocular mobility, just as mental concentration tends to be associated with steady fixation. In fact, this seems to be the case: whereas relaxed, bored, and receptive states are associated with a marked reduction in eye movements, active thinking is typically associated with a very high saccadic rate (Antrobus, Antrobus, & Singer, 1964; Lorrens & Darrow, 1964).

Again, the functional significance of these eye movements is not entirely clear. In some cases they seem to support what are more or less visual activities. The example given below is taken from Yarbus (1967, p. 201). The problem is to determine the length of the spiral, using the straight line as the unit of measurement. If you are allowed to step your fixations around the spiral as you measure off units in your mind's eye, the task is not that hard. However, if you are required to fixate on the center of the spiral while generating your estimate, the task becomes quite difficult. The effect can be more persuasively demonstrated by asking subjects to make spatial judgments on after-images. In this case, the image moves with the eye so that new fixations are impossible. Nevertheless, the subject's tendency is to run through a variety of physical contortions in his effort to move his point of fixation relative to the image, and tasks that are ordinarily quite easy may become next to impossible (Yarbus, 1967).



One function of eye movements during problem solving, then, is that they mediate spatial comparisons. Yet, they seem to have some more general, nonvisual significance as well. This notion is suggested by the fact that ocular activity increases with mental work even in congenitally blind people (Amadeo & Gomez, 1966). More compelling evidence has recently been provided by two Russian scientists, Zinchenko and Vergiles (1972, cited in Gould, 1976). Their subjects were asked to perform several tasks, ranging from visual search to problem-solving tasks, on displays that subtended 15° to 30° of visual angle. Like Yarbus (1967), they found that solutions came much more easily under normal viewing conditions than when the image was stabilized on the retina. Their most provocative discovery, however, was that when the image was stabilized and subjects were not allowed to move their eyes, most of the problems could not be solved at all. That is, eye movements seem to be a necessary component of the problem-solving process even when they are worthless for purposes of changing the visual input. Again, the implication is that if we restrict people's ocular activity we may interfere with their mental activity.

Patterns or sequences of visual fixations may also provide insights to the mind of the beholder. For example, given that the focal points of a complex, real-world display are determined by subjective definitions of the importance of its elements, then the pattern of fixations should reveal something about what the observer considers important. Yarbus (1967) has demonstrated the lawfulness of this relationship by asking observers to study Repin's painting, "An Unexpected Visitor," under different instructions:

> "For example, in response to the instruction 'estimate the material circumstances of the family shown in the picture,' the observer paid particular attention to the women's clothing and the furniture (the armchair, stool, tablecloth, and so on). In response to the instruction 'give the ages of the people shown in the picture,' all attention was concentrated on their faces. In response to the instruction 'surmise what the family was doing before the arrival of the "unexpected visitor,"' the observer directed his attention particularly to the objects arranged on the table, the girl's and the woman's hands, and to the music. After the instruction 'remember the clothes worn by the people in the picture,' their clothing was examined. The instruction 'remember the position of the people and objects in the room' caused the observer to examine the whole room and all the objects. His attention was even drawn to the chair leg shown in the left part of the picture which he had hitherto not observed. Finally, the instruction 'estimate how long the "unexpected visitor" had been away from the family,' caused the observer

to make particularly intensive movements of the eyes between the faces of the children and the face of the person entering the room. In this case he was undoubtedly trying to find the answer by studying the expressions on the faces and trying to determine whether the children recognized the visitor or not" (p. 192).

Russo and Rosen (1975) have demonstrated a particularly interesting application of visual fixation data. In their experiments, subjects' eye movements were monitored as they chose their most preferred of six used cars. On each trial, each of the subjects' options was specified by a value on each of three dimensions (make, year, and mileage), and all six sets of descriptors were arrayed simultaneously. During evaluative processing, fixation patterns consisted in oscillations between pairs of cars. This finding carries the important implication that preference in this situation, despite its explicitly multidimensional, multialternative format, was established through paired comparisons. Subjects' verbal protocols were consistent with this implication. The durations of the visual paired comparisons were directly related to the total judged utility of the pair. Moreover, when subjects were asked to make their choice through a procedure of successive elimination, the duration of the oscillating sequences reflected the difference in the utility of the alternatives under consideration. Russo and Rosen are also quick to point out that, given a longer menu representing a broader range of subjective values, many items would probably be eliminated immediately. One wonders how successfully Russo and Rosen's technique could be extended for purposes of assessing an individual's preference space for complex, multidimensional choice alternatives.

Although duration of the fixation correlates with the depth or complexity of stimulus processing, it tends to be a relatively uninformative measure of psychological activity. The reason is that an obligatory latency of about 200 msec. is attached to the initiation of a saccadic movement anyhow (Alpern, 1971). Since recognition of familiar stimuli may take as little as a few tens of milliseconds, there is considerable time for additional processing during the course of a normal fixation. Consequently, the duration of the fixation is only sensitive to encoding activities that are sufficiently complex to extend it beyond its normal term.

Gould (1967; Gould & Dill, 1969; Gould & Peeples, 1970; Gould & Schaffer, 1965, 1967) has extensively investigated the relationship between fixation durations and visual information processing. In these experiments, the subject was presented with a 3 x 3 array of patterns with the task of determining how many of the outer patterns matched the center or standard pattern. The standard pattern elicited the longest fixation; presumably, this reflected efforts to memorize or create a template of the standard. Further, matching or target patterns were fixated for significantly longer durations than nontarget patterns. By requiring one group of subjects to count the number of patterns that mismatched, rather than matched, the standard, Gould and Schaeffer (1967) were able to attribute this difference to two separate tendencies: (1) fixations were longer for target patterns than for nontarget patterns, regardless of whether they matched or mismatched the standard, (2) fixation durations increased with the similarity of the pattern to the standard, regardless of whether or not it was a target. Because this pattern of results obtained across a wide variety of stimulus domains, Gould has argued that it reflected basic, everyday aspects of human pattern processing.

Finally, the degree of pupillary dilation may also provide useful indices of higher order processes. For one, pupil diameter varies directly and reliably with the observer's interest in the information under fixation (Libby, Lacey, & Lacey, 1973). Further, Kahneman (1973) has argued that pupil diameter is a particularly interpretable index of autonomic activity, reflecting momentary changes in the subject's expenditure of general mental effort.

Rapid Serial Presentation of Foveal Displays

The effective rate of information presentation under visual display techniques generally depends on the observer's scanning rate. Although a single visual fixation may last as long as several seconds, there is evidence that the amount of information that is assimilated asymptotes within a few hundred milliseconds (Loftus, 1972). To the extent that this is true, longer fixations are a waste of time. Thus, the use of successive, foveal displays might offer a means of pacing visual information presentation for maximal efficiency.

Indeed, successive display techniques might provide a means of accelerating the rate of visual information processing beyond typical limits. Visual arrays are normally scanned at a rate of 3 or 4 fixations per sec., and people are physiologically incapable of executing more than 5 or 6 saccadic movements per sec. Information could be presented much more rapidly through successive display techniques since eye movements would be unnecessary. The potential utility of such techniques depends upon the rate at which people can process visual information, irrespective of eye movements, and this is the issue to be addressed in this section.

This problem was given considerable attention by early psychologists. The interest in the problem was initially motivated by a philosophical question: If the mind is unitary, how can it conceive of more than one idea or object at a time? The earliest "experiment" was merely a demonstration that in fact the mind could, and consisted in tossing handfuls of beans upon a table and asking subjects to estimate their number from a single glance. This soon evolved into an interest in the so-called "span of apprehension" or the ability to make instantaneous judgments of numerosity. The methodology became refined, and, perhaps most importantly, the tachistoscope was invented.

The tachistoscope offered unprecedented methodological flexibility in terms of both stimulus selection and temporal control. It was already known that visual information was acquired primarily during fixations and that the minimum time to change fixations was about 200 msec. As the tachistoscope provided a means of presenting a stimulus for less than 200 msec., it also provided a means of insuring that the subjects' responses would be based on information acquired during a single glance--or, in accord with the times, in a "single perceptual moment." The ability to manipulate display times led to the investigation of the speed of visual apprehension. Further, the scattered beans were replaced with systematically constructed arrays of dots, and eventually with letters, words, colors, and geometrical patterns.

The major findings of these early investigators are as follows. First, the span as measured by the number of reportable items was found to vary systematically with the coherence of the information in the display. For example, the span of apprehension is generally greater for digits than for unrelated letters, and for unrelated letters than for geometric patterns. Similarly, the typical span for any particular item type can be increased by clustering the items into a higher order pattern: more dots were reported if they were arranged in subgroups; more letters were recognized if they were arrayed as words; and more words were recognized if they were semantically interrelated. Second, the span was found to vary systematically between individuals, regardless of the item type. Finally, within the limits of a single fixation, the span was found to be relatively insensitive to the duration of the exposure; as long as the exposure duration exceeded some very brief minimum, the full span was generally reported.

The interpretation of these data has been tempered by two fundamental reservations. First, several investigators suggested that the subject may be able to see much more of a briefly exposed display than he can report--that the observed span of apprehension reflects limitations on verbal memory rather than visual perception (Cattell, 1885; Glanville & Dallenback, 1929; Woodworth, 1938). Second, it was repeatedly suggested that the sensory response or image evoked by a brief exposure might persist beyond the termination of the display; if this were true, then it would be inaccurate to equate the "time for perception" with the duration of the stimulus exposure.

These reservations were only recently borne out by empirical research (Sperling, 1960). Sperling began by replicating the earlier findings that the span of apprehension was invariant across a range of exposure durations. Then he devised an ingenious demonstration that this upper bound was due to memory rather than visual limitations. He presented his subjects with as many as twelve, unrelated alphanumeric stimulus items, arranged in three rows of four items each. Immediately following the offset of the display, he presented an auditory signal which indicated which row of the display the subject was to report. Under these conditions, the proportion of correct responses implied that subjects had at least twice as much visual information available as they were able to report. The results of subsequent investigations suggest that this information resides in the form of a virtual image or icon of the display. The icon normally persists for about 200 msec. beyond actual display termination. It can, however, be totally and immediately displaced or overwritten by a subsequent visual input (Averbach & Coriell, 1961; Haith, 1970; Sperling, 1963).

Sperling (1963) speculated that the purpose of the icon might be that of a visual buffer memory. The function of its persistence would be to maintain images across saccades. In this way, the visual world would switch cleanly from image to image instead of flashing on and off. In addition, it would provide a little extra processing time for each image. Since a new image totally displaces information in the icon, successively fixated materials will not visually interfere with one another.

If one wants to measure the speed of visual perception, then one must find some means of determining the subjective duration of the stimulus. The fact that the contents of the icon are displaced by subsequent visual stimuli provides a means of doing so. That is, the effective exposure duration of a visual stimulus can be estimated as the actual exposure duration of the stimulus plus either 200 msec. or the time until the onset of a second or masking stimulus, whichever is briefer. Thus, by setting the exposure duration of the stimulus at a minimal value and manipulating the latency of the masking stimulus, the rate of visual perception can also be determined.

Using this technique, Sperling (1963) discovered that for a brief period after stimulus onset, no information was available from the display; the duration of this period depended on various exposure parameters, such as the brightness of the stimulus field. After this period, however, letters were extracted at the approximate rate of one every 10 msec., asymptoting at four or five items, and this rate was relatively independent of exposure parameters. Sperling's interpretation of these findings was that items in the icon were serially scanned at a fixed rate, verbalized, and removed to an auditory rehearsal buffer. The asymptote of four or five items purportedly reflected the capacity limit of the verbal rehearsal loop.

This interpretation must be altered in view of more recent evidence. First, the items in the icon are apparently scanned or encoded in parallel rather than in series (Adams, 1975; Estes & Taylor, 1966; Gardner, 1973; Sperling, 1967). Second, the maturation rate of the percept is not fixed, but depends on the complexity of the stimulus. Basic visual features, like color, size, and position, become available almost immediately; in fact, these features are so readily evident that they can be used instead of Sperling's auditory tones for partial report procedures (Averbach & Coriell, 1961; Clark, 1969; Von Wright, 1968, 1970). Further, simple figures, like Landoldt Cs, can be encoded two to three times faster than digits (Allport, 1968), and familiar letter sequences are scanned faster than less familiar letter sequences (Adams, 1975; Mewhort, Merikle & Bryden, 1969). Third, visually presented items need not be verbally recoded in order to be remembered, but the same capacity limitations hold regardless (Scarborough, 1972; Sternberg, 1969; Wolford & Hollingsworth, 1974).

Even so, Sperling's basic notion, that information had to be recoded or "read-out" of the icon (though not necessarily verbally) in order to be retained, was correct. Turvey (1967) has argued that iconic information that is not recoded leaves no trace on the system whatsoever. He found that the probability of recalling a visual stimulus did not improve even after 54 repeated 50-msec. exposures of a single display which the subject had been required to attend to but not to report. This is in marked contrast to the effects of stimulus repetition on either short-term memory (Hebb, 1961) or long-term memory (Rundus, 1974). Moreover, unanalyzed iconic information seems to be virtually disconnected from the rest of the system. Neither the capacity of the iconic store nor the availability of its contents are affected by concurrent performance of other information processing tasks, and, conversely, the performance of other information tasks is not impaired by concurrent loads of iconic memory (Doost & Turvey, 1971; Turvey, 1966).

It follows that for purposes of determining the maximum feasible rate of display presentation, we should be concerned with the time required for the image to be encoded rather than with the time for it to reach full clarity. Moreover, encoding activities may continue after the stimulus has been masked. This is especially true when the masking stimulus consists of non-informative visual noise, as in the above-cited studies. This inference is supported by the fact that when the second stimulus is categorically similar to the first (e.g., both are letters or words), it maintains full masking capacity for at least 250 msec. (Merikle, 1974; Potter, 1976; Turvey, 1973). Importantly, this latency corresponds to the full duration of a normal visual fixation. The implication is that the speed with which people can identify the elements of a visual display cannot be increased beyond its natural limitations through rapid serial presentation of the displays.

If people are asked to detect or recognize prespecified visual events, rather than to identify and remember them given no foreknowledge, the situation is much different. Potter and Levy (1969) found that subjects' memory for pictures presented in rapid sequence depended strongly on the presentation time per picture. In their experiment, the probability of correctly distinguishing any of the 16 test pictures from 16 novel distraction items was only 0.16 if the pictures had been presented for 125 msec. apiece. From that point, accuracy rose nearly linearly with log presentation time, exceeding 90% only when the pictures had been presented for as much as 2 sec. apiece. By contrast, when subjects were asked simply to determine whether or not a previously seen target picture was a member of the sequence, accuracy asymptoted near 100% when the pictures were presented for as little as 167 msec. each. Moreover, the accuracy of target detection was equally good when the target picture had not been shown to the subject in advance, but only vaguely described, e.g., "a boat" or "two men drinking beer." Taking these results together with some from further research, Potter (1976) has concluded that a complex visual scene may be seen and understood within about 100 msec., but requires an additional 300 msec. of processing to be immunized from conceptual masking by subsequent scenes. Prior knowledge about a scene serves to prime the conceptual processing.

The ecological significance of Potter's results is that they reconcile the fact that we cannot remember unrelated scenes presented in rapid succession with the fact that we normally scan our environment in glances that last only one-third of a second. If the information in a glance is roughly anticipated, it is assimilated. Thus, our normal rate of scanning efficiently supports the processes of confirming or refuting our expectations about our environment and of searching for relevant information.

The significance of Potter's results with regard to C^3 technology is that, although rapid serial displays of visual information are not advantageous for purposes of initial information presentation, they may be very effective and efficient for purposes of information detection or retrieval. The latter suggestion is supported by a recent study by Sperling, Budiansky, Spivak, and Johnson (1971) in which they tried to push the detection function to its limit. Sperling et al. presented matrices, each of which was composed of from 2 to 25 unrelated alphabetic characters, in rapid succession. Somewhere, in one of the matrices, one of the alphabetic characters had been replaced with a digit, and the subject's task was to detect the digit. With practice, one of their subjects became able to locate the digit in a series of 16-item matrices, presented for only 40 msec. apiece; that is, this subject was apparently scanning the characters at a rate of 125 per sec.

What is more, Sperling et al. found that subjects could, with extensive practice, learn to detect any unspecified digit in a field of letters at comparable rates. Several investigators (e.g., Brand, 1971; Ingling, 1972; Posner, 1970) have found that adult subjects can discern whether a character is a letter or a digit before having determined its identity. Thus, one might attribute the performance of Sperling et al.'s subject to a peculiarity in the way we process alphanumeric items (Nickerson, 1973; Turvey, 1974). However, Shiffrin and Schneider (1975) have demonstrated that, with practice, subjects are able to detect any of a set of as many as 10 arbitrary items in a field of visually similar items, as easily as a single, physically specified item. This point is important for C^3 applications inasmuch as it suggests that, with practice, rapid serial displays could provide an efficient technique for presenting information to be scanned for several different arbitrary events at once. The fact that rapid serial displays can be perceived or understood if they are anticipated raises the possibility that they may also be useful for purposes of refreshing one's memory. To our knowledge, this application nas not been studied to date.

CHAPTER 2

MONITORING AND ATTENTION SHARING

Introduction

The human operator in a command/control/communication system may be called upon to monitor inputs (or stimuli) from a particular source, targets of a particular type, a particular attribute of complex stimuli, and outputs (or responses) in a particular category (Treisman, 1969) -- looking for patterns and trends as well as singular, discrete events -- all the while he is commanding/controlling/communicating. The job of simply monitoring particular targets, as exemplified by the radar or sonar observer, does not exist in a modern C^3 system. In today's or tomorrow's integrated intelligence/target-engagement system, the functions of surveillance and warning are largely automated, so that the important "signal events" for the human operator are those representing the operating status of the system, and are events that are more often inferred than directly sensed (Kibler, 1965).

Signal events must be inferred when a subsystem fails and the human must temporarily take over its functions; when a situation outside the capabilities of the automatic equipment arises and the human must act to extend the system's range of performance (Adams, 1965); and also when the system is functioning according to plan and the situation is evolving within anticipated limits. In a taxonomy of C^3 activities offered at a recent ARPA-sponsored conference, monitoring (item 3 in that taxonomy) feeds directly into what may be a quite complex process of event classification and choice of response, or decision-making (item 4), and is surrounded by resource allocation (1), scheduling (2), and creation and invention (5) (Pew, 1976). In short, while monitoring per se is as important in military settings as it ever was, it is now more complicated, and only part of the job.

Despite the inherent complexity of modern military monitoring, and of the setting in which it takes place, psychological research on monitoring behavior has continued primarily in the vein in which it began thirty years ago (Mackworth, 1948), when it was motivated by the radar observer, and it has, indeed, been regarded as research on "vigilance." With few exceptions among hundreds of studies, including hundreds in the last decade, the signals employed have been weak, brief, unidimensional, from a single source, and of fixed probability; the responses required have been related only to signal existence; and the observer has had nothing else to do (see reviews by Broadbent, 1971; Davies and Tunes, 1969; Mackworth, 1970; and the more recent bibliography compiled by Parasuraman, 1976a). These studies have not seriously involved selective attention, temporal or spatial integration of sensory information (trends or patterns), inference, interpretation, action, responsibility for control, or task interference. Again with a few recent exceptions, these studies have concentrated on time-dependent changes in performance, to the exclusion of situationally-determined effects, and to the exclusion of absolute levels of performance.

Nonetheless, we believe that the last decade of research on vigilance has provided a good basis for research on C^3

monitoring. Indeed, it now appears that the very simplicity of the classical vigilance task has enabled laboratory research to attain its primary objective, namely, the isolation and identification of some basic variables that operate in the complex, practical situation.

A key development in this respect has been the application of modern signal detection theory to vigilance. Signal detection theory, based on statistical decision theory, appears now to offer a sound theoretical framework that encompasses significant variables in applied monitoring tasks, and to provide a definition of functional relationships in quantifiable terms (Broadbent, 1971; Drury and Fox, 1975; Howell, Johnston, and Goldstein, 1966; Jerison, 1976; Mackworth, 1970; Parasuraman and Davies, 1976; Swets, 1976; Swets and Kristofferson, 1970). This theory seems thereby to fulfill the requirement that was identified by Jerison and Pickett (1963) and Kibler (1965) as essential to generalizing from laboratory to field. Also apparent is that much of the power of the theory remains to be tapped in studies of monitoring.

A second key development, accomplished in the past year or two, has been the careful and imaginative taxonomic analysis of monitoring tasks by Parasuraman and Davies (1976). These investigators, using the ARPA-sponsored taxonomy of tasks developed by Levine, Romashko, and Fleishman (1973), have shown how a consideration of the abilities required by different types of display can account for the major perceptual effects in vigilance experiments, and how that consideration also reveals consistent individual differences across tasks. Their results suggest, in addition, how processes of memory and effort may be related to perceptual processes.

In this paper we review the main outcomes of vigilance research influenced by signal detection theory (SDT), and examine the role that this theory and the new leads from the taxonomic analysis of vigilance tasks can play in studying complex monitoring. Our examination will be aided by a NATO conference held this year on "Vigilance II. Relationships among Theory, Physiological Correlates and Operational Performance."

We begin this review with the emphasis that SDT places on response strategies, or decision criteria, as distinguishable from perceptual efficiency (and observe in passing that latency data as well as accuracy data, and physiological recordings, support the distinction, in vigilance as well as in other signal detection studies). We establish that the criterion for a positive detection decision becomes more stringent over time in essentially all vigilance experiments, and that perceptual efficiency declines over time when and only when the monitoring (a) is done under time pressure and (b) is of a class of display that places a demand on memory.

We then see that, if displays are properly classified, individual differences are consistent over tasks. The next few sections discuss situational determinants of response strategies, the lack of effect of situational changes on perceptual efficiency, and the measurement of absolute levels of performance, as they have been considered in recent vigilance research. Industrial inspection is then discussed as supplying another research paradigm relevant to C^3 monitoring.

At that point we return to consider some implications of the finding that perceptual decrements occur when the display incorporates a memory requirement and the display's event rate exerts a time pressure -- specifically, how perceptual efficiency, and the decision criterion as well, may be affected by the effort required of the monitor, with the "cost of observing" as an intervening variable. Those relations bring us to some thoughts about selective and divided attention and interference among sub-tasks in the complex, practical task. In the process we describe specifically the few, recent experiments that have simulated aspects of the complex task: experiments that involve patterns, trends, changes in signal probability, multiple sources, and competing sub-tasks.

The concluding pages suggest and characterize two lines of laboratory research that should generalize rather directly to the field. The first focuses on the cost of observing for the monitor who has much else to do. This research area can benefit from a greater attention by psychologists to quantitative models of selective attention and control developed in the engineering literature. Our review of these formulations will be aided by another NATO conference held this year, entitled "Monitoring Behavior and Supervisory Control." The second line of research suggested focuses on the inference of signal events consisting of patterns and trends. This research area can further tap the power of SDT for behavioral studies of monitoring, by drawing on its models of pattern classification and deferred decision. These two suggested lines of research are particularly responsive to Kibler's (1965) call for laboratory research relevant to modern, complex man-machine systems.

Response Strategies vs. Perceptual Efficiency

The application of SDT to human observers has established that yes-no responses concerning signal existence reflect a decision criterion, i.e., a response strategy, as well as a perceptual capacity. The theory's primary analytical tool, the relative (or receiver) operating characteristic (ROC), shows how any given level of perceptual efficiency may be accompanied by any one of the full range of balances between the proportions of the two types of error. The ROC yields independent measures of perceptual efficiency and the decision criterion. The former is determined jointly by the observer's sensitivity and the strength of the signal, while the latter is determined jointly by the *a priori* probability of a signal and by the payoff matrix, i.e., the values of the two possible correct decisions and the costs of the two possible errors (Green and Swets, 1966).

The Initial Application of SDT to Vigilance

The initial result of applying SDT in experiments on vigilance -- in about thirty studies in the 1960's -- was to show that the so-called vigilance decrement, namely, a decline in the proportion of true or correct detections, usually reflected a progressively stricter decision criterion, rather than a decline over time in perceptual efficiency as assumed earlier. That is, the decline in the proportion of true detections was found to be reliably accompanied by a decline in the proportion of false detections; in most cases, the decline in the proportion of false detections was of a magnitude such that all of the decline in the proportion of true detections could be attributed to a criterion shift (Broadbent, 1971; Mackworth, 1970; Swets and Kristofferson, 1970).

A recent analysis of decision latency gives further evidence for a criterion change rather than a perceptual decrement. Whereas a decline in perceptual vigilance (or arousal) would lead us to expect an increase in the latencies of all response categories, the latencies of both correct and incorrect rejections (of "No" responses) have been found to decrease or remain stable over a work period. This result is consistent with the earlier interpretation of latencies in the SDT context that latencies are shorter as the perceptual evidence relative to signal existence (the observation value) is more distant from the criterial level of evidence required for a positive response (Carterette, Friedman, and Casmides, 1965; Emmerich, Gray, Watson and Tanis, 1972; Laming, 1973) -- and suggests that the criterion becomes more stringent as the task proceeds. Thus, both of the main performance indices, accuracy and latency, may be understood within the SDT framework and lead to the same conclusion about a criterion change (Parasuraman, 1976b).

A recent analysis of physiological recordings, specifically, of the late components of evoked potentials, also lends support to interpretations of vigilance changes in terms of criterion shifts. The basic finding relative to evoked potentials in detection experiments is that the amplitude and latency of the late components are systematically related to the strictness of the observer's decision criterion, as influenced by probabilities and payoffs in a binary-response task, and as estimated from the different response categories in a rating task (e.g., Paul and Sutton, 1972; Squires, Squires, and Hillyard, 1975a, 1975b). Davies and Parasuraman (1976) reported three experiments which suggest that the same measures provide reliable correlates of decision processes in vigilance; in one of their experiments, for example, a change over time to a stricter criterion, with constant perceptual efficiency, was accompanied by the expected change in late-component latencies.

A few experiments in the 1960's showed a decline in perceptual efficiency in addition to a criterion change. These exceptional experiments had two defining characteristics: they employed (1) high event rates, requiring almost continuous attention to the display, and (2) visual displays, which were thought of as loosely "coupled" to the observer relative to auditory displays, meaning that a waning of attention could result in events (signal and non-signal) being missed altogether (Broadbent, 1971; Mackworth, 1970; Swets and Kristofferson, 1970).

The New Taxonomic Analysis

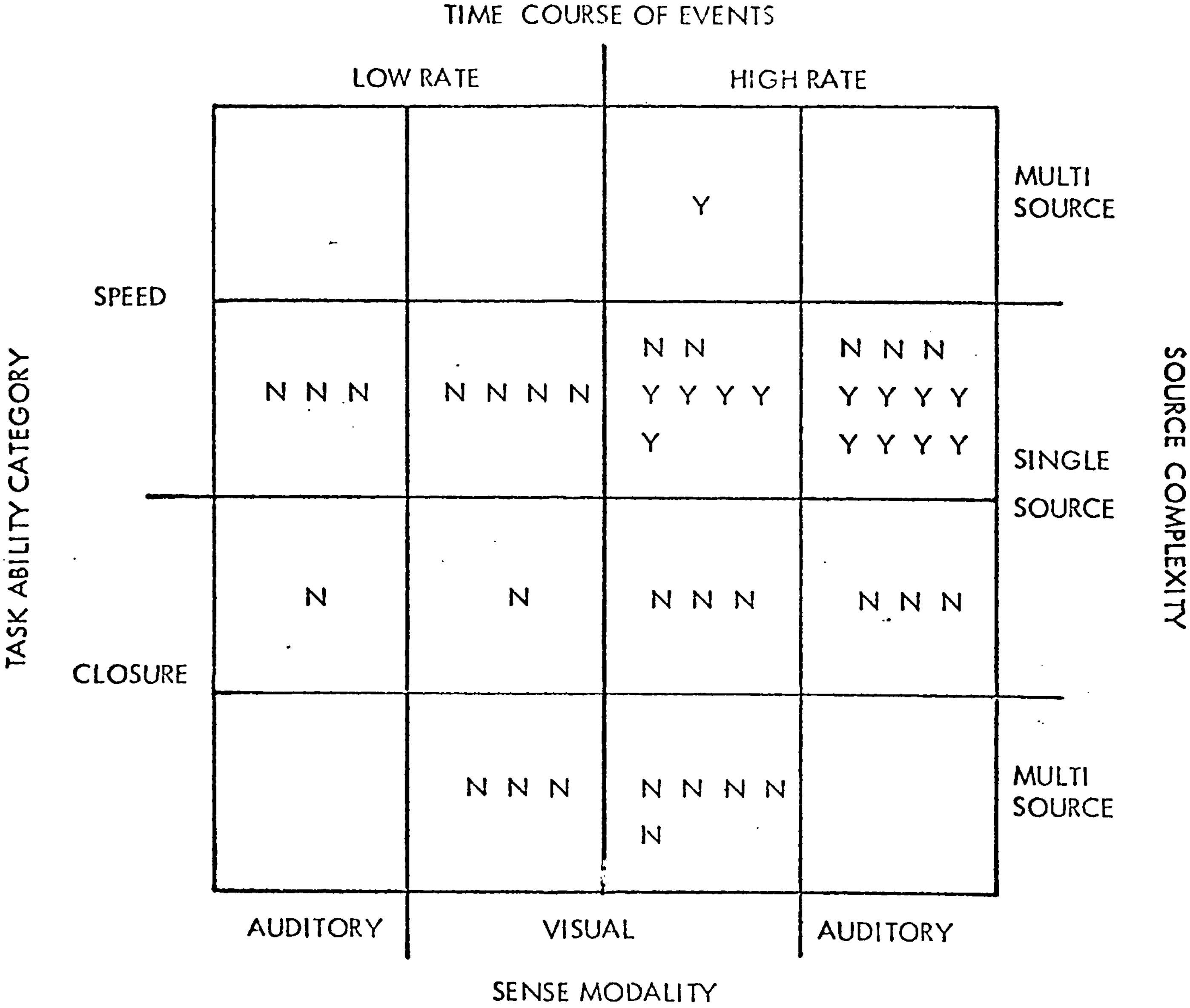
Another set of about thirty studies, conducted in the 1970's, is consistent with the earlier studies in presenting a reliable change in the decision criterion, but presents more experiments with a perceptual decline over time, including some with auditory displays (Parasuraman and Davies, 1976; Swets, 1976). However, the taxonomic analysis by Parasuraman and Davies supplies a new and current distinction between tasks that show a perceptual decrement and those that do not, encompassing all of the existing experimental results.

The main taxonomic dimension is now seen to be one that differentiates "perceptual speed" and "flexibility of closure." Described simply, tasks demanding perceptual speed are those requiring a rapid, successive comparison of stimuli or display states -- a change in a standard value, say, such that the standard value is no longer existing -- whereas tasks requiring flexibility of closure present a simultaneous comparison -- detection of a specified stimulus in a more complex surround. Examples of speed tasks are detection of an increase in the brightness of a flashing light, and a decrease in the duration of a repetitive tone. Closure tasks include detection of a disc of a different hue in a display of six discs, and a tone added to a noise background. The distinction between speed and closure tasks, and a distinction between high and low event rates (greater or less than 24 per minute), are sufficient to dichotomize present vigilance studies relative to a perceptual decrement.

The accompanying figure (Fig.1), devised by Parasuraman and Davies (1976), classifies 42 vigilance experiments according to these two dimensions, and also according to the dimensions of source complexity (single source vs. multiple sources of signals) and sense modality (visual vs. auditory). The Y's indicate experiments showing a perceptual decrement; the N's indicate experiments not showing such a decrement. We see that all instances of decrement arose from a successive comparison at a high rate.

Consistency of Individual Differences

Let us return later to an interpretation of this result, and consider now some collateral evidence for its significance: evidence from the consistency of individual differences in performance. Though earlier studies had not disclosed a basis for consistent individual differences across tasks, and, indeed, though a common view has been that individual differences are



TASK ABILITY CATEGORY	SOURCE COMPLEXITY	TIME COURSE OF EVENTS: LOW RATE		TIME COURSE OF EVENTS: HIGH RATE	
		AUDITORY	VISUAL	VISUAL	AUDITORY
SPEED	MULTI SOURCE			Y	
SPEED	SINGLE SOURCE	N N N	N N N N	N N Y Y Y Y Y	N N N Y Y Y Y Y Y Y Y
CLOSURE	SINGLE SOURCE	N	N	N N N	N N N
CLOSURE	MULTI SOURCE		N N N	N N N N N	

SENSE MODALITY

Fig. 1. Taxonomy of vigilance studies. Y's indicate studies showing a perceptual decrement; N's indicate studies showing no perceptual decrement. (From Parasuraman and Davies, 1976).

highly task specific, Parasuraman and Davies have neatly demonstrated that individual differences are task-type specific, when speed and closure types of display are differentiated. Specifically, they obtained correlation coefficients for various pairs of six tasks: Visual Speed 1 (VS1), Visual Speed 2 (VS2), Visual Closure 1 (VC1), Visual Closure 2 (VC2), Auditory Speed (AS), and Auditory Closure (AC).

Figure 2's upper panel shows correlations of d′, a measure of perceptual efficiency derived from SDT, and the lower panel shows correlations of the proportions of hits, or true detections -- for pairs of tasks laid out along the abscissa according to increasing compatibility. Classification across both the ability (speed-closure) and modality (visual-auditory) categories, denoted (VS/AC), leads to near-zero correlation; classification across the ability category and matching on the modality dimension (VS/VC) also produces a non-significant correlation. Significant correlations (above the line) are represented by the next two points, obtained from tasks matched on the ability dimension (VS/AS, VC/AC); the highest correlations are obtained from tasks matched on both dimensions (the rightmost two points). These results, together with other recent findings across sense modes (Tyler, Waag, and Halcomb, 1972) and across number of signal sources (Milosevic, 1974; Kennedy, 1976), suggest that we can devise selection tests for monitoring tasks, and extrapolate performance data from laboratory to practical tasks.

Situational Determinants of Response Strategy

Early vigilance research concentrated on time-dependent effects, regarded as decrements in perceptual efficiency, and so

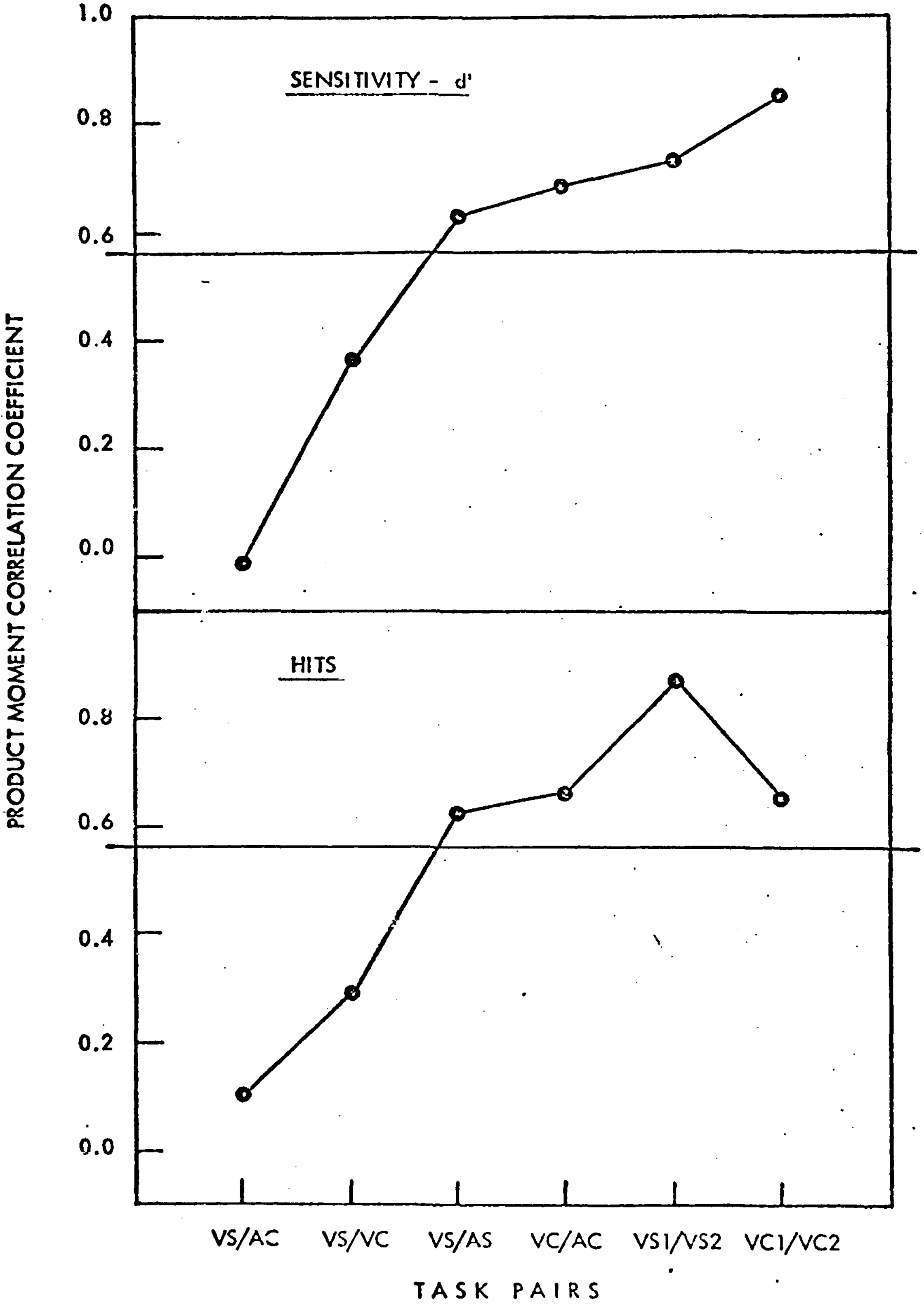


Fig. 2. Correlation coefficients for pairs of monitoring tasks, arranged according to increasing compatibility of the elements of a pair. See legend in text. (From Parasuraman and Davies, 1976).

failed to reflect the importance of situational changes in complex monitoring. Recent vigilance research -- acknowledging that time-dependent effects result from changes in response strategy, and often exclusively from such changes -- has emphasized situational variables that affect response strategy. Several studies have examined the effects of different signal probabilities, including abrupt changes in signal probability during a single experimental session, and a few studies have examined the effects of different stimulus-response payoff matrices.

The consistent result of experiments employing different conditions of signal probability is that the decision criterion adopted by the human observer, as indexed by the measure B (read "beta") derived from SDT, is almost perfectly correlated with the optimal decision criterion specified by the theory, though the criteria obtained generally shy away from the extreme optima (see Swets, 1976, and Swets and Kristofferson, 1970, for a review of a dozen such studies).

Williges (1973) and Embrey (1975) employed first a constant, then a changing, signal probability. In the first study, a signal probability of 0.1, with an optimal B of 10, led to obtained B's close to 10; and a signal probability of 0.5, with an optimal B of 1, produced empirical B's near 2. When the signal probability changed within a session from 0.1 to 0.5, the final B's obtained were also about 2. In the second study, an initial probability of 0.5 and a probability changing from 0.5 to 0.2, led to obtained B's very near the optimal values, namely, 1 and 4, respectively.

The point here, not to be lost among the numbers, is that vigilance research in the last decade, as influenced by SDT, has begun to bring into the laboratory the situational changes in probabilities that are characteristic of practical monitoring tasks. The same theory handles situational changes in decision values and costs. A half-dozen vigilance studies show that the decision criterion shifts in the expected direction as payoffs shift, when the payoffs are large enough to be significant for the observer (see reviews by Swets, 1976, and Swets and Kristofferson, 1970). Vickers (1976) has discussed some mechanisms of adaptive decision making in the context of SDT and vigilance.

The relevance of this line of research is suggested by discussion at the ARPA-sponsored workshop on perception and information processing (as summarized by Pew, 1976). It was observed there that the C^3 operator, particularly at the level of the commander, does not passively accept information but actively forms hypotheses that control his information-seeking and information-processing behavior, in accordance with probabilities and values and costs. An hypothesis may make the operator less attentive to a particular display as time goes on, so that his perceptual efficiency for that display appears to decline, or the hypothesis may serve to elevate his decision criterion relative to a signal from that display, or both. Stated at the same workshop was the view that the major failures of military command systems from their beginning (including, for example, Pearl Harbor) is their tendency to give insufficient weight to a low-probability event.

Do Situational Changes Affect Perceptual Efficiency?

The answer to that question is "probably not." That is to say, whereas situational changes clearly affect performance through alterations in selective attention, a matter to which we shall return, the motivating or energizing effects of situational emphases seem to have no beneficial effect on the perceptual efficiency of the attending observer.

Welford (1962), in fact, has developed the hypothesis that the observer's level of "activation" affects the decision criterion and not his sensitivity. Milosevic (1975) used skin resistance as a measure of activation in a vigilance setting, and found B (beta) to increase (indicating a progressively stricter criterion), along with an increase in skin resistance, while d′ (sensitivity) remained constant. Swets and Sewall (1963) and Watson and Clopton (1969) were unable to improve perceptual efficiency by motivational manipulation.

Absolute Levels of Performance

Several commentators (e.g., Howell, Johnston, and Goldstein, 1966) have pointed out that classical vigilance research has been unduly limited by its emphasis on time-dependent functions (decrements) rather than on absolute levels of performance. They write: "Many variables, such as those related to display complexity, may have more of a bearing upon over-all monitoring proficiency than they do upon maintenance of proficiency over time" (p. 130).

We observe here that SDT offers absolute measures of perceptual efficiency. And in the next breath that the standard measures supplied by the theory are based on error proportions, and so can be reliably determined only when the error proportions are large enough to be reliably measured in a realistic number of trials. However, Egan, Greenberg, and Schulman (1961) and Watson and Nichols (1976) have shown how measures of perceptual efficiency and of the decision criterion can be obtained from response latencies in a test with undefined trials -- thereby providing a methodology that permits working with somewhat stronger signals.

Industrial Inspection as a Research Paradigm

During the past six years SDT has been extensively applied in monitoring studies undertaken by industrial engineers in the interests of quality control. A good part of this work was described at a 1974 conference and collected in a volume edited by Drury and Fox (1975).

The industrial-inspection paradigm may be instructive for present purposes in part because there is little or no vigilance decrement in practical inspection problems. It is perhaps the freedom from concern for a decrement that has permitted the engineers to attend to other factors in monitoring, including physical and organizational variables.

The conceptual strengths of SDT have been treated explicitly and in detail in this context. Several authors have suggested that it offers a framework to integrate alertness, search, and memory with discrimination and decision in complex monitoring

(Adams, 1975; Buck, 1975; Drury and Fox, 1975). "Since the use of SDT permits the study of both stimulus and response variables, the entire inspection task may, at least in theory, be studied. The use of ROC curves should permit the comparison of viewing conditions, inspectors, equipment used, or visual target used to signal a defect" (Adams, 1975, p. 65). "The measures d′ and B ... are based only on error data so that the inspection of different products can be measured on the same scale, opening up new possibilities for absolute measures of job difficulty as well as their diagnostic value in improving performance" (Drury, 1975, p. 51). SDT′s "advantages are that it brings together the operational variables and allows their separate and interactive effects to be treated comprehensively in mathematical terms" (Drury and Fox, 1975, p. 97). Moreover, it "is most attractive as the vehicle for integrating human factors data with established quality control models .... Certainly in attempting to conceptualize the role of physical and organizational factors...it is invaluable and provides a rationale which makes the importance of these factors indisputable" (Drury and Fox, 1975, p. 98).

The economic or practical value of the theory seems also to be clearer in the industrial setting, perhaps because military values and costs are extreme. According to Wallack and Adams (1969), Drury (1973), and Drury and Fox (1975), the ability to predict various combinations of error probability along an ROC curve permits management to select the required "acceptable quality limit" with knowledge of the cost in relation to false detections, so that the need or not for a further check on the rejects is apparent. Sheehan and Drury (1971) illustrate the economic value of appraising inspectors of inappropriate criteria, of training and feedback to establish new criteria, and of feed-forward of fault probabilities. Strong support for the value of SDT to human-factors specialists in industry is the testimony of Chapman and Sinclair (1975): "The practical value [of SDT] arises from the fact that it allows economic justification for the application of ergonomics to inspection, and [from] the relative ease with which recommendations for improvement can be derived" (p. 241).

Memory, Time Pressure, and Effort

Consider now some implications of the finding that perceptual decrements occur when the display requires the ability of "perceptual speed" (i.e., calls for a successive comparison, or a change in some feature of a repetitive stimulus) and also requires almost continuous attention and decision making (i.e., presents the repetitive stimuli, or events, at a high rate). These implications, as spelled out by Parasuraman (1976b), point to a promising line of research on complex monitoring.

The first import of the successive comparison, in which a signal is specified only relative to another stimulus which is absent when the signal is present, is that memory is brought into play. In the terms of Norman and Bobrow (1975), monitoring tasks are "data limited," and the speed display suffers from limits on "memory data" as well as from the usual limits on the quality of "signal data." Performance will depend not only on the signal-to-noise ratio, but also on the quality of the memory references, i.e., of the stored representations of earlier inputs. Apparently, this demand on memory -- when coupled with a high stimulation rate, which might disturb the memory process -- brings about a perceptual decrement not observed with the "closure" (simultaneous-comparison) display.

Parasuraman has suggested that this focus on memory and time pressure may be related to Kahneman's (1973) ideas about memory, time pressure, and effort. In Kahneman's view, the effort invested in a task is determined largely by the intrinsic structure of the task; voluntary control over the amount of momentary effort exerted is quite limited. (In his example, one simply cannot work as hard retaining four digits for ten seconds as mentally multiplying two-digit numbers, even though the standard level of effort called for by either task does not give errorless performance.) He proposes that "Time pressure is a particularly important determinant of momentary effort," and that "Tasks that impose a heavy load on memory necessarily impose severe time pressure" (p. 27). We have a picture then of a high-rate successive-comparison display serving to mobilize a very high level of effort, so high that it cannot be maintained for long.

An enervating display may be said to have associated a high "cost of observing." Given the need to husband strength, to allocate effort over time, a large investment of effort in a given display may not be warranted by the payoff matrix of that display. A high cost of observing, in other words, may come to dominate the cost-benefit ratio, so that the display is not observed. Such may be the case even when there are no competing demands on attention, as usual in vigilance research. In the terms of one of psychology's conceptual schemes, used also in vigilance, the observer does not make an "orientation reaction" or "observing response" (Jerison, Pickett, and Stenson, 1965; Kahneman, 1973). Though more apparently applicable to vision, the concept of the observing response can apply also to audition.

Two recent experiments on vigilance that have required more of memory than has been typical should be mentioned. Johnston, Howell, and Williges (1969) presented visually an 8x8 matrix with each cell either blank or containing a series of two letters followed by a digit; the addition or deletion of one particular series constituted the signal, and additions or deletions of four other series served as noise. The hypothesis tested was that deletions demand more of memory than additions -- the relevancy of an added stimulus can be perceived, but the relevancy of a deleted stimulus must be remembered. Accordingly, perceptual efficiency (d′) was found to be lower for deletions than additions, and detection latency was greater. No perceptual effect over time was found, but the event rate was low (6 per minute). McCann (1969) required his subjects to check a written list of seven-digit numbers against a spoken presentation of the numbers; the signal was a one-digit discrepancy between a number presented in the two sensory modes. Here again, a low event rate was used (5 per minute), and there was no decline in perceptual efficiency.

Suggested Research: Cost of Observing

The cost of observing a particular display may be largely determined by the effort demanded by that display, but even a display that is effortless to observe may not be worth observing at certain times. In general, the cost of observing a display will also be affected by alternative demands on the operator′s attention. Observing one display may be sacrificed to observe another; and all observation may be suspended at times, as, for example, when the momentary premium is on commanding, controlling, or communicating, or on allocating, scheduling, or inventing. That is to say, the cost of observing a given display may be an "opportunity cost."

We wish to suggest, in broad outline, that one productive area of research on C^3 monitoring is one focusing on the monitor who has much else to do, and that the cost of observing can serve as a key concept in this research.

Some recent studies fall under this rubric, including a few in the literature on vigilance. The research program of Alluisi comes first to mind: ten studies on sustained multiple-task performance, representing about ten years of synthetic work by about 100 subjects, involving time-sharing among three monitoring tasks and three active tasks. The emphasis in these studies was on the effects of several independent variables on the two types of task, relative to baseline performance; the balance struck among the times allotted the various tasks was not analyzed. The general conclusion is that independent variables such as the time stress of continuous work, and sleep loss, have similar effects on monitoring tasks and active information-processing tasks, with respect to a decrement and recovery in performance (Alluisi, Coates, and Morgan, 1976).

More vigilance studies lately have used multiple sources of signals; here we will mention two presented at the recent NATO conference on vigilance. Colquhoun (1976) used one and four frequency channels in a sonar simulation, and found a decline in the proportion of true detections from 0.90 to 0.65 going from one to four channels. A decline over time was attributed to a criterion shift in both cases. In a second study, he told twenty-four observers to listen to a specified one of the four channels; a true-detection proportion of 0.75 when only the one channel was present declined to 0.45 when three noise channels were also present, implicating a masking process rather than a scanning process. The false-detection proportion, which fell over time, was higher over-all for four channels than one, but lower per channel. Craig (1976) used two visual targets, the two differing from a non-target stimulus on different attributes. No vigilance effects occurred when the two targets were presented as equally important; when one target was assigned priority, a decrement occurred for the more important target while none occurred for the less important target, the decrement attributable to a criterion shift. Craig's conclusion is that display complexity *per se* does not eradicate a vigilance effect.

We turn now to the engineering literature to identify quantitative models for selective attention to several simultaneous signal sources. Our brief review follows the lines of one prepared by Moray (1976) for the recent NATO conference on monitoring behavior and supervisory control; that conference also presented some current illustrations of the models. Having identified two "key developments" in vigilance research in the introduction to this paper, we should point out that the modelling work in the engineering literature constitutes a third development of the past decade that is key to research on C^3 monitoring.

Senders has considered how the observer's sampling of signal sources is affected by the information rate of the sources and the memory of the observer, and has elaborated two models of sampling behavior. According to the uncertainty model, the observer attends to various sources according to his uncertainty about their statuses (Senders, 1964); the probability (of signal) model applies when maximizing signal detection is desired

(Senders, Elkind, Grignetti, and Smallwood, 1965). Senders and Posner (1976) have further discussed queueing models for displays.

Carbonell (1966) considered how an observer's control of the process he monitored would serve to affect the information rate or predictability of a display's status. As Moray points out (1976, p.165), at this point we move from the study of attention to the study of skill, and extensions of this crossover are represented in Sheridan's (1970) paper on the concept of supervisor. Sheridan considers various response strategies derived from statistical decision theory as they might apply to selective sampling of sources and to distribution of attention between monitoring and control, with the cost of observing as a central concept. Observations of inputs have costs when alternative sources exist, and also when attention must be paid to outputs so that they may be skillful.

Moray (1976) reported psychological experiments based on these sampling models, involving various kinds of target, various target probabilities, selective attention and divided attention. One showed observers to have a good estimate of general and local statistical properties of signal sources, and to use those estimates in a near optimal manner to alter their decision criteria up to twice a second, as indexed by the B (beta) measure of SDT. Such estimates of probability, Moray suggests, may also be used in the observer's decision whether or not to make an observing response, that is, whether or not to attend. In the general picture that evolves from this line of thinking (Moray, 1976, p. 173), the observer uses probabilities and values in establishing response strategies that govern the decision he makes about the signal data he obtains from his observations, and also his decisions about when and where to observe, with the goal of predicting and adapting well enough to reduce reaction time and processing load.

Young's (1969) model of manual control also envisions such a process. We cannot begin to discuss the large literature on tracking here, but our virtually ignoring it does not deny that much of it is relevant to attention and C^3 monitoring. See, e.g., Levison, Elkind, and Ward (1971) on task interference. In a related study, Allen, Schwartz, and Jex (1976) consider decision theory as a model for car-driving behavior at a signaled intersection.

Another large literature that in the main is not discussed here is the psychological literature on "selective attention." That work is less relevant to C^3 monitoring than its name suggests, because it has emphasized structural questions at the expense of functional questions (Moray, 1976). Because various positions taken were difficult to undermine by experimentation, many studies throughout the 1960's could be directed to the question of whether selection was achieved by a discrete switching mechanism or a continuous attenuation (see Long, 1976), and to the question of whether the mechanism operates at the level of sensory analysis or response selection (Shulman and Fisher, 1972). As these authors point out, research on the rational allocation of attention, and on related functional questions, falls outside the scope of the current structural models.

However, there is increasing concern in psychological research on attention for problems of capacity allocation. Moray's work has been mentioned. Shulman and Fisher found that manipulation of signal probabilities and payoff matrices led their subjects to distribute their attention between two information sources in proportion to the expected value of attending to each. Taylor, Lindsay, and Forbes (1967) proposed that the additive properties of the square of one of SDT's measures of discrimination, d′, enable quantification of the processing capacity devoted to discrimination when two or more discriminations are attempted simultaneously; they concluded that approximately 85 percent of the available processor capacity is used for shared discriminations, and that 15 percent of the capacity is required to control the sharing procedure.

The time-sharing (or divided-attention) paradigm has been used to determine the information-processing load of one task by means of the decrement that task causes in the simultaneous performance of a second task. An example is the work of Wickens and Gopher (1975), who also manipulated the subject's attention-allocation policy by assigning variable task priorities, and, hence, variable costs of observing. They found that over-all error in an information-processing (tracking) task increased, and that reaction time in a ten-choice task increased, as the relative priority of each task decreased from 100 percent to 70 percent to 50 percent to 30 percent. Finer analyses of the tracking data showed that the introduction of the second task, and the initiation of sampling behavior, was accompanied by a lower effective signal-to-noise ratio (i.e., a perceptual decrement) and by a reduction in response gain (which the authors suggest is analogous to a response-bias, or decision-criterion, effect). The gain measure -- which corresponds to physical measures of response effort, and therefore relates closely to attention by the hypothesis stated earlier -- was further sensitive to differential task priorities.

In a second experiment, Wickens (1976) analyzed a tracking task time-shared with an input task (auditory signal detection) and an output task (application of a constant force). Only the parameters measuring internal noise and response bias (not one measuring processing delay) were sensitive to time-sharing conditions, and these only to the output task. The latter result suggests a more severe limitation of attention in output than in input processes, even though all six subjects felt the input (detection) task to be considerably more demanding than the output (force application) task.

An attention operating characteristic analogous to the ROC of SDT has been discussed by Karlsbeek and Sykes (1967), Kinchla (1969), and Sperling (1976). The notion here is simply that a given amount of attention can be distributed between two tasks in any given proportions, depending upon relative task priorities as reflected in signal probabilities and a payoff matrix. All three articles presented experimental verification of the hypothesis. Further discussion of the assumptions and interpretation of such operating characteristics, sometimes called "performance operating characteristics," is given by Kantowitz and Knight (1976) and Norman and Bobrow (1976).

The brief review in this section is meant to indicate that allocation of attention to monitoring tasks can be quantified, and related to independent variables in the over-all task structure, primarily the costs and benefits of observing. The last decade has produced general and quite adequate models for allocation policy, and a few experimental demonstrations that human monitors use such a policy. We think it evident that a programmatic experimental effort undertaken now, to obtain parameters for various functional relationships, would enhance significantly our understanding of monitoring in C^3 systems.

Suggested Research: Evolving Signal Patterns

The C^3 monitor, as we have said, must collect information from several sources, to infer signal events that consist of patterns of information, and he must integrate the information over time as the patterns develop dynamically. In this concluding section, suggesting another line of research, we attempt to show that recent developments in SDT provide an initial framework for a program of experimentation on such complex monitoring behavior.

Consider first signal detection over time, and then we will expand the discussion to include signal classification, or pattern recognition, over time. We will see that the two processes -- detection and recognition -- proceed in a correlated manner over time, and that models exist to predict recognition performance from detection performance. We shall, for the sake of brevity, go lightly over the quantitative particulars.

Two theoretical treatments of the observer accumulating signal information over time are of interest: one considering observation for a predetermined length of time (Birdsall and Roberts, 1965a), and the other considering a variable length of observation as influenced by the incoming information -- a procedure called "deferred decision" (Birdsall and Roberts, 1965b). In both cases, observation time is a function of the expected signal strength, the signal probability, the expected value of the decision (payoff matrix) and the incremental cost of observing as time proceeds. These same variables, of course, affect detection accuracy and response bias.

Preliminary applications of both models to human observers exist (Swets and Birdsall, 1967a). Application of an earlier model similar to that of deferred decision, involving sequential analysis, was reported by Swets and Green (1969); earlier human studies with several observations preceding a decision, with the number fixed by the experimenter, were reported by Swets, Shipley, McKey, and Green (1959). In general, the models predict human behavior: (1) the human observer accumulates sensory information at the rate specified in the theory, and (2) as concerns the balance of time and accuracy that determines the termination of observation, the human observer is sensitive to changes in the appropriate variables, and his performance is correlated with the optimal performance specified in the theory.

Interest in the signal recognition process, within the framework of detection theory, was evidenced early in the theoretical and experimental study of the two-alternative recognition problem (Tanner, 1956), and of the three-alternative (two signals and noise alone) problem (Swets and Birdsall, 1956). In another transitional step, the detection model for 1-of-M (multiple) orthogonal signals (Peterson, Birdsall, and Fox 1954; see also Nolte and Jaarsma, 1967) was successfully applied to speech perception (Green and Birdsall, 1964).

The SDT model for pattern recognition over time, incorporating the earlier work mentioned above, was described in Nolte's (1967) article on the adaptive optimum receiver. This model treats extended observation of, and deferred decision relative to, one of M possible signals. The observer is treated as storing updated probability estimates separately for each signal under consideration, so that detection and classification proceed simultaneously as parts of a single process, and the bases for both kinds of decision are available. The classification decision is determined by the largest probability; the detection decision is determined by comparing the largest probability to a decision criterion for a positive response, a procedure that gives a more reliable indication of signal presence than would be available if the observations of the various signals were combined in some way.

Some results of a preliminary, unpublished application of this model, relating only to detection performance for a fixed number of observation intervals, are shown in Fig. 3, in which the SDT measure of accuracy or perceptual efficiency d′ is plotted against the number of successive observation intervals, for the average of three observers (Swets and Birdsall, 1967b). The square data points at the top of the figure represent a baseline condition: the average values of d′ for eight different signals (in this preliminary case, eight frequencies of a brief tone) when only one signal (or noise alone) was presented throughout the (six) successive intervals of a trial and throughout a given series of trials, and the signal to be presented was known to the observer. The theoretical line fitted to the first point, called "constant 1-of-1," represents an increase in d′ proportional to the square root of the number of observation intervals, as predicted for this condition.

The data points of major interest are the triangles, representing d′ when any one of the eight signals (or noise alone) could occur on a given trial of a series, but with the

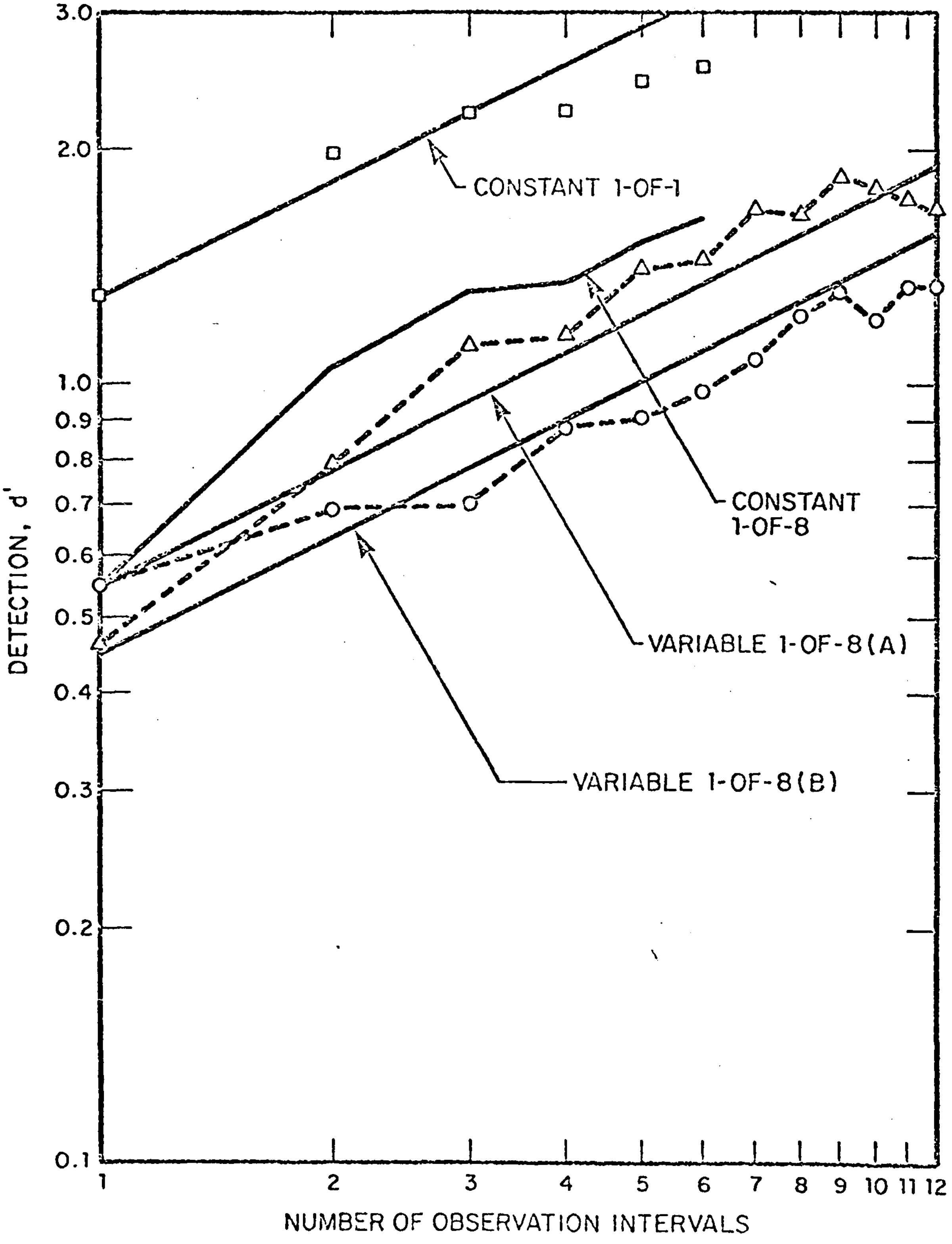


Fig. 3. Detection performance, d', for eight signals, as a function of the number of observation intervals; average of three observers. The three sets of data points and the four predictions from theory (solid lines) are described in the text.

same signal (or noise alone) occurring throughout the (twelve) observation intervals of a given trial. The theoretical line labelled "constant 1-of-8" is predicted from the square data points, in accordance with the model for 1-of-M orthogonal signals and the adaptive-optimum-receiver model. The prediction may be seen to exceed human performance, but generally to follow its contour.

The circular data points represent another baseline condition, one in which any one of the eight signals could occur in any interval of a given trial as well as in any trial; the observer, then, can only combine the observations at each frequency band at each observation interval. The two curves labelled "variable 1-of-8" (A and B) represent SDT's optimal predictions based on two different methods for combining observations over signal channels. Both curves exceed human performance, but one comes quite close.

Another condition of this experiment used only two signals but collected classification data as well; the proportion of correct recognitions varied from near chance at the first observation interval to near perfect at the twelfth interval, as detection performance varied through approximately the same range.

The last piece of theoretical and experimental work to be mentioned shows another way of predicting classification from detection, developed in the context of radiology (Starr, Metz, Lusted, and Goodenough, 1975). The notion here is to subtract a quantity from the ordinate (proportion of true detections) of the empirical ROC curve for detection of 1-of-M orthogonal signals, the exact quantity depending on M, in order to obtain a predicted

ROC curve for classification, i.e., where the ordinate represents the proportion of signals correctly detected and correctly classified. How well this model works, when the classification is a matter of visual localization among quadrants based on a single observation interval per decision, is shown in Fig. 4. Application to Lindner's (1968) data on recognition of two frequencies of tones predicted the results for four observers to within one or two percent (Green, 1976). This model remains to be applied to the recognition of complex patterns developing over time.

The brief review in this section is meant to indicate that a good deal of theory, and some preliminary experiments, exist relative to evolving signal patterns, and that a programmatic effort extending both theory and experimentation, to enable detailed predictions for classification of complex patterns, would be an appropriate way to extend our knowledge about monitoring in command/control/communication systems.

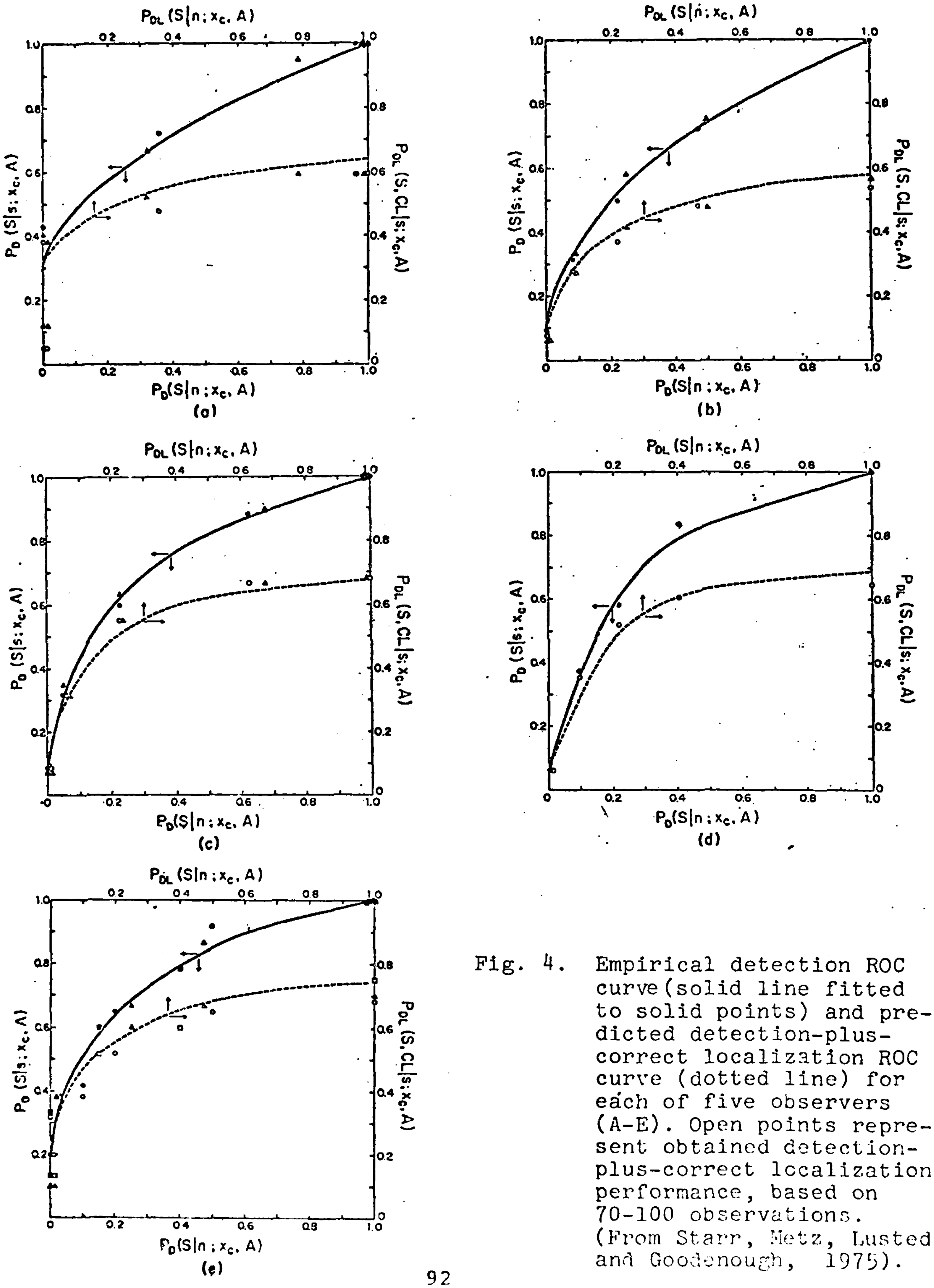


Fig. 4. Empirical detection ROC curve (solid line fitted to solid points) and predicted detection-plus-correct localization ROC curve (dotted line) for each of five observers (A-E). Open points represent obtained detection-plus-correct localization performance, based on 70-100 observations. (From Starr, Metz, Lusted, and Goodenough, 1975).

CHAPTER 3

MEMORY AND INFORMATION STORAGE

The ability to store and retrieve information is basic to any organism or system that must perform intellective functions. In the absence of human memory there could be no such things as perception and thought. Perhaps for this reason memory has been more intensively studied by psychologists than any other aspect of human experience or performance. In spite of the attention it has received, memory remains an enigma. Some determinants of performance of certain types of memory-dependent tasks have been identified, and some functional relationships between stimulus and performance variables have been determined, but exactly how memory works is not known. Even how many memory systems there are, remains an unresolved question.

In this chapter we review work on memory that has been done, largely during the relatively recent past. Most of the work reviewed is basic research, because most of the work that has been done on memory has been motivated by theoretical, as opposed to applied, interests. That is not to suggest that the results of the work have no applied implications; however these implications are probably less direct and obvious in this area than in several of the others covered in this report.

Short-term versus Long-term Memory

The notion that memory is a two-component system, composed of a conscious short-term store (STS) and an unconscious long-term store (LTS) dates back at least to the nineteenth century. Although such concepts were largely forgotten during the realm of the behaviorists, two process memory theories were revitalized by Hebb in 1949. Since then such theories have enjoyed extremely widespread acceptance by psychologists interested in human information processing.

The essence of these theories is that after sensory information is physiologically encoded, it may briefly reside in the brain in that state. However, if the information is to be retained longer than a very few seconds, it must be read into the STS; if it is to be remembered for a relatively long time, it must be subsequently transferred to the LTS. Whereas the LTS is a permanent memory with a practically unlimited capacity, the capacity of the STS is strictly limited in both the number of items it can hold and the temporal persistence of those items. Unless the contents of the STS are actively maintained through rehearsal, they are lost in a matter of seconds. However, as long as the information resides in the STS, it may be transferred to the LTS. In this way, the STS processes are responsible for long-term memories.

Currently Atkinson and Shiffrin's (1968, 1971) version of the two-process memory theory is probably the most popular, but it is far from unique. Almost every theorist who contributed to Norman's book, Models of Human Memory (1970), subscribed to the basic view that stimulus information flows from a sensory register to a STS, where it is recoded and temporarily retained, and then, given sufficient attention, it is transferred to a LTS.

There is considerable empirical support for the division of memory into short- and long-term stores. Neurological evidence for two-process memory is primarily derived from the memorial deficits of patients with bilateral hippocampal lesions; these people exhibit an inability to form new long-term traces despite the fact that they suffer no impairment in immediate memory (Milner & Teuber, 1968).

Apart from the neurological evidence there are considerable behavioral data demonstrating process differences between the short- and long-term stores. First, whereas information in the STS seems to be exclusively encoded along acoustic or temporal dimensions, coding in the LTS involves semantic, frequency, and redundancy factors, but not acoustic factors (Baddeley, 1970b). Second, while the amount of material retained in the LTS is sensitive to the rate at which the material was presented, the amount retained in the STS is not (Glanzer & Cunitz, 1966). By contrast, delaying recall or manipulating the order of report affects short-term memory, but not long-term memory (Murdock, 1967); these data are consistent with the view that information in the STS is rapidly lost if not actively maintained or transferred to the LTS. Finally, whereas information that is in the STS can be rapidly and accurately accessed through a serial exhaustive search of its contents (Sternberg, 1969), information retrieval from the LTS involves an often laborious and presumably nonexhaustive search and is a function of the sought item's frequency of previous occurrence (Underwood & Shulz, 1960).

While the distinction between short-term and long-term memory is a widely accepted one, and one for which supportive evidence can be cited, its validity has not been demonstrated unequivocally. Moreover, even assuming the validity of the distinction, the boundary between these two components is fuzzy at best; it is not the case that something is either remembered for a very short time, or for a very long time, if at all. It seems not unlikely to us that a one-component theory of memory might be developed that would account for what is known about memory-dependent performance as well as, or better than, any existing two-component theory. However, theory development is beyond the scope of this report, and acceptance of the distinction between short-term and long-term memory facilitates the review of previous research inasmuch as most of the theorizing on which that research has been based has taken this distinction as a given.

Auditory Encoding in Short-Term Memory

The use of auditory coding in the STS is both intuitively obvious (e.g., verbal rehearsal) and empirically well substantiated. For example when interitem similarity is high, verbal learning is retarded because of interference (Kintsch, 1970a). If items in the STS are encoded along acoustic dimensions, then remembering acoustically similar lists should be more difficult than remembering acoustically dissimilar lists. In fact, this is found to be true. Conrad and Hull (1964) found that the memory span for acoustically confusable items was differentially small. Similarly, Baddeley (1966, 1970a) found that acoustic similarity had an adverse effect on immediate memory for both verbal lists and paired-associates. Wickelgren (1965a) demonstrated the same phenomenon through a retroactive interference paradigm. In his experiment, subjects first listened to a list of four memory letters, and then to an interference list of eight more letters. The greater the acoustic similarity between the letters in the interference and memory lists, the greater was the retroactive interference.

A close analysis of auditory confusions in a STS shows that they may be predicted from the linguistic structures of the stimuli. More specifically, Wickelgren (1965b, 1966) determined that vowels and consonants are not encoded as units in the STS, but, rather, are encoded in terms of their distinctive features. His subjects copied memory sets of six aurally-presented items which varied only in their vowel (1966) or only in their consonantal components. After the sequence had been presented, they covered their transcriptions and began ordered recall of the memory set. Copying during stimulus presentation was required so that memory errors could be separated from perceptual errors. As it turned out, errors in neither the perception (Miller & Nicely, 1955; Wickelgren, 1965b, 1966) nor the memory (Wickelgren, 1965b, 1966; Murdock, 1967) for letters could be explained by the all-or-none confusion or forgetting of the individual sounds. Rather, errors in both are the product of confusing or forgetting the appropriate values of either voicing, nasality, or place (manner) of articulation, and for a given letter, retention or loss of each of its constituent features is independent of the fate of the others.

The first evidence that people often recode visually presented stimuli into an auditory form if they are expected to remember those stimuli even for relatively short periods of time came from the analysis of substitution errors made by people in verbal-recall tasks (Conrad, 1964; Wickelgren, 1965a, 1966). The general finding was that the errors were more readily accounted for in terms of acoustic similarities between the erroneously reported items and the correct items that were missing than in terms of visual similarities between these items.

There are many other examples of acoustical confusions in short-term memory for visually presented items. In an experiment by Laughery and Pinkus (1966) subjects were presented either a visual or auditory series of six or eight letters at one of three presentation rates (20, 60, or 180 items per minute). The letters on a given trial were either acoustically similar or dissimilar. Recall performance was depressed with acoustically similar letters. Further, while auditory presentation resulted in better performance, the difference was accentuated at shorter presentation intervals. It was adduced that visual items had to be translated to an auditory form before they could be maintained.

In a similar experiment, Laughery and Harris (1970) sequentially presented eight letters to subjects either auditorily or visually. After each sequence, the subjects turned over a card with eight letters on it and compared it with their memory of the previous sequence. On some of the cards, the experimenter had substituted certain letters of the sequence with either visually or acoustically similar letters. The subjects' task was to correct the card such that it was a perfect representation of the sequence. Substitutions which were acoustically similar to the correct letter were often undetected, but visually similar substitutions had no such effect. Similarly, Baddeley (1966) and Cimbalo and Laughery (1967) found that short-term memory was depressed for acoustically similar items but not visually similar items, regardless of presentation modality.

Such findings demonstrate quite conclusively that subjects often encode even visual stimuli auditorily when they know they are going to be called upon to retrieve them at some later time.

Why such a visual-to-auditory recoding occurs is not fully understood, although there are some plausible conjectures. If the subject has to make a verbal response in a recall task, then a recoding is necessary at some point in order to articulate the visual stimulus with the verbal response. If this were the primary reason for the recoding, the recoding could take place just before the emission of the response, and the stimulus could be held in memory during the retention interval in a visual form. Alternatively, it could be recoded immediately after reception and the output code retained during the retention interval. The assumption that recoding is done simply because it is a necessary step before the response can be evoked does not favor either of these possibilities over the other.

Another plausible explanation of why the recoding is done is to facilitate rehearsal of the material during the retention interval. This interpretation gains credence from the fact that, as Dainoff (1970) has pointed out, experiments that have provided evidence of auditory recoding have typically involved memory loads that approach the immediate memory span (say 5 to 7 items). Evidence for such recoding has not been so readily obtained in studies in which the memory load has been small (e.g., a single item). Why it should be easier, or more effective, to rehearse material that is represented in memory in an auditory form than to rehearse material that is represented visually has not been explained. In particular, if rehearsal amounts to "listening" to memory images of the names of items, why should it be easier to listen to acoustic images than to "look" at visual ones? It seens to be generally assumed that rehearsal is easier with auditory stimuli, although several investigators have presented evidence that strongly suggests that people can generate visual as well as auditory images.

One possible explanation of why the visual-to-auditory recoding occurs is that the information cannot be retained in visual form--that it must be translated into an auditory code if it is to be stored at all. The common finding that material is retained better after aural than after visual presentation lends credence to the idea that auditory short-term memory is more robust in some sense than is visual short-term memory. The generality of this difference is supported by the fact that it has been found with tasks ranging from free recall (Murdock & Walker, 1969), to paired-associate learning (Baddeley, 1970), to serial learning (Craik, 1969), to probed recall (Murdock, 1968), and finally to bisensory memory tasks (Dornbush, 1968, 1970, 1971). Moreover, the difference seems to be located in the STS: the effect is typically most pronounced in the recency portion of the serial position curve, or, in the bisensory memory task, when auditory items are reported first.

The utilization of auditory codes to facilitate retention of visual information is also demonstrated by Glanzer and Clark's finding that people's abilities to reconstruct an array of black and white geometric forms was highly related to the number of words which they used to describe the figures (Glanzer, 1961; Glanzer & Clark, 1963). Similarly, Mackworth (1963) found that the speed with which subjects could name elements of a given stimulus set (either digits, letters, colors, or shapes) was predictive of the number of elements which they could recall from that set after its brief presentation. These experiments are consistent with Glanzer's (1961) verbal-loop hypothesis: the subject processes perceptual information through covert verbalization and the difficulty of the verbalization determines the difficulty of retaining the stimuli in the recall task.

Sperling (1963, 1967) has also proposed that at least for brief exposures of visual stimuli, memory consists of an auditory information store and a rehearsal loop. His argument is that while only 10 milliseconds of visual exposure is necessary to recognize a letter, report of the letter is usually delayed for several seconds and can be delayed indefinitely. However, since the visual icon fades within a fraction of a second, the subject's ultimate report cannot be the translation of a visual memory. Further, despite the rapidity of visual scanning, the number of reportable items within a large range of stimulus exposure durations, is exactly limited to the capacity of the auditory information store. These facts, coupled with the role of auditory similarity in accounting for confusions, have been interpreted as favoring the view that auditory storage, with its unique capacity to reproduce and thus restore its images through rehearsal is the only feasible means of prolonged retention.

Evidence Regarding the Retention of Visual Information in Short-Term Memory

While the evidence that visually presented information is often recoded into auditory form for short-term retention is very strong and the conclusion that such recoding occurs has hardly been challanged, the idea that such recoding must occur because the information cannot be retained in visual form is a much less widely accepted one. A variety of experimental results supports the general notion that visual information can indeed be retained in visual form for durations as long as those usually associated with auditory STS.

Effects of Visual and Auditory Similarity

One of the basic approaches that has been used to study short-term memory codes has been to observe the effects of systematic variations in the auditory and visual similarity of the memory items. Several experiments have varied the degree to which the items that must be compared resemble each other, either in terms of visual features or in terms of the sounds of their names. The rationale for this type of manipulation is the assumption that varying similarity with respect to one modality should affect performance only if the items are compared with respect to that modality.

This approach is illustrated by an experiment by Posner and Chase (Posner, 1967) in which they attempted to analyze the temporal course of acoustic encoding by comparing the kinds of confusions generated in visual and memory search conditions. In the visual search condition, an array of one to four comparison letters and a target letter were presented simultaneously. In the memory condition, the comparison array was presented for ten seconds and succeeded by the target letter. The subject's task in both conditions was to respond as quickly as possible as to whether the target item was a member of the comparison array. Three sets of letters were used: neutral, visually confusing, and acoustically confusing. The expectation, following Sperling (1963), was that the search rate in the visual task would be solely a function of the visual confusability of the arrays, whereas the search rate in the memory search task would reflect only acoustic confusions. Contrary to this expectation, the search rate in both the visual and memory conditions was uniquely affected by visual confusability; acoustically similar letters were searched as rapidly as were neutral letters. Shiffrin and Gardner (1972) have replicated this effect in a forced-choice detection paradigm.

An experiment by Wood (1974) also illustrates this approach. Wood covaried the auditory and visual similarity of pairs of letters in a same-different task. For one set of stimuli, the items of a "different" pair were auditorily similar but visually distinct (e.g., C, Z); for the other set, the items of a different pair were visually similar and auditorily distinct (e.g., O, C). The items of a pair were presented serially with about 1 second between onset times (the duration of each stimulus was about 360 msec.), and the subject's task was to say whether the second letter was the same as the first. Reaction times were longer for the auditorily similar pairs when the letters were presented auditorily and for the visually similar pairs when they were presented visually, suggesting that the stimuli were coded and matched in the modality in which they were presented.

Still another experiment that illustrates the strategy of varying auditory and visual confusability is one by Dainoff (1970), who used a Posner-type letter-name matching task to explore the time course of auditory and visual memory representations. As in the paradigm orginally used by Posner and Mitchell (1967), some of the "same" pairs were in the same case and therefore visually identical, while others were in different cases and therefore visually dissimilar. For half of the subjects the letters that were used were slected from a set whose names were auditorily confusable (b, p, d, t); for the other half the letters had names that were auditorily distinct (f, i, m, q). Stimulus onset asynchrony (SOA), the difference between the times of onsets of the two letters on a trial, was varied between 0 and 2 sec.

"Same" RT for visually identical letters was faster than "same" RT for visually dissimilar letters that had the same name and faster than "different" RT, the difference tending to decrease with increasing SOA. "Different" RTs were longer for the letter pairs that were chosen from the auditorily confusable set than for those chosen from the auditorily distinctive set, and this difference tended to increase with SOA. "Same" RTs tended to be shorter for letter pairs chosen from the auditorily confusable set than for those chosen from the auditorily distinctive set, and this was relatively independent of whether the letters corresponded visually or only in name, and of SOA.

Dainoff interpreted these results as favoring the view that whether the same-different decision is based on a comparison of visual or auditory memory codes, depends on how long the first item must be maintained in memory. In particular, the probability that a visual code will be used decreases, while the probability that an auditory code will be used increases, over time.

Effects of Stimulus Degradation on Encoding versus Comparison Operations

The results of Sternberg's (1969) first studies of memory scanning suggested that when subjects are asked to determine whether or not a given item is a member of a memorized subspan list, they execute a serial exhaustive search for a match in the memory set. In keeping with an acoustical encoding hypotheses, subjects did report silent rehearsal of the memory items even though both the memory and test items were visually presented. A serial search through the rehearsal set for a name match seemed a feasible strategy. However, whereas subvocal rehearsal proceeds at a maximum rate of ten items per second (Sperling, 1963; Landauer, 1962), the average search rate was about thirty items per second. In order to verify the acoustic nature of the representations of the memory set, Sternberg presented a visually degraded test stimulus. If the memory search involved name matching, then while the degraded test item should be more difficult to identify, and thus produce increase in the intercept of the search function, there should be no change in its slope: once the name of the test item had been ascertained, the comparison process should proceed normally. By contrast, if the stimuli were not verbally encoded, but comparisons were between visual images, the degradation of the test item should directly affect the slope of the comparison process but have no effect on the intercept. The results showed that at least with relatively unpracticed subjects, degraded test stimuli produced increases in both the slope and the intercept of the reaction time function. Since the increased slope presumably reflects comparisons of the physical attributes of the test stimulus, the increased intercept must indicate that the encoded test stimulus may not constitute an exact image of the test item, but may be the product of a clean-up process. If the test stimulus for the comparison process is visually represented, so too must be the memory items to make comparisons possible. The implication is that even when an array of letters must be retained across many trials, it may still be visually encoded.

Control of Sufficiency of Auditory or Visual Code

Perhaps the most straightforward way to study visual or auditory memory codes is to use stimulus materials that the subject *must* code in visual form or in auditory form, as the case may be.

Unfortunately, it is not clear that there are materials for which a recoding from one form to the other would be patently impossible. One can, however, design experimental situations in which the amount of recoding that is necessary in order to compare two stimuli is controlled to some degree. The mixed-case, letter-matching procedure developed by Posner and his associates (Posner, 1969, Posner & Keele, 1967; Posner & Mitchell, 1967; Posner, Boise, Eichelman, & Taylor, 1969) is an example of such an approach. In this procedure, the subject is asked to decide whether two letters (which may be presented simultaneously or sequentially) have the same name. In some instances the letters are presented in the same case; in other instances, one is in upper case and the other in lower. The assumption is that, if the decision is based on a comparison of names, performance should be insensitive as to whether the letters that are being compared are in the same case. Thus, an effect of case on matching is taken as evidence of matching of visual characteristics. In other words, a visual match is sufficient when the characters are in the same case; however, it is insufficient when they are in different cases. The basic finding, with this paradigm, has been that the time required to decide that two simultaneously presented letters have the same name is less if the letters are visually identical (AA) than if they are visually dissimilar (Aa). If the letters are presented sequentially, the advantage of visual identity diminishes as the interval between the letters is increased, going to nearly zero within two or three seconds. The paradigm and variations of it have been widely used and the results by and large are consistent with the notion that the decision is made on the basis of a visual match, if a visual match is sufficient and if the retention interval is very short (less than 1 or 2 sec.) and that the names are matched otherwise. To account for the fact that the match of names takes longer than the match of visual characteristics, it is generally assumed that the match of names requires some processing activity (identification of items, retrieval of names from memory) that a visual comparison does not.

Phillips and Baddeley (1971) have criticized the logic by which the rate of decay of visual short-term memory has been inferred from results obtained with Posner and Keele's (1967) physical-versus-name matching paradigm. What they object to is the equating of the interval over which physical identity facilitates same-name decisions with the duration of the visual memory trace. They point out that if the name code is developing as the visual trace is fading, the time at which the former becomes the preferred basis for decision is not necessarily the same as the time at which the latter is no longer accessible.

In an effort to get an indication of decay rate for the visual trace that is not obscured by the involvement of a name code, Phillips and Baddeley did a same-different study with stimuli for which, presumably, a name code would not be developed. Stimuli were 5 by 5 matrices of randomly filled and unfilled squares. On each trial, the second matrix that was presented either was the same as the first, or differed from it with respect to a single square. The first stimulus was exposed for 0.5 sec., and the retention interval was varied from 0.3 to 9.0 sec. Percent correct decreased and RT increased, regularly (both nonlinearly) over the entire range of retention intervals. Phillips and Baddeley concluded that some visual information is retained for several seconds.

A different way of controlling the sufficiency of a visual representation was used by Nickerson (1975, 1976). Subjects are shown two "noisy" visual characters on each trial and are asked to decide whether the second character is the same character as the first. Two kinds of noise are used: One type is referred to as correlated, and the other as uncorrelated. In the correlated noise case, the same noise pattern is imposed on both characters that must be compared on a given trial. In the uncorrelated noise case, different noise patterns are imposed on the two characters that must be compared. When the noise is correlated, the same-different decision can be made on the basis of a test for congruence, because the letters are the same if and only if the patterns, noise and all, are visually identical. With uncorrelated noise this is not the case; no two patterns are ever the same; therefore, a test for congruence will always fail. If a visual representation of the first pattern is not retained and the only thing that is stored in the subject's memory at the time that the test pattern is presented is the name of a character, then the time required to decide whether the second letter is the same as the first should be independent of whether the noise patterns that are used with the two letters are identical or not. In other words, if the stimuli are compared by matching their names, performance should be independent of whether or not the noise is correlated. Therefore, differential performance under the two noise conditions could be taken as evidence that visual information is involved in the process by which the same-different decision is made.

In two experiments in which this task was used, performance was better (faster and more accurate) with correlated than with uncorrelated noise patterns. Increasing the amount of noise degraded performance under both the uncorrelated and correlated noise conditions, but more in the former case. Finally, the difference between the effects of correlated and uncorrelated noise decreased as the duration of the interval between the two patterns increased from 0 to 8 sec.; however, there was a difference even with the longer intervals. These results were taken as evidence that visual representations of the stimuli can be retained for several seconds at least, and may be used as a basis for same-different decisions even when the stimuli are nameable, provided the situation is such that congruence testing is an effective way to proceed.

Interference with Auditory Recoding and Subvocal Rehearsal

One of the methods that has been used to investigate the question of whether visually presented information can be retained in the absence of auditory recoding and rehearsal is that of attempting to interfere in some way with the rehearsal process that is assumed to be necessary to keep the auditory trace intact. An experiment by Parks, Kroll, Salzberg, and Parkinson (1972) illustrates this strategy. Subjects were shown two letters, each visually and briefly, separated by 8 sec. The task was to indicate as quickly as possible whether the second letter had the same name as the first. Both upper-case and lower-case letters were used; on some trials, the cases of the two letters being compared were the same and on other trials they were different. During the retention interval, the subject was required to "shadow" letters spoken at the rate of 2 per sec. If one is willing to assume that rehearsal is essential to the maintenance of an auditory trace, and that the shadowing task makes rehearsal impossible, then, in order to perform this task, the subject would have to retain visual representations of the memory items. As it happened, subjects did perform this task quite well. Moreover, correct "same" response times were faster and error rates were lower when the two visual stimuli were in the same case than when their cases differed.

This finding is consistent with the following model. A visual representation of a memory item is retained during the retention interval. When the test item occurs, a test is made between its visual characteristics and those of the memory item. If the representations match, a positive response is evoked; if they do not match, a further test is required inasmuch as the fact that the patterns differ visually does not establish that they have different names. Presumably, this further testing involves identification (naming) of both patterns and a comparison of their names. Thus, positive RTs associated with the same letters in different cases should be longer than those associated with the same letters of the same case. In the absence of any reason to assume that the time required to compare names will depend on the outcome of the comparison (same or different), we would expect <u>all</u> of the "different" RTs to be similar to the "same" RTs that are obtained with characters whose cases differ, but this is another issue.

In the aggregate, the results obtained by Parks et al. (1972) are consistent with this general view of the process. In one particular, however, they suggest that the situation is more complicated than this. Although tests of significance were not performed relative to this point, it appears that of considerably more importance than whether the characters being compared agreed in case was the case of the test item. In particular, RT was shorter when the test item was an upper-case character than when it was a lower-case character. Indeed, the memory item-test item combination lower-upper produced a shorter RT than the combination lower-lower. This result suggests that at least some of the subjects retained an upper-case representation of the memory item irrespective of the case in which it was presented.

Several other experiments by Kroll, Parks, Parkinson, and their other associates (Kroll, Parks, Parkinson, Bieber, & Johnson, 1970; Parkinson, Parks, & Kroll, 1971; Kroll, Parkinson, & Parks, 1972) have also shown that if the subject is engaged in a shadowing task during the retention interval, auditory memory declines rapidly whereas memory for visual or audio-visual stimuli may show no loss even over 12 seconds. Presumably, the shadowing task generated auditory retroactive interference and prevented verbal rehearsal, so it selectively impaired auditory memory.

Further evidence of the relative insensitivity of visual memory to auditory interference comes from an experiment by Scarborough (1972) who presented stimulus trigrams either aurally, or visually, or in both modes simultaneously. The interpolated task was counting backwards by threes from an aurally presented 3-digit number. As he had predicted, aurally presented trigrams were recalled the worst. Further, visual and visual-aural trigrams were recalled equally well, supporting his contention that auditory and visual retention can be mediated by different memories which are largely independent of each other.

Control of Auditory Codability of Visual Stimuli

One would expect that auditory recoding of visual stimuli would be most likely to occur for visual patterns that have highly overlearned, or readily accessible, names. Letters and numbers are examples of such patterns, and it is such stimuli that have been used in most of the experiments that have provided evidence that recoding occurs. But, what about patterns that do not have readily accessible names? How are such patterns retained? In particular, how does their fate in memory differ from that of patterns which do have readily accessible names?

A problem that one encounters in trying to do research on this issue is the fact that it is not easy to find visual stimuli that do not have, or will not quickly acquire, names. One can use nonsense figures; however, one cannot permit a subject to see any given figure very often, because he will invent a name for it, and what was a nonsense figure on its first occurrence may become a familiar pattern with a name very quickly. In an effort to cope with this problem, Nickerson (1972) used computer-generated patterns that changed on every trial. The subject's task in this experiment was to decide whether the second of two sequentially presented visual stimuli was the same as the first. The visual patterns were formed from subsets of the dots from a 7 x 5 rectangular dot matrix. Three types of stimulus pairs were used: letters, random dot patterns what were easy to differentiate, and random dot patterns that were difficult to differentiate. The time between the two stimuli on a trial was varied from 0 to 8 sec. It was assumed that the letters, but not the random patterns, were readily nameable. Reaction times were much longer and error rates much higher for the difficult random patterns than for the letters. Performance with the easily discriminated random patterns, however, resembled that with the letters more than it did that with the difficult random patterns. On the whole, the results were consistent with the idea that the main difference among the three sets of stimuli was their visual discriminability. There was little evidence that memory representations for the letters and for the random dot patterns decayed at different rates.

Manipulation of Time Required for Visual to Auditory Recoding

Another approach that has been taken involves manipulating not auditory codability per se, but the time required to effect the encoding. Presumably, the time required to encode a visual letter auditorily includes the time required to identify the visual pattern and to retrieve its name from long-term memory. Any manipulation that would complicate the recoding process should differentially affect performance, depending on whether or not the recoding is done.

Kellicutt, Parks, Kroll, and Salzberg (1973) conducted an experiment in which letters were used as stimuli, but in some cases they were presented in an unfamiliar orientation for the purpose of impeding their identification and hence increasing the time of the recoding process. In this experiment subjects were asked to compare two <u>pairs</u> of letters with respect to their names on each trial. One member of each pair was in upper case and the other in lower case; matching letters sometimes corresponded in case and sometimes they did not. On some trials all letters were in their normal orientation, and in others they all were in a backward (left-right reflected) orientation. The retention interval was 8 sec., during which subjects shadowed spoken letters. Data were reported only for "same" trials, that is trials on which the two items of the test pair had the same names as those of the memory pair.

Performance was better when corresponding items of the two pairs were in the same case than when they were in different cases, and it was also better with stimuli in normal than in backward orientation. The investigators had hypothesized that orientation would have a detrimental effect when the characters were in different cases, but not when they were in the same case. The idea was that, given the same case, the pairs would be matched without (or before) being identified, so a backward orientation (which, presumably, would impede identification but not visual matching) would have no effect. In fact, it did have a very sizeable effect on the RTs of the subjects in the main study (30 subjects, 196 trials each). Two subjects were given additional practice, however, and during the last 160 of 640 trials the effect of orientation was present only when the cases of the corresponding items of the pairs differed; it was not there when the cases were the same.

Investigation of Spatial Position Effects

Space is to vision what time is to audition. Visual stimuli may be said to exist in space; auditory stimuli are more appropriately described as occurring over time. As evidence of visual coding in short-term memory, therefore, one might look for effects of spatial relationships among stimuli. An example of such an experiment is one by Nickerson and Pew (1973). On each trial, the subject had to make a decision about two successively presented pairs of visual stimuli. The decisions were of three different types: (a) Do the two pairs have at least one item in common? (b) Do they have both items in common? (c) Are they identical? In some cases the stimulus pairs that were compared were English letters; in other cases, they were random dot patterns. For all three types of decisions, performance proved to be highly sensitive to whether matching items, when they occurred, occupied the same positions in both memory and test displays. The effect was greater when the stimuli were random dot patterns than when they were letters, but it was obtained in both cases.

Manipulation of Expectancy for Auditory or Visual Stimuli

If one can retain either a visual or an auditory representation of a memory item, which he chooses to retain on any given trial might be expected to depend on the modality in which he expects the test item to occur. Tversky (1969) trained subjects to associate a different nonsense word with each of eight simple line drawings of faces. After having learned the names of the faces to a criterion, the subjects were given a same-different judgment task. On each trial two stimuli were presented successively, and the subject's task was to indicate whether the second stimulus had the same name as the first. All four possible combinations of verbal and pictorial stimuli occurring as first and second items were used. For any given experimental session, the first stimulus was invariably verbal or invariably pictorial; the second stimulus could be either of the two types, verbal or pictorial, but the subject knew that one of these possibilities was more probable than the other. In particular, one form was scheduled to occur on 79% of the trials, and he was told which form that would be. Reaction time was shortest when the second stimulus was in the "expected" form, whether or not this form happened to correspond to that of the first stimulus. Tversky interpreted this finding as evidence that subjects could encode the first stimulus either as a pictorial image or as an auditory name, and that they chose the form that corresponded to the one in which they expected the test item to occur.

Conclusions re Auditory and Visual Short-Term Memory

On balance, the evidence favors the conclusion that information can be retained in short-term memory in both auditory and visual form. The extent to which information is recoded and stored in a form in addition to that in which it is presented seems to depend on the nature of the task and, in particular, on what use the subject expects to have to make of the information that he is trying to retain. If, for example, he is going to have to make a same-different judgment he is likely to try to retain the first stimulus that is to be compared in the form in which he expects the second one to be presented.

Other things being equal, the tendency to recode visual stimuli into auditory representations seems to be stronger than the reverse. The fact that a recoding has occurred, or that the recoded representation is more accessible than the original one, does not demonstrate, however, that the original representation has been lost, or could not have been retained. Indeed, several of the results mentioned in the foregoing sections have been taken as evidence that information often is retained both in the form in which it was presented and in a recoded form.

Why should the tendency to recode be stronger for visual than for auditory stimuli? Undoubtedly, one reason is that, in most experiments that have compared the retention of visual and auditory stimuli, verbal stimuli and recall methods have been used. Ternes and Yuille (1972) have shown that even with nonverbal visual stimuli, the expectation of verbal recall induces a verbal coding process which is disrupted by verbal interpolated activity. By contrast, when a recognition response was expected, only the traces of visual verbal stimuli were disrupted by auditory activity. Extrapolating, in many experiments where visual materials were found to be sensitive to auditory variables (e.g., acoustic similarity), verbal recoding was probably induced by the design.

The differences between the retention of visual and auditory materials may not reflect any general differences in the two systems at all. Ingersoll and DiVesta (1972) have found that the relative memorial advantage of visual or auditory materials was entirely predictable from an individual's preferred modality of first report in a bisensory memory task. Those who tended to report visual items first performed better with visual materials in a missing-units task, while the opposite was true for those who tended to report auditory items first. Overall, performance on the missing-units task was better with auditory stimuli, but only because three-fifths of the subjects preferred the auditory modality.

Murdock and Walker (1969) have argued tht the differential ease of recalling auditory stimuli is related to the fact that both recall and auditory memory are inherently temporally ordered, but visual memory is spatially ordered.

Another factor that may induce recoding, particularly in the visual-to-auditory direction when recall is required, is that of overloading. In experiments in which nonverbal responses are required for visual verbal stimuli, the accuracy of visual recognition is typically found to be a function of visual factors, or at least to be unaffected by acoustic similarity factors (Chase & Calfee, 1969; Chi & Chase, 1972; Posner, 1967). Yet exceptions to this general rule exist. Kirsner and Craik (1971) found that if the stimuli consisted of a list of eight 2-syllable words, the time to decide whether or not the test stimulus had been a member of the list reflected separable contributions of auditory and linguistic memory, but no traces of visual memory. Comparing these data with data like Chase and Calfee's (1969) on the one hand, and those like Laugherty and Harris' (1970) on the other, we may conclude that the differential lack of visual factors was real, and resulted from a predominance of linguistic coding strategies when visual memory is overloaded.

The number 4 is constantly emerging in studies of short-term memory. For example, Sperling (1960) found that whole recall from briefly presented visual arrays was limited to about 4 items, and further (Sperling, 1963) that acoustic confusions did not occur until the number of items to be retained exceeded 4. Similarly, Kirsner and Craik (1971), in the study just noted, found that visually presented letters did not have to be recoded as linguistic strings until their number exceeded 4. Johnson (1970) and Simon (1970) deduced that the number of chunks that can be maintained in the STS was about 4, and both cite massive lists of data to support their estimates. Most of Miller's (1956) estimates also varied around 4 chunks, but, apparently, for the sake of consistency with absolute judgment data, he simply preferred the number 7. Further, Mandler (1967) repeatedly found that the optimal number of items or subcategories per category was very close to 4, and Dirlam (1972) mathematically derived that this should be so. Last, but not least, both the recency and primacy effects subtend about 4 items.

The recurrence of the number 4 in all of these situations seems to reflect one thing: any given active memory system can simultaneously maintain no more than about 4 items. After that, additional input results in interference. It is, therefore, imperative that the subject find another, and preferable more stable, memory code, and it is the establishment of this code that is reflected in the cognitive load of supraspan lists. Further, all of the hypothetically different memory stores ranging from the sensory stores to the STS, to the LTS, only reflect different levels of encoding. The differential longevity of their contents is only because the more highly coded the information, the less vulnerable it is to interference.

Memory for Complex Visual Patterns

So far, the discussion of visual memory has focused on memory for relatively simple stimuli (letters, words, simple patterns). Given that the visual system is capable of registering very complex patterns, there is always the possibility that performance limitations obtained with simple stimuli could result as much from the simplicity of the stimuli as from the limitations of the system. It has been shown that if subjects are allowed to inspect each of a series of hundreds of photographs of complex meaningful patterns for just a few seconds apiece, they are able to discriminate them from previously unseen pictures with nearly perfect accuracy (Nickerson, 1965; Shepard, 1967; Standing, Conezio, & Haber, 1970). What's more, the distinctiveness of the pictures in the inspection series may persist for weeks (Nickerson, 1968; Shepard, 1967).

While the results of these studies are provocative, they seem to raise more questions than they answer. Presumably, the process of recognizing a stimulus involves checking the familiarity of its constituent features. If familiarity exceeds some criterion, the stimulus is "recognized." Both the kinds of features that are checked and the value of the criterion must depend largely on the specific nature of the task. In all of the above studies, the stimuli were both complex and meaningful. They were taken from magazines with no stated selection criteria except that they be diverse. Consequently, we can say very little about the nature of the features on which performance was based. In Nickerson's (1965, 1968) studies, the subject's task was to indicate whether or not he had seen each picture previously during the experiment. In the other two studies (Shepard, 1967; Standing et al., 1970), the subject was always presented with a pair of pictures, one new and one old, and asked which of the two he had seen before. Given the diversity of the stimuli, neither procedure would seem to force a very stringent response criterion upon the subject, and, therefore, neither allows us to say much about the nature of the stored information on which performance was based.

In order to determine the most useful applications of pictorial displays for C^3 operations, a more precise understanding of memory for complex meaningful patterns is needed. For example, under what presentation conditions will subjects' memories for pictures be better or worse? Is there anything special about the way subjects examine a display when they are trying to remember it, as opposed to scanning it with no intention of remembering it? How would performance have changed if the stimuli in these studies had been more similar to one another? What kinds of similarities (e.g., elements of detail, spatial composition, meaning or subject matter) would have been most disruptive? Is the memory trace for a display essentially visual or is it abstract or propositional in nature? Does it consist of isolated details or a schema of the scene as a whole? How are spatial relations encoded? How is the encoded information forgotten; does it gradually fade away or is it lost in discrete chunks? How resistant is it to interference from extraneous verbal or pictorial input? The answers to most of these questions are unclear.

Encoding of Complex Displays

What determines how well a picture will be remembered? To answer this question, Loftus (1972) presented subjects with simultaneous pairs of pictures. In order to manipulate memorization strategies, he assigned a value to each member of the pair corresponding to the amount of money the subject would win or lose if he succeeded or failed to recognize the picture during testing. Each pair was presented for 3 sec., and the observer was free to distribute his attention across pictures as he chose. Eye movements were monitored.

Recognition accuracy increased with the absolute value of a picture as well as its value relative to the other member of the pair, as expected. Further, Loftus found that the recognizability of a picture was positively correlated with three inspection parameters--total viewing time, number of fixations, and duration of fixations--which were, in turn, correlated with one another. Somewhat surprisingly, when these inspection parameters were picked apart, only the total number of fixations was a good predictor of recognizability. Again, the problem of controlling the observer's visual fixations arises.

Despite Loftus' failure to find a direct association between the duration of fixations and recognition performance, the fixation period does seem to be integrally related to encoding activities. It has been shown that pictorial elements tend to be fixated for longer durations (1) when the subject is trying to memorize the display (Loftus, 1972; Tversky, 1974), (2) when the subject is examining the picture for the first time (Mackworth & Bruner, 1970), (3) when the subject is examining visual detail (Antes, 1974), and (4) when the image is blurred (Mackworth & Bruner, 1970). Moreover, using a more rigorous recognition test, Tversky (1974) found a positive relationship between the duration of visual fixations and picture memory.

The amount of attention allocated to the encoding process has been found to affect the strength of the memory trace for a picture, irrespective of scanning activities. For example, even with number of fixations controlled, if the subject is engaged in a secondary task during picture examination, subsequent recognition performance suffers (Loftus, 1972). Conversely, by varying both the exposure time per picture (0.25 to 2.0 sec.) and the duration of the blank interval (1.5 or 3.0 sec.) between pictures in the inspection series, Tversky & Sherman (1975) demonstrated that picture memory improves with the length of the total encoding interval, regardless of actual picture presentation time. Thus, while total exposure time per se seems to have little direct effect on picture memory, these studies suggest that, at least within the bounds of a few seconds, longer exposure times may be indirectly advantageous as they support multiple fixations and undistracted encoding activities.

The Nature of the Internal Representation

Given a thorough understanding of the ways in which various inspection parameters affect people's ability to recognize or recall a visual pattern, we should be able to devise presentation procedures that will make optimal use of that ability. Even so, human memory will be fallible. If we are to anticipate or safeguard against the kinds of forgetting and confusions that are most likely to occur, we need to understand the way in which visual information is internally represented.

For a long time, psychologists seemed to favor the view that visual information was not retained in a visual format at all.

They claimed that in order for information about a visual event to be remembered, it had first to be recoded as a set of verbal descriptors. This notion seemed to be supported by a variety of experimental findings, several of which were reviewed in a preceding section. For example, when people are asked to recall lists of visually presented stimuli, their false recollections often consist of responses that are acoustically similar to the names of forgotten items (Conrad, 1964; Conrad, Freeman, & Hull, 1965); the number of unrelated visual items that can be remembered corresponds to the verbal memory span for their names (Sperling, 1963); and the more words that are needed to describe the items in a visual array, the fewer the items that can be remembered (Glanzer, 1961; Glanzer & Clark, 1963). Moreover, the notion was consistent with the theoretical tenet that information could only be firmly implanted in long-term memory through the process of verbal rehearsal (Atkinson & Shiffrin, 1968; Murdock, 1967; Peterson & Peterson, 1959).

If visual information could, in fact, only be retained in the form of a verbal label or description, we would expect visual memory to be quite unreliable. Memory for complex scenes would be crude, especially given limited inspection times, and confusions among different items with the same name or similar verbal descriptions would be frequent.

It has been shown that people do, indeed, have a tendency to mistake novel items for familiar items with the same, or acoustically similar, names (Bencomo & Daniel, 1975; Conrad, 1964). However, they are also prone to confuse visually similar, but nominally distinct, items (Bencomo & Daniel, 1975; Palmer, 1975; Posner, 1967; Tversky, 1969). Such findings as these have led psychologists to support what is known as the dual-coding hypothesis of visual information processing (Paivio, 1971). Very simply, this hypothesis states that easily labeled, pictorial items tend to be represented by both visual and verbal codes or traces in memory. By contrast, verbal stimuli do not generally elicit imaginal codes.

The dual-coding hypothesis provides at least a partial explanation for the typical resiliance of people's memories for visual information. It seems that memories may become relatively inaccessible or that they may become partially confused or merged with traces of similar events, but that information, once stored, is rarely forgotten, in the true sense of the word. If the dual-coding hypothesis is correct, pictorial information is bound to be more retrievable than otherwise comparable verbal information. First, if a pictorial event results in two, as contrasted with one, memory trace, then the a priori odds that a record of its occurrence will remain accessible are better than for comparable verbal material. Second, even if one of the traces of a pictorial event is effectively lost through interference, the critical information may survive in the other.

Even when there is no need for long-term remembrance, the dual-coding hypothesis predicts certain advantages for pictorial as opposed to verbal inputs. Information in working or short-term memory is particularly susceptible to interference from physically similar events (Allport, Antonis, & Reynolds, 1972; Atwood, 1971; Brooks, 1967, 1968; Reitman, 1971; Salthouse, 1974). Thus, memories for verbal information may be virtually wiped out by irrelevant verbal activity. But to the extent that graphic information is encoded both visually and verbally, neither visual nor verbal interference alone can obliterate the memory (Pellegrino, Siegel, & Dhawan, 1975).

The practical implications of the dual-coding hypothesis are obvious. But, before going as far as to endorse the use of graphic information whenever possible, one should consider its limitations. Most especially, the compulsion to label verbally a pictorial event is much stronger for, if not limited to, isolated and easily identifiable items. The ability to remember more nebulous forms can generally be improved by attaching meaningful descriptive labels to them, but this procedure introduces errors of its own inasmuch as people's memories for the form tend, with time, to drift toward forms more accurately described by the label (Daniel, 1972).

Moreover, if people do, as a rule, try to generate verbal descriptions of complex visual scenes, their descriptions are, at best, very sketchy. As compared with people's impressive ability to recognize complex visual scenes, their ability to describe or reconstruct such scenes from memory is disappointing (Haber, 1970). Perhaps this should be taken as evidence that memories for complex visual scenes are generally vague and incomplete. Perhaps recognition studies to date have just been too insensitive to reveal the paucity of precision and detail. Alternatively, Haber has interpreted the discrepancy between recall and recognition performance as an indication that people's visual memories for complex scenes are generally very good, but are not readily accessible to verbal or reconstructive output modes.

In support of this contention, it has been shown that the amount of information about a visual display that can be recalled may increase with contemplation. Haber and Erdelyi (1967) found that recall of pictorial information was improved after 30 min. of verbal free-associative activity, but not after 30 min. of irrelevant nonverbal activity (dart-throwing); thus, the effect cannot be simply attributed to a recovery of the memory with time. Further, Erdelyi and Becker (1974) were able to argue that the effect was not entirely due to a relaxation of reporting criteria by demonstrating that multiple-recall trials and, moreover, interpolated thinking, resulted in increases in the recall function for pictures but not for words.

Haber and Erdelyi's findings are somewhat comforting at a philosophical level. If people really are bad at expressing what they remember about visual events, then it is encouraging to know that they can remember more than they say. However, their findings are frustrating from a practical point of view. In many situations, it matters little how much a person knows unless he can make that knowledge public. The relationship between visual memory and its reportability clearly needs more research. For example, how great is the discrepancy? What aspects of the memory tend to be most accessible? What kinds of distortions are liable to occur in reports or reproductions? What techniques can be implemented to increase the accessibility of visual memories?

The inexpressable nature of visual memories also complicates the task of determining what is and is not retained. Given the inadequacy of verbal reports and graphic reproductions, recognition tests become the most obvious means of assessing visual memory. But recognition tests may grossly overestimate the information that is actually stored in memory. In theory, remembering is normally a two-process operation: the information is first retrieved or called into consciousness, and then its familiarity or appropriateness is assessed. Under recognition procedures, the information is provided such that only the second process need be executed (Kintsch, 1970b). Thus, information that is unavailable to the subject when he is left to his own devices may be fully recognizable. In keeping with this notion, Dallet, Wilcox, and D'Andrea (1968) have shown that people's ability to recall visually, or conjure up, an image of a complex scene is only roughly correlated with their ability to recognize the scene if testing occurs immediately after presentation; if testing is delayed by one week, recall and recognition scores are virtually unrelated.

What is worse is that recognition responses are fairly inscrutable measures of the goodness of the visual trace for a complex array. If the subject says yes that he recognizes a picture, how can we tell what aspects of the picture are familiar to him? If he says no, what aspects of the picture look new or wrong? The problem is only exacerbated by the fact that it may not do any good to ask him.

Still, it should be possible to get a general idea of what people remember through experimentation. One means towards this end would be through recognition tests in which the similarity of the test items to the target items is systematically manipulated. In this way, the nature and intricacy of the subject's memory for the target items could be roughly deduced from the kinds of test items that he does and does not discriminate from the target items. Within this framework, the visual similarity of the target and test items could be manipulated on at least two dimensions. That is, one could alter either the spatial organization of the array or the nature of its component elements. Surprisingly little research of either sort has been done.

Spatial Memory

People's memory for spatial organization has often been investigated with nonsense forms or random dot patterns (Attneave, 1957; Caldwell & Hall, 1970; Pick, 1965; Posner, 1969). The use of such arbitrary patterns, instead of meaningful visual patterns, insures that manipulations of their spatial components will not be cued by changes in the meaning of the relationships or interactions between their figural components. These studies indicate that people have an impressive ability for recognizing spatial distributions. Moreover, they indicate that their recognition decisions tend to be based on very complex internal representations or schemata of the array as a whole. The exact location of any specific element does not matter as much as the spatial relationship of each of the elements to each other.

Posner and Keele (1968, 1969) may have provided the most persuasive support for this position. They generated a set of nonsense dot patterns by systematically distorting the arrangements of several base or prototype patterns. These distortions were used as stimuli in a paired-associate learning task wherein all of the variations of a particular prototype required a common response. After subjects reached criterion with these patterns, they were transferred to a new set which included: (1) the patterns they had just learned, (2) the prototypes of those patterns (which they had never seen before), and (3) control patterns that were of varying similarity to the prototypes and the learned patterns. Subjects were, of course, more accurate with the learned patterns, but were very nearly as good with the prototypes. The ease with which the control patterns were learned depended on their distance from the old patterns, but even those that were equidistant with the prototype resulted in almost twice as many errors as the prototype. This pattern of results was not only maintained but, in fact, became more dramatic when the transfer task was delayed for a week after training. Apparently, what subjects had learned was not information about specific dots or dot clusters, but information about the patterns as wholes.

It is not at all clear that findings with nonsense patterns can be extrapolated to complex meaningful stimuli: but very little has been done by way of testing people's sensitivity to spatial parameters of real-world displays. It would be useful to know both how readily people can detect various spatial distortions of real-world displays and how easily they can recognize a given scene under various ecologically permissible transformations, for example, as seen from a different vantage point.

There are some data indicating that, just as with nonsense patterns, people may be more attuned to the Gestalt or complex of spatial relationships within meaningful patterns than they are to the specific spatial coordinates of their elements. For example, people can apparently recognize mirror images of complex scenes as well as the originals (Dallet, Wilcox, & D'Andrea, 1968; Standing, Conezio, & Haber, 1970). Interestingly, in most cases, they are also able to discern whether or not a picture has been reversed (Blount, Holmes, Rodger, & Coltheart, 1975; Standing et al., 1970).

However, Dallet et al. (1968) also found that people were much less able to recognize displays that were turned upside-down. Possibly, this difficulty is due to a disturbance of meaningful rather than spatial dimensions of the arrays. Alternatively, it may be that people are more adept at adjusting for some spatial transformations than others. In a three-dimensional world where we normally stand on our feet, it is adaptive to be able to recognize objects from literally diametric perspectives. It is not nearly so useful to be insensitive to whether or not things are upside down. Moreover, Pick (1965) has demonstrated with nonsense forms that experience with different spatial transformations positively transfers across visual events. That is, if people are familiarized with the effects of a particular transformation on a given set of items, then they can more readily identify items not in the set under the same transformation.

Again, the problem in doing such experiments with coherent, meaningful stimuli is that it is very difficult to alter their spatial relationships without altering them in other important ways. The use of nonsense forms stands as one means around this problem. However, this work may have limited generality inasmuch as the only information in a nonsense form is spatial information. An alternative approach is to use arbitrary arrangements of meaningful elements.

The latter tack has resulted in very different results. In particular, it seems that information about the identity and the position of an item are processed by separate mechanisms (Adams, 1975; Finkel, 1973; Salthouse, 1974). Positional information may be processed primarily by what Neisser (1967) has termed preattentive mechanisms. These are the mechanisms which act to parse a scene into units and thereby enable further processing. By contrast, item identification presumes the preattentive mechanisms and involves focal or selective attention.

In keeping with this distinction, studies involving tachistoscopic presentation procedures have repeatedly demonstrated that spatial information is processed more easily than, and prior to, identity information. In fact, partial report procedures often rely on this fact (Averbach & Coriell, 1961; Sperling, 1960; Von Wright, 1968). Further, with only a few items in the memory set, subjects will recall the position of an item correctly if they can recall its identity, but the converse is not true (Dick & Dick, 1969). Finally, memory for positional information is very good even when the number of items exceeds the immediate memory span, or processing time is severely limited. However, if item identification is emphasized under either of these conditions, the positional information tends to be lost (Finkel, 1973; Wolford, 1975).

The vulnerability of positional information in memory is somewhat counterintuitive. After all, variations on the basic method of loci have been among the most common tricks of mnemonists (Bower, 1970; Luria, 1968; Paivio, 1971). These methods would only work if positional information were easier to recall and were integrally associated with identity information in memory. Under what conditions will this be the case?

The classic method of loci involves the one-to-one association of each of the to-be-remembered items with each of a standard, structured set of places or mnemonic bins. For example, the places might correspond to the rooms and courtyards of a palace. Then each of the to-be-remembered items would be imaginally located in one of the prescribed sites in the palace. At the time of recall, the mnemonist imagines that he is walking about the palace by a standard route, and reports the item that he finds in each of the critical locations.

Bower (1970) has analyzed the classic method of loci into a set of nine distinct components. These are:

1. There is a known list of "cues."
2. The cues are memory images of geographic locations.
3. Cues and items on the list to be learned are to be associated during input of the list items.
4. Associations are to be made in one-to-one pairings.
5. Associations are to be effected through imaginal elaboration, specifically by use of visual imagery.
6. The imaginal construction should be unusual, bizarre, striking.
7. If the list items are studied a second time, the same items should be placed at the same loci; even if ordered output is not required, constant ordered input is desirable.
8. At the time of recall the person must cue his recall of the list items.
9. The recall cues must be the same as, or similar to, those he thought of while studying the items.

After Bower, we can evaluate each of these components in terms of its mnemonic utility, and, in that way, determine which is essential, which is superfluous, and, more importantly, why the method of loci works.

1. <u>A known list of cues</u>. The important aspect of this item is that the list of places or mnemonic bins must be readily available to the mnemonist during both the encoding and retrieval of the to-be-remembered items. If the cue list also required effort to generate, the system would hardly be economical. Notably,

the cue list need not be memorized. It is also effective to associate the to-be-remembered events with a series of externally generated events as long as the series is completely available during both input and output episodes. In fact, much of the research on paired-associate learning can be viewed from this perspective.

2. The cues are memory images of geographic locations. This restriction is apparently unnecessary. The important property is rather that the list be well structured so that none of the cues will be overlooked during recall. This property is intrinsic to a well planned path through a palace, for example, inasmuch as one must pass through each location to get to the next. However, any well structured string of concrete items will do. Bower cites, as an alternative example, the pegword system where the cues are concrete nouns that rhyme with numbers, e.g., "one is a bun, two is a shoe, three is a tree, etc." Under this system, the first to-be-remembered item is imagined in some interaction with a bun, the second with a shoe, and so on. As with the method of loci, the pegword system insures that no item should be inadvertently omitted on recall unless the person cannot count. A fixed sequence of externally generated events holds the same advantage.

3. Cues and list items are to be associated during list input. This point is essential, but should be made more strongly. The cues and the items must be purposefully associated during list learning. Clearly, knowing a list of locations or pegwords does not help unless it is properly applied. Further, as shown by some of the forementioned studies, externally determined associates of the to-be-remembered items may even be suppressed unless the memorizer recognizes their relevance.

4. One-to-one pairings of cue and item. Bower (1972) has shown that multiple items can be effectively hooked to a single pegword if they are elaborated in a grand, interactive imaginal scheme. However, the advantages of one-to-one pairings are that they insure, first, that the order information is preserved and, second, that no item will be inadvertently omitted.

5. Use of imaginal elaboration, especially visual imagery, and,

6. Unusual, bizarre imagery. These two points go together. Visual imagery is not essential--rhythmic constructions, rhymes, and other nonvisual associations may also be effective. Wollen, Weber, and Lowry (1972) have also shown that the bizarreness requirement is unnecessary. What is necessary is that the to-be-remembered items and the mnemonic cues be integrally associated. For visual mnemonics, the trick is to create an image wherein the cue and the to-be-remembered item are interacting. Elaboration is necessary inasmuch as the to-be-remembered item must be recoded to fit the mnemonic. Bizarreness may help to distinguish the current list from previous lists if the mnemonic is frequently reused.

7. Repeating each item on the same locus. This simply reinforces the association between the item and the cue, thus making the cue more effective.

8. At recall, the person must use cues, and

9. Recall cues must be similar to those studied. These two points are important inasmuch as the cues serve to locate and organize the list items in memory. In addition, they underscore the basic weakness of using external rather than memorized

mnemonic cues because only the latter have guaranteed availability at the time of recall.

In summary, the reason that the method of loci works is not because there is anything intrinsically more memorable about spatial locations. Rather, it is one instance of an effective class of mnemonic systems. The critical components of a good mnemonic system are that (1) the cues are readily available, (2) it is well structured, and (3) the to-be-remembered items are integrally associated with the cues. Such systems work because they provide a means of retrieving or locating the items in memory and because they impose an extrinsic organization on the items such that recall can be ordered and complete.

Conceptual Organization and Visual Memory

While memory for a visual display is clearly influenced by stimulus factors and inspection parameters, there are other, higher-order cognitive processes that are probably even more important. One variable that surely makes a difference is the subject's conceptual familiarity with the displayed information. The classic example of this relationship comes from studies of chessboard perception (Chase & Simon, 1973; de Groot, 1965). If master players are allowed to study a legitimate game position for about 5 sec., they are able to reconstruct it almost perfectly from memory. However, the number of correctly placed pieces falls off sharply for less and less skilled players. This difference cannot be attributed to differences in either general familiarity with the chessboard or visual memory capacity: when the pieces are randomly placed on the board, the masters perform just as badly as the beginners, nor can it be entirely attributed to the masters' familiarity with the specific configurations tested.

Rather, the difference seems to lie in the ability of the player to organize the board as a coherent pattern.

Eisenstadt and Kareev (1975) have provided a cogent demonstration of this point. Their experiment was based on two different, two-person board games: GO and GOMOKU. Although the two games involve the same type of board and playing pieces, they involve different rules and strategies. Kareev and Eisenstadt first taught their subjects how to play both games. For the experiment each subject was shown a particular configuration of pieces and asked to generate the best move for one of the players. The critical aspect of the procedure was that, for a given configuration, the subjects were sometimes told that it was a playing position from a game of GO and sometimes that it was a position from a game of GOMOKU. After the subject had declared a move, he was unexpectedly asked to reconstruct the configuration of pieces from memory. Kareev and Eisenstadt found that if subjects had analyzed the board as a GO configuration, they remembered more of the pieces that were crucial for the game of GO; conversely, if they had analyzed the board as a GOMOKU configuration, they remembered more of the pieces that were crucial to the game of GOMOKU. Since the displays were the same in both cases, the difference in memorability can only be attributed to differences in subjective organization.

This kind of pattern perception deserves more attention. Although the studies cited above concentrated on the perception of board games, one would expect to find the same sorts of familiarity factors operating in the perception of maps, circuit diagrams, indicator boards, instrument panels, and schematics in general. It is important, in and of itself, that the expert can assimilate the array more readily than the novice. But one would further like to know, for example, how pattern familiarity affects the viewer's capacity for other cognitive operations. Mewhort (1972) has shown that under tachistoscopic conditions letter strings that conform more closely to English orthography absorb more processing capacity during initial perception than do letter strings that are orthographically irregular. However, after several hundred milliseconds the situation reverses: the regular strings become less taxing than the irregular strings. Mewhort cogently argues that this shift in processing demand corresponds to a switch from encoding activities to retention activities. Whereas the irregular stimuli are simply encoded as an ordered string of letters, the regular stimuli additionally invoke higher order codes and, consequently, require more encoding effort. Yet, as the result of the higher order codes, the regular stimuli are more coherent, occupy fewer units or "chunks" of memory, and are, therefore, easier to retain than the orthographically irregular stimuli. Mewhort's (1972) findings are provocative, but their generality cannot be taken for granted. Letter strings are very special visual stimuli for the normal, literate adult. Under appropriate presentation conditions, letter strings can be clearly perceived within just a few milliseconds. By contrast, the chess players discussed above studied the board for a few seconds--a difference of three orders of magnitude.

If Mewhort's results can be qualitatively generalized to more complex visual stiuations, such as the reading of chessboards, maps, or wiring diagrams, it is not clear how the time course of the processing capacity effect would be altered. Possibly, the experienced viewer, as compared to the naive viewer, would need to allocate a differential proportion of his cognitive capacity to the encoding process throughout the course of array examination. It seems intuitively more probable, however, that encoding activities per se would be spread out across the examination period. If this is true, then the experienced viewer might, in effect, enjoy a constant advantage in processing capabilities over the naive viewer. That is, not only would the experienced viewer have more capacity to manipulate or contemplate internalized information between encoding episodes, but he would also have more and better organized information to work on.

A second reason why the generality of Mewhort's (1972) results may be limited also derives from his use of letter strings as stimuli. Words are such highly overlearned visual stimuli for the literate adult that they tend to be perceived more quickly or accurately than single letters (Reicher, 1969; Wheeler, 1970). The implication is that whole words can provide more information to the reader than can be derived in sum from their individual letters. This extra information cannot be entirely attributed either to superficial cues such as word shape or to active guessing strategies. Rather, it seems to be rooted in the reader's familiarity with different spatial distributions and combinations of letters. This knowledge seems to be automatically and effortlessly invoked during stimulus processing such that the letters of relatively frequent strings mutually facilitate one another (Adams, 1975).

This phenomenon, whereby a cluster of discrete visual elements is more perceptible than any of the elements alone, has been demonstrated with schematic faces (Homa, Haver, & Schwartz, 1976) and with visual arrays that capture some of the properties of good figures, in the Gestalt sense (Pomerantz & Garner, 1973; Weisstein & Harris, 1974), as well as with words. What is not clear is whether the phenomenon is restricted to a small set of extremely familiar and basic kinds of visual schemata, or whether it may be an aspect of familiar perceptual situations in general. For example, is it this phenomenon that underlies the chessboard effect? Assimilating a multidimensional array as a whole may be vastly more efficient than separately processing and then recombining its individual elements. It would be useful to know the extent to which this sort of perceptual integration could become operative on such complex arrays as maps, aerial photographs, or circuit diagrams, and the nature and extent of experience with such stimuli that would be required.

A word of caution is in order here. The advantage of the experienced observer in terms of his capacity to perceive and remember an array derives from his ability to organize or group the information in the display according to some familiar schema. This is equivalent to saying that the experienced observer initially is strongly biased to interpret or encode the array in a habitual way. Further, since his mnemonic advantage over the naive observer is precisely due to the cohesiveness of his mental representation, he may also find it particularly difficult to reinterpret or manipulate the information in the array in novel ways. The implication is that in some situations, display interpretation might be best accomplished by a team of observers, including one person with extensive experience and one who is moderately naive.

At the extreme, conceptual information may even distort or override visual information about a situation. Norman and Rumelhart (1975) have described several examples of this phenomenon. In one case, they asked people who lived in university housing to draw floor plans of their apartments. All of the apartments were alike, and all of the subjects had lived there for some time. Nevertheless, the subjects' floor plans included systematic distortions that seemed to derive from inferences about structural regularities of buildings. In another case, graduate students were asked to draw pictures of the building where they had their classes, offices, and laboratories. Again, there were systematic errors that could only be attributed to nonvisual misconceptions about the building. The students were then asked to distinguish between an accurate drawing of the building and various distorted renditions. Interestingly, some of the distorted versions were again systematically preferred to the most accurate drawing.

Stevens (1976) has argued that such spatial distortions are often the predictable byproducts of what is generally an efficient and economical memory system. According to his model, our memories for spatial relations are hierarchically organized. The spatial relations between any two loci are only directly encoded if the two loci are part of the same superordinate spatial unit. Information about the relative locations of sites that belong to different superordinates is not explicitly stored, but is available through inference or reconstruction: the individual must combine his memories for the relationships between the superordinate units and the locations of the two relevant sites within their respective superordinate units.

This kind of hierarchical organization is economical in that it allows for the requisite storage capacity of the system to be greatly reduced. For example, if one stores the fact that North America is in the Western Hemisphere, one need not redundantly store the facts that every country, state and city within the continent are in the Western Hemisphere. Instead, one can infer that information from their superordinate unit, North America. However, to the extent that information is overgeneralized or overregularized within the system, systematic errors should result.

Stevens has capitalized on these errors in order to demonstrate the validity of his model. In one experiment, he asked subjects to indicate the direction from one real-world location to another. As predicted, the judgments were distorted toward the relative locations of the relevant superordinates. For example, San Diego, California, was judged to be southwest of Reno, Nevada, rather than southeast, as it actually is; presumably, this was because subjects knew that California is basically west of Nevada. Similarly, the Atlantic entrance to the Panama Canal was judged to be due-east of the Pacific entrance, rather than northwest, as it actually is. Stevens found the same sorts of distortions when subjects were asked to make directional judgments about simple maps they had studied in the laboratory.

While the spatial distortions that Stevens discusses are more or less predictable, and reflect a generally useful property of human memory, there are other, more insidious kinds of distortions that can arise through related, nonvisual experiences. Loftus (1975) has shown that linguistic information about a visual event can alter the memory considerably. She showed her subjects video tapes of various events and then assessed the accuracy with which they could answer questions about those events. The critical aspect of her procedure was that she would introduce new information to the subject verbally, either prior to, or during, questioning. For example, in one case her subjects saw a film of eight demonstrators disrupting a classroom. After the film, she asked the subjects a question either about the leader of the "twelve demonstrators" or about the leader of the "four demonstrators." A week later the subjects were asked how many demonstrators there actually had been. The estimates of the subjects who had heard the "twelve" question in the interim were significantly higher than those of the subjects who had heard the "four" question.

Loftus has shown that more subtle suggestions may be just as influential. In another experiment (Loftus & Palmer, 1974), she showed subjects a film of a traffic accident and then asked them how fast the two vehicles were going when they "smashed." Subjects' estimates of the vehicles' speeds, given the word "smashed," were significantly higher than when that word was substituted with less potent verbs, like "collided," "bumped," or "hit." Moreover, through the use of carefully worded questions, Loftus has not only succeeded in distorting subjects' memories for what they did see, but, further, has gotten them to testify that they saw things that were never really there. Again, the suggestion may be fairly blatant, as in asking about "the X." Or it may be quite subtle: subjects more readily testified that they had seen broken glass in a film of an automobile accident if they had previously been querried about the cars' having "smashed" into one another than if they had been querried about the cars' having "hit" each other.

Loftus's explanation for these findings is simply that people tend to store all information about a given event in a single, complex integrated trace. Since knowledge gained both prior to and subsequent to the event in question may be inextricably entwined with information acquired directly from the event itself, the memory may be substantially, and yet unknowingly, distorted.

How can such distortions be safeguarded against? For one, Loftus's work clearly suggests that we should be careful to use neutral words and avoid presuppositions when interrogating observers. But, what if the damage has already been done? What if the observer's memory has already been distorted through conversations or other extraneous but related experiences?

This brings us back to the question of whether visual experience may be retained in some essentially visual format that is inaccessible and, therefore, immutable through the other modalities. Some psychologists have argued strongly against this notion (Pylyshyn, 1973). But, as discussed above, there is also evidence that it may be correct. If so, then the "visual" distortions described in this section may be happening to the verbal or conceptual trace of the event. The visual trace may remain intact like a buffer copy of the event. If this were true, and if some way of accessing the visual trace could be figured out, then it would provide a check against conceptual distortions.

CHAPTER 4

DECISION MAKING AND THINKING

Many investigators have tended to conceptualize both decision making and thinking as processes involving sequences of steps. Dewey (1910), for example, attempted to identify the stages by which "reflective" thinking proceeds in practical situations. After studying a number of students' reports of "simple, but genuine cases of reflective experience," he concluded that

> "Upon examination, each instance reveals, more or less clearly, five logically distinct steps: (i) a felt difficulty; (ii) its location and definition; (iii) suggestion of possible solution; (iv) development of reasoning of the bearings of the suggestion; (v) further observation and experiment leading to its acceptance or rejection; that is, the conclusion of belief or disbelief" (p. 72).

Numerous other conceptualizations of problem-oriented thinking, and decision making, have been proposed, some of which identify fewer steps and some many more (e.g., Johnson & Jennings, 1963; Schrenk, 1964; Wallas, 1926).

This "episodic" view of thinking is very useful for many purposes, and has provoked much research. There are alternative ways to conceptualize how problem solving and decision making get done, however, that may provide a more suitable frame of reference for thinking about these processes in the context of C^3 system operations. A busy commander, or command staff, tends to be engaged in a continuous stream of activity; conceiving of that stream as divided into discrete episodes is, for some purposes, somewhat artificial. The final section of this chapter presents an alternative, less episodic, conception of the operator's task, and points to some of the human factors questions such a conception raises.

In the intervening sections of this chapter, we review selectively the results of research relating to several aspects of decision making and thinking that we feel are particularly relevant to the design and operation of C^3 systems, and that represent problems on which further research has a reasonable chance of effecting significant improvements in C^3 operations.

Some of the best known and most interesting kinds of experiments on the psychology of thinking and decision making will not be discussed here. Many important experiments in this field were done to study the influence of factors that can be manipulated in the laboratory, but not in the design of a C^3 system. In particular, some factors that have large effects on problem solving and decision making can be manipulated only by an experimenter who knows the solution to the problem or knows what the correct decision is. In the practical situations that arise in a center for command, control, and communications, the "right answer" is not known--either by the designers who planned the center or by the personnel who operate the center when the situation arises. If the right answer were known, then, from a practical point of view, the problem would be solved or the decision would be made.

Man as an Intuitive Statistician

Recently, a number of studies have been conducted in an effort to characterize the capabilities of people as intuitive statisticians and logicians, and to discover the sorts of rules and heuristics they use when evaluating data and predicting future events. In addition to being intrinsically interesting, the results of some of these studies are relevant, we think, both to the design of data acquisition and retrieval systems, and to the development of problem solving and decision aids.

The results of laboratory investigations of man's capabilities as an intuitive statistician and comparisons of his performance with that prescribed by normative statistical theory have potential relevance to descriptive modeling of behavior in operational settings and to the development of data stores. A classic paper by Peterson and Beach (1967) highlights those aspects of human performance that are thought to be of importance and that, with some further research, could yield to quantitative expression.

The long-range goal of the work described by Peterson and Beach is the development of a theory of human behavior in uncertain environments. Of prime interest are those environments that can be characterized statistically and those tasks for which a model of ideal performance can be specified. For our purposes, the cited paper provides a convenient catalog of observer performance in two major areas: (1) estimates of the characteristics of samples of data, and (2) the use of samples to infer characteristics of parent populations. The first of these is referred to by the authors as "intuitive descriptive statistics"; the second as "intuitive inferential statistics." Experimental findings relating to both of these areas may be summarized as follows:

Intuitive Descriptive Statistics

a. Judgments of proportion. The relation between mean estimates and sample proportions is described well by an identity function with a maximum deviation of the mean estimate from the sample proportion of 0.3 to 0.5 and an average deviation very close to zero. There may be a very slight overestimation of low proportions and an equally slight underestimate of high ones, depending upon experimental conditions.

b. Judgments of means and variances. Subjects familiar with the concept of a mean of a distribution give highly accurate estimates, though there is some variance as a function of sample variance, sample size and rate of presentation of data. There are systematic discrepancies between intuitive judgments of sample variance and actual variance. In general, as means increase, estimates of variance decrease.

Intuitive Inferential Statistics

a. Inferences about proportions. In experimental tasks requiring sequential revision of subjective probabilities, subjects are typically found to be conservative in their behavior, revising their estimates in the appropriate direction as additional data are acquired, but at a rate considerably less than is justified.

b. Inferences about means. Inferences about the central tendencies of symmetrical distributions are generally accurate. In J-shaped distributions, estimates of medians and modes are accurate, but those of the mean are biased toward the median.

c. Inferences about correlations. Evidence suggests that subjects do not pay attention to all cells of a 2 x 2 contingency table when inferring correlation. That cell in which favorable outcomes appear together or those associated with the positive diagonal seem to receive primary consideration. As the tabular dimensions increase beyond 2 x 2, inferences become more nearly like those assessed by normative correlation statistics.

d. Determining sample size. In experiments in which a subject must infer which of two populations is being sampled, he has continuous control over the number of samples taken, a cost is assigned to each sampled datum, and the dependent variable is the number of data purchased prior to making a decision, it is typically found that although performance is influenced in the proper direction by cost/payoff structure, the magnitude of the influence is less than that prescribed by ideal models. When the ultimate size of the sample is decided ahead of time by the subject, the cost/payoff structure appears to have no effect on behavior. When prior probabilities are reduced by defining more alternatives, subjects (appropriately) purchase larger samples of data. When only two alternatives are involved and prior probabilities are made disparate, fewer data are purchased. With any number of alternatives or distributions of priors, however, purchasing performance is conservative relative to that dictated by the optimal (Bayesian) rule.

Observations such as those above are useful in model building activity in two different ways: (1) they help to establish the forms of relationships between independent variables and performance in situations requiring quantitative inference, and (2) they help to define the ranges over which these relationships can be expected to obtain. Because of the frequency with which the phenomenon of conservatism in sequential sampling has been demonstrated, we believe there is sufficient justification for attempting to capture its basic form in descriptive models of human performance. For any given situation, of course, some prior empirical work must be accomplished to define its range.

With respect to the other indicators of suboptimal performance, we feel less certain. Perhaps only additional research within the precise contexts of given operational tasks can verify their existence and significance.

Estimation Heuristics

Considerable work has been done on intuitive statistics in recent years by Kahneman and Tversky (1972, 1973; Tversky & Kahneman, 1971, 1974). They have been concerned in particular with identifying the heuristics that people use to arrive at estimates of statistical parameters. Two such heuristics that they have described are the principles of representativeness and availability. According to the representativeness hypothesis, people select or order outcomes on the basis of the degrees to which these outcomes represent the essential features of the evidence they have examined. In doing so, they frequently ignore the prior probabilities of those outcomes and/or the degree of unreliability inherent in the evidence. Where the representativeness of the outcomes is, in fact, matched to the likelihoods, this behavior leads to a statistically appropriate selection (or ordering). In situations where they are not matched, however, the selection (ordering) will be inappropriate.

An important prediction from this hypothesis is that, whereas the processes of evaluating a set of data and of predicting an outcome on the basis of the data are independent and should entail differing amounts of uncertainty when the data are less than perfectly reliable, people will perform as though the processes and uncertainties were similar. The essential difference between evaluating an input and predicting an outcome is highlighted in an example provided by the authors:

> "Suppose one is told that a college freshman has been described by a counsellor as intelligent, self-confident, well-read, hardworking, and inquisitive. Consider two types of questions that might be asked about this description: (a) Evaluation: How does this description impress you with respect to academic ability? What percentage of descriptions of freshmen do you believe would impress you more? and (b) Prediction: What is your estimate of the grade point average that this student will obtain? What is the percentage of freshmen who obtain a higher grade point average?"

The authors point out that there is surely greater uncertainty concerning the second question than the first; hence, the prediction should be closer to 50% than the evaluation. They predict, however, that subjects will produce approximately the same estimate for both questions since they will consider the best prediction to be that score which is most representative of the input data.

The second prediction resulting from the representativeness hypothesis is that the degree of confidence one has in his prediction from a set of data reflects the degree to which the outcome he has selected is more representative of the set than are other possible outcomes. Thus, one might have more confidence in the prediction of an overall B average in future courses on the basis of B grades in two separate introductory courses than he would have given an A and a C in those courses. This, the authors note, is incompatible with the common multivariate model of prediction in which predictive accuracy is independent of within-profile variability.

The tendency, if it exists, to predict solely on the basis of the input data and to ignore the prior probabilities of chosen outcomes may also be at the root of commonly observed failures to appreciate regression phenomena. In Kahneman and Tversky's view, the regression effect typically violates the intuition that a predicted outcome should be maximally representative of the input information.

According to the availability hypothesis, a person judges the frequency or probability of an event or class of events on the basis of the ease with which relevant instances are brought to mind. An example provided by Tversky and Kahneman (1974) is that of assessing the risk of heart attack among middle-aged people by recalling such occurrences among acquaintances.

As in the case of inference via representativeness, the predictions one is led to under this heuristic are frequently correct, since instances of frequent events are typically recalled better and more quickly than instances of less frequent events. Associated with its use, however, are certain biases which occasionally produce inappropriate judgments: some of these are as follows:

a. Biases due to retrievability of instances. A class whose instances are easily recalled will, using this heuristic, appear more frequent than a class of equal frequency whose instances are less easily recalled.

b. Biases due to effectiveness of search set. In tasks requiring estimates of the relative frequencies of different words, the availability heuristic will lead to the judgment that abstract words are more numerous than concrete words.

c. Biases of imaginability. In instances in which one must assess the frequency of class whose instances are not stored in memory one may identify a rule or algorithm that can, in turn, generate instances for consideration. The ease with which instances can be generated then b ecomes the criterion for frequency assessment. Depending on the characteristics of the rule and the prior likelihoods of the instances generated, the assessment may or may not be accurate.

d. Illusory correlation. This term refers to a tendency to overestimate the frequency of the co-occurrence of events that are naturally associated with each other.

We have noted a number of the empirical expectations that follow from the hypothesis that representativeness and availability heuristics frequently underlie the evaluation of data and the prediction of outcomes. These and others of perhaps lesser importance in the current context can be summarized as follows:

a. Judgments of the probabilities of future events will tend to be representative of the input data and not to reflect the prior probabilities associated with those events.

b. Judgments of the probability of obtaining a particular result in a sample drawn from a specified population will be relatively insensitive to the size of the sample.

c. The confidence one has in a prediction will depend almost completely on the degree to which that prediction is representative of the input data and will be insensitive to the reliability of the data.

d. The normal tendency for the performance of people (or machines) to oscillate about some mean level--hence, for a better-than-average or poorer-than-average performance to be followed immediately by an "average" performance--will fail to be recognized. One possible outcome of this failure may be that punishment will be perceived to be more effective than reward.

e. When judging the relative size of classes by the availability of their instances, a class whose instances are easily recalled will appear larger than one whose instances are less easily recalled.

f. If the frequency of words is judged by availability of contexts, abstract words will be judged to be more numerous than concrete words.

g. The frequencies of co-occurrences of natural associates will tend to be overestimated.

h. The frequency of conjunctive events will tend to be overestimated; the frequency of disjunctive events will tend to be underestimated.

Findings of these types have potentially great relevance to the descriptive modeling of human performance. They would appear to be applicable to many situations in which an operator must predict the future state of a system on the basis of noisy input data, or where he is required to perform fault-isolation and trouble-shooting tasks. We expect that for such situations the basic forms of the relationships between input data and evaluation and prediction performance would be relatively easy to determine empirically, though values suitable for precise estimation in either analytic or simulation models might require a relatively extensive program of parametric research.

The point should also be made that since the behavioral phenomena of interest relate largely to man's capabilities as an intuitive statistician, there exists, in the form of predictions derived from normative statistical theory, a ready standard against which to compare his performance. The ability to achieve this comparison can aid directly in assessing the relative merits of alternative system aids to the evaluation and prediction performance of the human.

Where the problem becomes much more complex is in situations for which normative statistical theory is not applicable. The accuracy of estimates of the probability of occurrence of unique or near-unique events (nuclear world war), for example, cannot be determined by an appeal to normative statistics. And the results of studies of performance on tasks involving estimates of relative frequencies should not be extrapolated uncritically to probabilistic situations for which probability and relative frequency are not synonymous. Much work remains to be done on the question of how well people estimate probabilities of this type.

Intransitivity of Preferences and Non-Independence of Beliefs and Desires

Practically all prescriptive models of utility begin with the assumption that each of the possible outcomes associated with action alternatives open to the decision maker at the time of choice has two important parameters: (1) a value denoting the decision maker's "taste" or preference for the given outcome, or that outcome's real (e.g., monetary) value, and (2) a probability of occurrence. Where the scales of values and probabilities assigned over the set of possible outcomes can be made to correspond to certain "rules," identification of the prescribed choice is, at least in principle, easily achieved. The rules, as formulated by Savage (cited in Becker & McClintock, 1967) may be summarized as follows:

Rule 1: Transitivity. If, in a choice situation, the decision maker prefers Outcome A to Outcome B and Outcome B to Outcome C, he should prefer Outcome A to Outcome C.

Rule 2: Comparability. The decision maker should be willing to compare two possible outcomes and decide either that he prefers one to the other or that he has no preference between them.

Rule 3: Dominance. If the decision maker determines that, *under every possible condition*, a choice of one of his alternative actions results in an outcome at least as desirable as that which would result from the choice of a second alternative action, and results in a more desirable outcome *under at least one possible condition* than would the second action, the second action should not be preferred to the first.

Rule 4: Irrelevance of nonaffected outcomes. If the decision maker determines that, for a particular state of the world, two or more of the actions open to him result in the same outcome, his preferences among such actions should not be affected by the outcome associated with that state.

Rule 5: Independence of beliefs and rewards. The decision maker's statement concerning the likelihood of occurrence of a given outcome should not be affected by what he hopes will occur.

Despite their intuitive appeal, at least some of these rules seem actually be to honored more in the breach than in the observance (Nickerson & Feehrer, 1975). Empirical studies, for example, frequently show that, whereas the decision maker may prefer Outcome B to Outcome A and Outcome C to Outcome B, he may also prefer Outcome A to Outcome C. On occasion this intransitive pattern will persist even when the decision maker is confronted by the experimenter with the apparent inconsistency.

Other violations occur with respect to Rule 5, a common finding being that decision makers attach higher estimates of probability of occurrence to those outcomes that they consider more desirable. Violations of the remaining rules occur occasionally, but are, perhaps, of less immediate concern to the formulation of decision aiding techniques than those associated with transitivity and non-independence.

There are at least two possible approaches to solving the intransitivity/non-independence problems. One is to attempt to train decision makers to be aware of such occurrences and to teach them methods for achieving transitivity and independence. A second (not mutually exclusive) approach is to generate procedures that accommodate intransitivity and non-independence when they occur, and thus allow the decision maker to achieve whatever orderings he feels comfortable with.

Bayesian Inference with Reliable and Unreliable Data

Given a set of mutually exclusive and exhaustive hypotheses, (H_i) and datum, D, Bayes rule expresses the probability that a given hypothesis, H_j, is true as a function of $(p(D|H_j)$, the probability that D will be observed given that H_j is true, and $p(H_j)$, the probability that H_j is true prior to the observation of D. That is,

$$p(H_j|D) = \frac{p(D|H_j)p(H_j)}{p(D)}$$

where

$$p(D) = \sum_{i=1}^{n} p(D|H_i)p(H_i);$$

n equals the total number of hypotheses in (H_i).

For a sequence of observations, this formulation is applied recursively, the value of $p(H_i|D)$ computed as the result of one observation serving as the $p(H_i)$ for the next.

Since Bayes rule specifies an optimal policy for aggregation of data against hypotheses and for the sequential revision of probabilities distributed over alternative hypotheses as data are examined, it has served as an important normative framework with which to evaluate the performance of real decision makers. Comparison of man's actual data aggregation performance against the dictates of Bayes rule has become one of the leading preoccupations in modern decision-making research. These efforts have been directed at finer understanding of one of the earliest and most persistent findings in the Bayesian decision-making area, viz, that humans appear to assign a less extreme probability to the likelihood of a given hypothesis after examining a given datum than is justified on the basis of the prescriptive rule. This basic finding and its implications for the training of decision makers have been discussed by Nickerson and Feehrer (1975).

Recently, there has been considerable interest in extending the Bayesian model to situations in which the decision maker may be in doubt as to the actual occurrence of an event on which his inference must be based. Such situations can arise in at least two different ways: (1) During an observation period, the occurrence of one event may be mistaken for the occurrence of another; (2) the report of an observer of a correctly made observation may be in error. In the first of these cases, the decision maker must decide whether the reported event or some alternative event occurred on the basis of an observer's expressed degree of certainty. In the second, the decision maker must make an inference concerning the reported event on the basis of the estimated credibility of the reporting source. Both situations are of the "cascaded inference" type, requiring that the decision maker discount the nominal diagnostic value of a datum as observed or reported in the light of the observer's certainty, or credibility, and then, in a second step, determine the support for the alternative hypotheses provided by the adjusted datum.

Approaches to the formulation of prescriptive and descriptive models for cascaded inference are presented in Nickerson and Feehrer (1975). Of that set, models concerned with inference in the context of faulty reporting are perhaps of greatest interest to the design and operation of C^3 systems because of the extent to which they may be generalizable to any situation in which the decision maker is faced with "noisy" reporting systems. Though relatively little effort has been made so far in this direction, there seems no reason, in principle, why these formulations could not be employed in the study of information acquisition and reporting operations in any system where estimates of the reliabilities of communication channels utilized by the decision maker can be made.

We present below a portion of Nickerson and Feehrer's discussion that outlines the prescriptive Bayesian models of Schum and DuCharme (1971) for four situations in which the nominal impact of the report of a datum must be altered in light of the estimated credibility of the reporting source. This presentation is followed by portions of our earlier discussion of descriptive models in this area.

The Prescriptive Models of Schum and DuCharme

These authors have presented formulations for cases in which there are two possible hypotheses, H_1 and H_2, two possible data events, D_1 and D_2, and two possible reports of the event that occurred, d_1 and d_2. The decision maker's task is to determine $p(H_i|d_j)$. According to Bayes rule:

$$p(H_i|d_j) = \frac{p(d_j|H_i)p(H_i)}{p(d_j)}$$

which can be manipulated to obtain:

$$p(d_j|H_i) = \sum_k p(D_k|H_i)p(d_j|H_i \cap D_k)$$

if $p(D_k|H_i)$ and $p(d_j|H_i \cap D_k)$ are known. Given this, a likelihood ratio (Λ) adjusted for the diagnosticity of the datum can then be derived.

$$\Lambda = \frac{p(d_j|H_i)}{p(d_j|H_k)}$$

The authors distinguish four decision "cases" which differ with respect to the symmetry of the impact of observing a given event on $p(D|H)$ and $p(d|D)$ and derive the optimal Λ for each. A summary of each of these follows:

Case I: Symmetric $p(D|H)$: $p(D_1|H_1) = p(D_2|H_2)$

Symmetric $p(d|D)$: $p(d_1|D_1) = p(d_2|D_2)$

$$\Lambda_{s,s} = \frac{pr+(1-p)(1-r)}{(1-p)r+p(1-r)}$$

where

$p \equiv p(D_i|H_i)$

$1-p \equiv p(D_j|H_i)$, $j \neq 1$

$r \equiv p(d_i|D_i)$

$1-r \equiv p(d_j|D_i)$, $j \neq 1$

Case II: Asymmetric $p(D|H)$: $p(D_1|H_1) \neq (D_2|H_2)$

Symmetric $p(d|D)$: $p(d_1|D_1) = p(d_2|D_2)$

$$\Lambda_{a,s} = \frac{p_1 r+(1-p_1)(1-r)}{p_2 r+(1-p_2)(1-r)}$$

or, equivalently,

$$\Lambda_{a,s} = \frac{p_1+k}{p_2+k}$$

where

$$k = \frac{1-r}{2r-1}, \quad r \neq .5$$

$$p_1 \equiv p(D_1|H_1)$$
$$1-p_1 \equiv p(D_2|H_1)$$
$$p_2 \equiv p(D_1|H_2)$$
$$1-p_2 \equiv p(D_2|H_2),$$

and r and 1-r are defined as above.

Case III: Symmetric $p(D|H)$: $p(D_1|H_1) = p(D_2|H_2)$

Asymmetric $p(d|D)$: $p(d_1|D_1) \neq p(d_2|D_2)$

$$\Lambda_{s,a} = \frac{pr_1+(1-p)(1-r_2)}{(1-p)r_1+p(1-r_2)}$$

or, if $p \neq 1$ and $r_2 \neq 1$,

$$\Lambda_{s,a} = \frac{c\left[\frac{p}{1-p}\right]+1}{c+\left[\frac{p}{1-p}\right]}$$

where

$$c = \frac{r_1}{1-r_2}.$$

Case IV: Asymmetric $p(D|H)$: $p(D_1|H_1) \neq p(D_2|H_2)$

Asymmetric $p(d|D)$: $p(d_1|D_1) \neq p(d_2|D_2)$

$$\Lambda_{a,a} = \frac{p_1 r_1 + (1-p_1)(1-r_2)}{p_2 r_1 + (1-p_2)(1-r_2)}$$

or, if $r_1 = (1-r_2)$

$$\Lambda_{a,a} = \frac{p_1 + b}{p_2 + b}$$

where

$$b = \left[\frac{1-r_2}{r_1-(1-r_2)}\right]$$

and the remaining parameters are defined as above.

Descriptive Approaches

Actual observation of decision makers attempting to infer states of the world on the basis of unreliable data have led to a number of empirical models, the most significant of which are due to Snapper and Fryback (1971), Gettys, Kelly, and Peterson (1973), Phillips and Edwards (1966), and Funaro (1974).

Snapper and Fryback suggest that in dealing with unreliable data, decision makers first estimate the likelihood ratio as though the data were completely reliable, then adjust this ratio by multiplying it by a reliability quotient, and, finally, apply the adjusted ratio to the calculation of posterior odds.

Stage 1: compute $\lambda = rL$

Stage 2: compute $\Omega_1 = \lambda\Omega_0$

where

r = the reliability of the report
L = the likelihood ratio
$\tilde{\Omega}_1$ = the adjusted posterior odds
Ω_0 = the (unadjusted) prior odds

According to Gettys, Kelly, and Peterson's model, the decision maker estimates posterior odds assuming the most likely event is true, and then adjusts the odds to reflect the reliability of the source.

Stage 1: compute $\Omega_1 = L\Omega_0$
Stage 2: compute $\tilde{\Omega}_1 = r\Omega_1$

Edwards and Phillips present evidence suggesting that posterior odds are described by a single-stage process, as follows:

$$\Omega_1 = L^c\Omega_0$$

where c varies with L.

Funaro has recently attempted to evaluate the predictive power of Snapper and Fryback's model and of that of Gettys, Kelly, and Peterson, using both L and L^c as unadjusted likelihood ratios in each case. A symmetric $p(D|H)$--symmetric r task (Schum & Du Charme, Case I) was used. Subjects were required to revise odds' estimates under both single-stage (perfect-source reliability assumed) and cascaded-inference conditions. Values of c were estimated separately for individual subjects from data obtained in the single-stage conditions.

The results were not consistent with any of the models described above. They were predicted best by another model that Funaro proposed. This "empirical model" assumes that subjects accurately estimate Λ, and then apply this estimate to the revision of odds with the same degree of effectiveness, or ineffectiveness, with which they apply L in single-stage tasks. The conclusion appears to be inconsistent with the results of Youseff and Peterson (1973) who found that odds' estimates made under cascaded conditions were consistently excessive relative to those made in single-stage tasks, given $\Lambda = L$. Funaro notes, however, that subjects in his experiment could have acquired a direct appreciation for Λ from the proportion of successes and failures in a series of reports obtained from the source during the course of the experiment. (In a symmetrical $p(D|H)$ chips-in-urn situation, one can unambiguously define a "success" as the drawing--or in this case reporting--of a chip of the predominant color.) To the extent that subjects were able to develop a direct awareness of Λ, the effect would have been to eliminate the need for a two-stage process and to transform the task into the simpler problem of revising odds on the basis of totally reliable data. The suggestion is an eminently plausible one, and the possibility that this is in fact the way unreliable data are often accommodated in real-world situations deserves further study.

Constraints on Generality of the Bayesian Framework

Despite the acknowledged productivity of the Bayesian framework, it has a number of properties that should be recognized in assessing its generality and usefulness for real decision-making situations. These are dealt with at some length in Nickerson and Feehrer (1975) and will only be summarized here.

1. The rule concerns only that aspect of decision making related to hypothesis evaluation.

2. It requires uncertainty to be expressed in terms of a mutually exclusive and exhaustive set of phyotheses.

3. There is no provision for adding new hypotheses to an existing set of hypotheses once evaluation has begun.

4. Bayes rule does not determine the "truth value" of a given hypothesis, only with its statistical support.

5. Incorrect assignment of prior likelihoods may be devastating when hypotheses are evaluated on the basis of few data.

6. Bayes rule itself does not provide a direct criterion for "stopping" in a sequential data-acquisition situation.

It is difficult to identify any system, however carefully designed and operated, that does not occasionally provide decision makers with data whose reliability is suspect. When this is the case, and Bayesian analysis can be applied without considerable oversimplification of the decision-making situation, models of the types reviewed above may be very productive. Some hint of this is provided by the recent efforts of Schum (1975a,b) to extend his analysis to jury proceedings, an area long neglected by behavioral research and theory.

Finally, within the framework of a research program aimed ultimately at enhancing the performance of C^3 functions, there are a number of questions raised by evidence accumulated in Bayesian research so far. Some of them are:

a. Can decision makers be trained to perform at levels that more closely approximate optimality in those operational situations in which conservatism is manifest?

b. Whether or not their performance could be improved through training, could a system be configured to help decision makers know when their performance is suboptimal?

c. Are there decision-making contexts in which performance evaluated as suboptimal with respect to the Bayesian metric is optimal (or at least more nearly optimal) with a different, and equally acceptable, yardstick?

d. How accurately can decision makers perform when they are familiar with optimal rules for aggregating data against hypotheses and for adjusting diagnostic impact when reliability information is available?

Studies of Concept Attainment

Much of the research on how people form and test hypotheses has been done under the rubric of concept formation or concept attainment. The typical task in this sort of experiment is to discover some "concept," which is usually defined in terms of a conjunction, disjunction, or complex function of stimulus features (e.g., "red and square," "red and small and not square," or "blue and square").

In the usual laboratory experiment on concept attainment, there is nothing probabilistic about the stimuli with which the subject is confronted. A card that contains two small figures within a single border is a positive instance, any other card is not. The cues--number of figures, size, and number of borders--

combine in a perfectly deterministic fashion to specify whether the stimulus is included within the class that defines the concept. Moreover, there is nothing subtle about the cues themselves; a stimulus card clearly does or does not have a single border, the figures are clearly small or they are not, and so on.

The concepts an operator of a C^3 system must identify in deciding what the state of the world is are often not so clear-cut. They are more like the concept a line-backer in a football game uses when he tells himself, "This is a passing situation." In the first place, the cues are probabilistic; there is no one cue, and no combination of cues, that determine unequivocally whether the situation is an instance of the concept "a passing situation." In the second place, the cues themselves may be more obscure than the presence or absence of a border on a stimulus card.

Both of those factors have been found to retard concept attainment. Bruner, Goodnow, and Austin (1956) did an experiment in which the stimuli were drawings of people who differed in sex, kind of clothes, and so on. Concept attainment was considerably slower with these stimuli in which the cues, though not especially obscure, were not so blatant as in the typical experiment. In other experiments, Bruner et al. also found that concept attainment was greatly retarded if the cues were probabilistic, i.e., if there was no cue or combination of cues that unequivocally determined whether the stimulus was or was not an instance of the concept. Since then, there does not seem to have been much laboratory research on concept learning under circumstances that are realistic in the sense that the cues are obscure and probabilistic.

Memory Limitations in Concept Formations

One of the most severe limitations on a person's ability to form a concept is the limitation imposed by his memory. He simply cannot remember all the positive and negative instances he has seen. The classic research of Bruner et al. on concept attainment was primarily on the information-processing strategies that subjects adopt to circumvent the limitations of memory. Some of these experiments used the selection procedure, that is, the subject chose a stimulus, was told whether it was a positive or negative instance, chose another, and so on. The sequence of checks revealed fairly directly what exploratory strategy he was following. Similar strategies were also detected by more indirect means in experiments that used the more familiar reception procedure--that is, experiments in which the subject saw the stimuli one-at-a-time in an order the experimenter had chosen in advance.

The question for human factors research is how a person's memory might be assisted so that he could employ a more efficient strategy and attain a concept after seeing a smaller number of positive and negative instances. This is a field in which a considerable amount of research has already been done (Bourne, Ekstrand, & Dominowski, 1971, pp. 234-237), both demonstrating the utility of an external memory in concept learning and, to some extent, revealing the manner in which the subject is able to use the information stored there. In a sense, the practical problem is usually a problem of display: the computer can store information about previous situations easily enough; the problem is how to display the information so that an operator can detect any regularities there may be. One possible approach is Chernoff's (1973) faces, but there must surely be others.

Strategy in Seeking Information

Quite apart from failures of memory, people often fail to use good strategy in seeking information. The questions they ask are not always those that would be the most revealing, or would provide the information at the least cost. This was shown in the concept-formation experiments of Bruner et al. that were just cited, and also in a variety of experiments on trouble-shooting, diagnostic problems, and guessing games (Johnson, 1972, pp. 163-168).

The human factors question is how operators might be trained to use good strategy in seeking information, and whether there are ways in which a computer could assist them to seek information efficiently.

Remembering what Tacks have been Explored

There are circumstances in which it is very difficult for a person to remember which courses of action he has evaluated, and with what results. This is especially evident in the kind of thinking Newell and Simon (1972) have analyzed in great detail--the kind required in playing chess, in proving theorems, and in solving cryptarithmetic puzzles. In chess, for example, the evaluation of a course of action sometimes requires a lengthy mental exploration of what the consequences of the action would be. After that exploration, the player may find he has trouble remembering what other actions he has evaluated, and with what result. Similarly, in solving a puzzle, it is quite easy to get confused about what possibilities one has tried and rejected.

Inability to recall what courses of action have been evaluated is a special case of what Newell and Simon describe as inability to return to a previous "knowledge state." A person working on a complicated problem cannot remember everything he has done, and so he cannot return at will to any arbitrary point in his past thinking. In fact, Newell and Simon (p. 815) estimate that their subjects usually retained only enough information to permit a return to one or two "knowledge states." This limitation on a person's memory for his own thoughts is a crucial one. Newell and Simon (pp. 814-818) point out that it puts sharp constraints on the problem-solving strategies people can use. Attempting to provide a problem solver with a complete record of his own thoughts is, presumably, out of the question, but if one could supply him with, or help him to maintain, a partial record, what part would be the most valuable? We suggest that the most useful information would be a list of actions he had evaluated, and the evaluations he had given them. Although this would be a potentially very fruitful subject for research, it is not, so far as we know, a problem on which research has been done.

Finding and Evaluating Courses of Action

This aspect of decision making has two parts: first, identification of all courses of action that are promising enough to be worth evaluating, and, second, evaluation of those alternatives for the purpose of picking the best. The first process is emphasized in those laboratory experiments on problem solving in which the evaluation of a course of action is trivial. A particular action obviously solves the problem or does not; the point is to discover an action that does. The second process is emphasized in those experiments on decision making in which the list of alternatives is given to the subject, and his task is only to choose the best one from among them.

Remembering the Consequences of Previous Actions

The value of utilizing case histories in teaching such things as law, medicine, military strategy and tactics is well known. And, at least in the first two of these instances, continued access to precedents and records of earlier diagnoses and treatments serves an important role in the day-to-day operations of practitioners.

It is not clear, however, that, beyond the initial grounding in military history received as part of a military-school curriculum, commanders ever have the opportunity to search for precedents that could aid them in the solution of current problems. As a result, a commander may be required again and again to confront specific tactical or strategic aspects of an engagement as though they were unique, while, at the same time, he feels that those aspects must have been considered before and that he might benefit from reviewing the conclusions reached.

One possible conclusion we could come to is that a way should be found to provide the commander with the time and resources necessary to search for occasions on which either he or a counterpart has faced similar circumstances in the past. Unfortunately, however desirable that might be on occasion, it is clearly impractical unless ways can be found to capture and describe the structure of case histories and current problems economically. Search for precedents might then proceed efficiently on the basis of some well-defined template-matching process.

A system that aided the decision maker by enabling him to search for similar past situations in an efficient manner might be of further benefit if it could help him to assess the specific similarities and differences between the current situation and those selected by the template-matching process. It might even be argued that the more imprecise the matching process, the more important provision of such a capability would be. This is almost certain to be the case if the imprecision of the matching process results from reduction in the dimensionalities of past and current situations. Where this was the case, however, it also seems clear that more information concerning the past situation would have to be potentially available than was employed in the matching process itself, if a further similarity/difference analysis was to be performed.

Some research exists that indirectly supports an argument that the effort required to define and specify a precedent-seeking aid might be justified in terms of the effects it could have on decision making (e.g., Fischhoff, 1974a,b; Fischhoff & Beyth, 1974; Wohlstetter, 1962). All of these studies draw attention to the fact that situations and events judged in hindsight--that is, after they have occurred--are perceived quite differently (presumably, because of knowledge of their outcomes) than they were perceived in foresight. An example of what Fischhoff (1974b) called "creeping determinism" is provided by his quote of the historian, Florovsky.

> "The tendency toward determinism is somehow implied in the method of retrospection itself. In retrospect, we seem to perceive the logic of the events according to a recognizable pattern with an alleged inner necessity. So that we get the impression that it really could not have happened otherwise" (p. 288).

A series of three experiments was run by Fischhoff in an effort to answer two questions concerning the differences between hindsight and foresight: (1) "How does the receipt of outcome information affect judgment?" and (2) "How aware are people of the effects that outcome knowledge has on their perceptions?" (p. 288). The results of these experiments suggest strongly that people are very much influenced by a knowledge of the outcomes of situations they are called upon to judge. As suggested by the "creeping-determinism" hypothesis, the effect was one of finding that the observed outcome was inevitable, given the situation, and that it was, in fact, apparent in foresight. The judges were, moreover, generally unaware of the effect that outcome knowledge had on their perceptions.

Actions Suggested by the Machine

A common problem in decision making is that of considering too limited a range of actions and overlooking alternatives that would be more effective than those considered. The sort of inflexibility that this represents has been investigated repeatedly in laboratory experiments. Many of these experiments are classified as experiments on problem-solving set (Bourne et al., 1971; Johnson, 1972). Perhaps the best known are Luchin's (1942) which involve problems in which the subject is asked how one could measure a certain amount of water using combinations of jars of certain sizes and subtraction or addition operations. The salient point is that having learned to use a particular method to solve each of a series of problems, subjects were unlikely to discover a simpler method when they encountered problems for which it would have been appropriate.

The same type of inflexibility, or something very similar, has been observed in experiments on what Duncker (1945) has called "functional fixedness" (Bourne, et al., 1971; Johnson, 1972). The problems that typically have been studied in these experiments can be solved only by using a familiar object in an unusual way, for example, by thumbtacking a box to the door to make a shelf, or by tying a relay to a string to make a pendulum. The subject will fail to solve the problem if his conception of the "function" of a relay is so"fixed" that he cannot think of using it as a weight.

In command and control situations in which the individual has the task of identifying a variety of possibilities from which to choose a course of action, it seems obvious that this sort of rigidity could be detrimental. One way to free the individual's imagination might be to have the machine present him with a variety of suggested courses of action that are quite different from each other. Even if none of the suggestions was acceptable in itself, they might stimulate his thinking and provoke him to conceive of better alternatives. The question of whether, or under what circumstances, this technique would be helpful, is an empirical one. So also is the question of whether the suggestions would have to be sensible to be useful; it is not clear, for example, that bizarre and impractical suggestions might not provide the stimulus to thought that could sometimes lead the system user to produce a better set of his own.

Commitment

An individual who has made a choice among decision alternatives and has made that choice "public" is, in several significant respects,- operating in a different mode than is one who has either not yet made a selection or at least has not yet announced it to the world. Certain dimensions of this difference have been recognized for some time; others are only slowly becoming apparent. What has been learned to date may have important implications for decision makers who must exercise their skills in rapidly changing situations, and for the design of procedures to support that activity.

One of the earliest differences to be observed was reported by Gibson and Nicol (1964). Subjects in this study assumed the role of field commanders who were required to predict the direction from which a guerilla attack would come on the basis of continually updated map displays of enemy patrol activity. Following each update in the sequence, the subject was required to estimate his degree of belief concerning from where the attack would come by distributing a set of twelve points over the four possible cardinal directions (N, E, S, W). Thus, a distribution [3,3,3,3] would indicate a belief that, as of that time, the alternatives were equally likely, while a distribution [3,2,6,1] would indicate a belief that the attack was most likely to occur from the South, least likely to occur in the West, and slightly more likely in the North than in the East.

As the experiment progressed and the expectations of subjects were built up (as indicated by increasing values being assigned to a particular alternative), positive indications of patrol activity in one or more low-valued directions were increased in frequency. This had the effect of maintaining the judgment process in a continual state of flux and made it possible to study the transitions from one belief state to another.

Two outcomes of this study are of interest in the present context. The first is that more information was required by the subjects to change their belief states, once these had developed, than was required by them to arrive at those belief states in the first place. In the authors´ words: "Once a person has decided that a certain state of the world exists, it is going to take more contraindicative information to change that decision than it took information to originally arrive at that decision from a point of neutrality" (p.11).

The second finding of interest is that there was a significant interaction between the time at which belief states changed and the amount by which they changed. Again, in the authors´ words,

> "This seems to indicate, at least within the limits of this experiment, that a major change in conditions is perceived far less readily by an observer who has been developing a conception of the situation over a long period of time than by a person who has been developing a conception of the situation over a shorter period of time" (p.13).

As the authors point out, these results represent an extension of the findings reported by Edwards (1965) that suggest the decision maker is an inefficient processor of information when compared against the Bayesian ideal. The results seem to show that the degree of inefficiency grows with the amount of time spent in establishing a commitment to one or another hypothesis and that, once formed, and expressed, this commitment is relatively resistant to contraindicative evidence.

The basic phenomenon of commitment has been demonstrated many times in a wide variety of contexts and some of the behaviors associated with the phenomenon in more complex situations have been described. Hoyt and Centers (1972) for example, found that subjects who were publicly committed to a given position tended to write essays espousing a more moderate position than did subjects who had made no (public) commitment. Salancik and Kiesler (see Kiesler, 1971) in a study of learning and retention of attitudinally related word pairs found that committed subjects recalled a significantly larger number of consistent pairs and a significantly smaller number of inconsistent and irrelevant pairs than did uncommitted subjects.

One of the most impressive studies of "real world" decision making that highlights the effects commitment may have on subsequent cognitive processes was performed by Soelberg (1967) on prospective MIT Sloan School students in the throes of making choices among employment opportunities. Through the use of a non-obtrusive interview technique begun shortly after the set of job offers was first considered and continued through the decision period until an unequivocal selection was made - in some cases, a period of ten weeks or more - this investigator was able to show that commitment to a particular alternative may be made very early in the process and that most post-commitment activity is directed toward justifying the choice made. Just how early may be judged from the experimenter's comments regarding the dynamics of the decision process:

1. "The decision maker believes a priori that he will make his decision by weighting all relevant factors with respect to each alternative, and then 'add up numbers' in order to identify the best one. In fact, he does not generally do this, and if he does it is done after he has made an 'implicit' selection among alternatives."

2. "Evaluation during the search phase takes the form of screening each alternative along a number of non-compared goal dimensions; no evidence of factor weighting is apparent at this stage."

3. "...when he ends his search for new alternatives, he will have more than a single alternative in his 'active roster'."

4. "When the subject terminates his search for new alternatives before his search resources run out, he will already have identified a favorite alternative..."

5. "At the point of search termination a person generally will not have compared his alternatives with one another, will not possess a transitive rank ordering of alternatives, and will refuse to admit that his implicit choice has been made."

6. "Confirmation processing aims to resolve the residual uncertainties and problems connected with the choice (i.e., preferred) candidate and to arrive at a decision rule which shows unequivocally that the choice candidate dominates the confirmation (non-preferred alternative in a two alternative scheme) candidate - Pareto dominance being the ideal goal strived for."

7. During confirmation processing a great deal of perceptual and interpretational distortion takes place in favor of the choice candidate, to the detriment of the confirmation candidate, goal attribute ´weights´ are arrived at, or changed, to fit the perceived data and the desired decision outcome."

"The decision is made when a satisfactorily Pareto dominant decision rule has been constructed..."

9. "Search for new alternatives end a significant period of time before the decision maker is willing to admit having made his decision."

10. "When the decision maker ends his search for new alternatives he will report significant uncertainty about which alternative he will select as his choice."

It is tempting to conclude that human propensities to form commitments early in a decision process and to remain, thereafter, relatively refractory to the possible impact of subsequent data render them important targets in the design of systems and procedures that will aid decision making. This, we feel, is an appropriate conclusion, but the nature of the system or procedure to be developed may be in doubt.

One reason for this doubt is that many (perhaps most) complex situations are of a nature that makes a definition of "ideal" performance very difficult to specify. In these situations, questions as to whether commitments are formed too early and/or abandoned too late may be unanswerable in any objective sense. One may only be able to observe their onset and follow their dynamics, as Soelberg has done.

A second reason for doubt is that early and/or long continued commitments, when based on intuitions that may be contraindicated by available data but that ultimately prove to have been correct, may be regarded as desirable. Like the judgmental heuristics discussed earlier in this report, they may represent valuable "shortcuts" to problem solving when conditions are right.

Still a third reason is that, whatever else might be said for and against commitment behavior, it does represent a stabilizing influence in the decision making process. Appropriately or not, it enables a decision maker to simplify complex problems by reducing their dimensionality, thereby, decreasing the amount of data that have to be processed. In this sense, commitment as a phenomenon has the same sort of ecological validity attributed to "availability" and "representativeness" heuristics.

From the point of view of decision aid design, then, it would be desirable to have a means for detecting the presence and duration of a commitment and for helping the decision maker decide whether or not that commitment - or, indeed, *any* commitment - is appropriate at that time. The emphasis here might be placed on making the decision maker aware of his behavior and conscious of the possibility that he might ignore or undervalue information that is crucial to the decision, rather than directing his attention explicitly to one or another hypothesis best supported by data available to the system. Alternatively, a set of procedures might be devised that places equal emphasis on both approaches and that enables the decision maker to base his strategy on the merits of the situation (e.g., on the extent to which he believes the system "understands" his value system and contains all data of importance to his problem).

Unfortunately, much of the guidance that would be required to develop such an aid is not available. It is not clear, for example, what sorts of interrogation procedures would be appropriate for detection and tracking of commitments. One can easily think of several ways these could be accomplished if one did not need to be concerned with the extent to which the procedure intruded itself into the decision making task, *per se*, or the extent to which it represented such an annoyance that its use would be curtailed. But, clearly, these are very important constraints. One goal for research, then, is the discovery of non-obtrusive, "cooperative" decision-process sampling techniques.

Decision Making under a Tactical Disadvantage

In his study of behavioral aspects of tactical decision making, Sidorsky (1972) reports a result which, though unverified to date, may have implications for both the training and aiding of decision makers in systems. The task was one in which a (simulated) commander of an ASW attack submarine must decide when to fire a missile at an approaching enemy submarine. The problem posed by the situation is that, although the probability of a successful hit on the target increases as a function of decreasing range, the probability that the presence of the commander's own ship will be detected and fired upon also increases with decreasing range. The commander's decision, then, must be based on a tradeoff between the probabilities of hitting the enemy and being hit by him. Further complicating his decision is the relative tactical advantage afforded his own or the enemy ship as a result of differences in depth, local bathythermal conditions, etc.

Sidorsky found that when subjects perceived themselves to be in an advantageous position or when an advantage did not exist for either ship, the quality of their decisions was relatively high - that is, the range at which they chose to fire their missile was close to the optimum hit/detection tradeoff point. On the other hand, when subjects perceived themselves to be at a disadvantage, decision quality deteriorated markedly.

The response method chosen for the experiment and the analytic framework used for evaluation of decision quality provided little opportunity for inferences concerning why the behavior occurred. Nonetheless, if it stands as a predictable piece of behavior - that is, if the phenomenon can be replicated over a range of simulated tactical encounters - it raises some interesting research questions. Among these are:

a. Does the quality of decision performance scale with the perceived degree of disadvantage?

b. How highly does perceived disadvantage (or advantage) correlate with objective disadvantage (or advantage) in those situations for which the latter factor can be defined and scaled?

c. Would it be desirable to provide to commanders decision aids designed explicitly for use when they are operating at a perceived disadvantage? What types of aids would be appropriate for these cases?

We are inclined to think that the behavior exhibited by Sidorsky's subjects may be one example of a more general class of suboptimal behaviors associated with performance under stressful conditions. If so, research on this phenomenon also provides an opportunity to work within a context in which dimensions of stress might be objectified and measured.

Group Decision Making; The Risky-Shift

It is a working assumption in many organizations that the "product" (decision, evaluation, problem solution, etc.) delivered by a group of individuals differs from that delivered by the individuals working alone. Sometimes these differences are considered desirable and the organization makes an explicit effort to utilize group problem solving approaches. Sometimes the differences are considered undesirable and critical problems are left to individual attention.

What are the alleged differences? Personal experience and folklore seem to suggest that one is the time taken to reach a conclusion, and a generalization often heard is that a problem in need of a quick resolution should never be left in the hands of a committee. A second possible difference is that groups operating under an appropriate set of ground-rules may deliver a greater number of creative ideas per unit of time than would their members working individually. Thirdly, as dedication to the concept of trial by jury attests, the collective opinion regarding evidence may more nearly represent the truth value of that evidence than does the opinion of a single individual. From a practical point of view, it seems that there may be problems so complex as to preclude solution by the efforts of any individual. In these cases, whatever opinions one might have concerning group productivity, there appears to be no alternative to the use of groups of persons with the required cross-section of knowledge and skills. How best to utilize the collective talents of any given group and how to interpret the group's output, however, are questions the answers to which are not always apparent.

We believe that some C^3 problems are of the type that require the combined efforts of collections of specialists. One could take the position, for example, that the entire strategic or tactical intelligence gathering and analysis activities of a given C^3 system are group problem solving enterprises. Moreover, one could take the position that although many C^3 problems are cast, for purposes of research, as individual challenges - the intelligence analyst's "problem" in assessing the aggregate meaning of discrete indicators, the commander's problem in choosing an optimal course of action, the monitor's problem in interpreting raw sensor data - this orientation may reflect the availability of analytic tools and methods more than the realities of the tasks under study.

In this section, we discuss one phenomenon of group decision making that may be relevant to military decision making where consensus is required before a plan of action can be implemented.

"Rarely in the history of social psychology has a single study stimulated as much research as the master's thesis by Stoner (1961) which reported the discovery of the 'risky shift.' Its conclusion that groups are riskier than individuals was widely interpreted as being contrary to the findings of previous research on the effects of groups on individuals. It challenged conventional wisdom, and it appeared to have implications for those responsible for making decisions involving risk" (p.361).

This assessment by Cartwright (1971) in a review of risky shift research 10 years after Stoner's finding and its first replication and extension by Marquis (1962) hardly overestimates the impact that the work has had on both empirical and theoretical research on group decision making. In addition to introducing a new topic area with an easily administered experimental paradigm and results with a high degree of replicability into the mainstream of social psychology, investigators of risky shift appear to have had at least an indirect influence on interests and procedures in other areas of group research (see, for example, the study of the "group polarization phenomenon" by Myers & Lamm, 1975).

The vast majority of risky shift studies have, following Stoner, employed a subset of the 12 Choice-Dilemma Problems (CDP) originated by Wallach and Kogan (1959) for their studies of individual differences. Each of these problems briefly describes a particular situation in which the decision maker is faced with two alternative courses of action differing qualitatively in attractiveness and probability of successful outcome. Examples of two of these are presented in complete form by Marquis (1962) and Wallach, Kogan and Burt (1967) respectively:

<u>Example number one:</u> "Mr. A., an electrical engineer who is married and has one child, has been working for a large electronics corporation since graduating from college five years ago. He is assured of a lifetime job with a modest, though adequate, salary, and liberal pension benefits upon retirement. On the other hand, it is very unlikely that his salary will increase much before he retires. While attending a convention, Mr. A is offered a job with a small, newly founded company with a highly uncertain future. The new job would pay more to start and would offer the possibility of a share in the ownership if the company survived the competition of the larger firms.

Imagine that you are advising Mr. A. Listed below are several probabilities or odds of the new company's proving financially sound. Please check the lowest probability that you would consider acceptable to make it worthwhile for Mr. A to take the new job.

- The chances are 1 in 10 that the company will prove financially sound.
- The chances are 3 in 10 that the company will prove financially sound.
- The chances are 5 in 10 that the company will prove financially sound.
- The chances are 7 in 10 that the company will prove financially sound.
- The chances are 9 in 10 that the company will prove financially sound.
- Place a check here if you think Mr. A should not take the new job, no matter what the probabilities" (pp.8-9).

Example number two: Judge E. is attempting to decide on the most appropriate disposition of the case of Johnny L., the 16-year-old son of a well-liked businessman who has been a leader in the community. Johnny has been convicted of breaking and entering a local store along with an older boy. This is Johnny's first major offense but he has been in a great deal of trouble of a less serious nature. The Judge has only two options. He can send the boy to reform school for a year or he can put him on probation. The judge has reports which suggest that Johnny may well continue following the example of some older boys with bad records if he is put on probation. Johnny has been seen with these boys for several years and over this period has seemed to become more and more closely involved with their activities. On the other hand, the only available reform school has a somewhat poor reputation, partly because the administrators seem to be not as inclined as is desirable to work for the rehabilitation of the boys who are sent to the institution. Some of the inmates, rather than reforming, turn into consistent lawbreakers.

Please put a check next to the ONE statement that best expresses your opinion.

- I strongly favor that Judge E. send Johnny to reform school for a year.
- I moderately favor that Judge E. send Johnny to reform school for a year.
- I slightly favor that Judge E. send Johnny to reform school for a year.
- I slightly favor that Judge E. put Johnny on probation.
- I moderately favor that Judge E. put Johnny on probation.
- I strongly favor that Judge E. put Johnny on probation.

The typical experiment is administered as a within-subjects-design. Each S is required to make an initial estimate of acceptable risk either by checking the appropriate line on the questionnaire or by estimating the minimum odds he would require with respect to the success of the more attractive alternative before recommending it as a course of action. Ss are then brought together for a discussion of the problems. Depending on the conditions under study, the group may or may not be required to reach consensus on the alternative to be selected. Following the discussion (and announcement of group choices, if appropriate) each subject makes estimates of acceptable risk as earlier.

The results from such studies, as described by Marquis, have two major characteristics of interest. One is a decrease in variance between mean scores on different problems as a result of group discussion. A second characteristic of interest is the extent of changes in mean scores of extreme groups ("most risky"), ("most cautious") relative to those of less extreme groups. Here, as in other studies, the major shifts in acceptable risk are associated with individuals who take the most extreme initial positions.

It is important to note in this summary of typical results that, although little importance is assigned by most (particularly early) researchers to their occurrence, shifts in the cautious direction are sometimes observed, as are occasions of zero change. Major reasons for the lack of interest in these two outcomes seem to be that they occur with very low frequency relative to changes in the risky direction, that they are only occasionally replicable, and, in the case of the shift in the direction of increased cautiousness, that they are of lower absolute magnitude.

Four explanations, of "risky shift" behavior have been proposed.

a. <u>The "Diffusion of Responsibility" Hypothesis.</u> Perhaps the earliest hypothesis advanced to account for the risky shift is that individual participants feel less personal responsibility for failure of a risky option chosen by the group than they do for failures of similar options chosen by them alone. As Marquis points out,

> "Certainly insurance companies and bookies behave in this fashion, ′laying off′ part of a large transaction with cooperating operators. And children playing ′follow-the-

leader′ take greater risks in the group than they would individually" (p.16).*

Proponents of the diffusion hypothesis consider, generally, that there are four links in the causal chain: (1) The creation of "affective bonds" between group members during and as a result of group discussion; (2) diffusion of responsibility for a possibly unsuccessful outcome, resulting in (3) a reduction in fear of possible failure, and hence, (4) an increase in acceptable level of risk.

b. <u>The "Leadership" Hypothesis.</u> A second hypothesis is that individuals to whom higher than average levels of risk are acceptable tend also to be more influential and persuasive than other individuals. As a result, their impact during group discussion is greater than that of more conservative discussants.

A recent variant of this hypothesis - "Rhetoric-of-Risk" (cf. Dion, Baron & Miller, 1970) - considers degree of persuasiveness to be an intrinsic characteristic of the <u>position</u> taken, rather than a characteristic of the person taking it. As two of its proponents suggest,

> "There are two related aspects of the risky position that may give the proponent of such a position a disproportionate weight in open discussion: (1) the ′rhetoric of risk′ is more dramatic, and (2) the conflicts and uncertainties entailed in accepting the riskier alternative might lead the

---

*Actually, the second of these "examples" is more in accord with the "Leadership" hypothesis (discussed next).

> proponent of such alternatives to state his arguments with a heightened intensity and amplitude. In short, he may have the advantage of a more potent language, more intensively produced" (Kelley & Thibaut, 1968, p.82).

This position is somewhat more parsimonious than the classic explanation since it does not require a distribution of risk preference or leadership quality across the members of the group.

c. The "Pseudo Group Effect" Hypothesis. This hypothesis holds that it is the additional time provided to each individual by virtue of the group discussion, not the discussion itself, that accounts for the shift. In Bateson's (1966) words...

> "...a person might not want to commit himself to a decision until he has had plenty of time to think about it. This should be especially true if the problem were either complex or important. With a complicated problem, the person might feel he wanted an opportunity to weigh up all the evidence. With an important problem, even a fairly simple one, he might be reluctant to make a hasty decision. In both cases he might feel inclined to give an initially cautious response, meaning 'Don't know' or 'not yet ready to decide'. After he had time to study the problem, however, his initial caution might evaporate, and the result would be an increase in riskiness" (p.121).

d. The "Risk-as-Value" Hypothesis. The final major hypothesis is that our culture assigns a positive value to the acceptance of risk and that each individual initially adopts a level that he assumes will be higher than the average for the group. When, during group discussion, others are found to be willing to assume greater risk, the levels assumed by the (relatively) conservative decision makers increase.

All of these hypotheses seem intuitively plausible. Probably most readers can recall from their own experience, situations in which there seemed to be a feeling of "I'll take a chance if you'll take a chance" - perhaps an example of "responsibility diffusion." One can imagine other situations in which group members consciously "throw in their lot" with one or more individuals who have reputations for bold, decisive and successful action. And surely there are instances in which opportunity to become more familiar with a problem - to consider the relative worths of possible outcomes and to realize that few decisions are, in fact, irrevocable (or at least, that the outcomes to which they lead, irreversible) - result in willingness to accept a decision alternative with an intrinsically higher degree of risk than one would have accepted initially. Finally, one may be able to recall occasions where a readiness to accept a level of risk greater than that acceptable to a group of one's peers was positively reinforced both by the group and by outsiders.*

A major problem with these hypotheses is that one or more may apply in some degree on some occasions and none may apply on others. Perhaps, then, the major error to be make in pursuing them scientifically is to assume that one of the four (or another

---

*It is interesting in this regard to reflect on the differences between popular opinions of Field Marshall Montgomery, who was held to be a methodical tactician with a low acceptance for risk, and of General George Patton who was seen to be impulsive and the very essence of a high stakes' gambler.

not yet specified) must be the correct one. As Davis, Laughlin and Komorita (1976) have noted,

> "...the choice shift is after all a group-individual difference, a comparison already honored by 75 years of experimentation and periodic rediscovery. It seems unlikely that all 'differences' on all tasks under all social conditions would be due to the same social process. Different social processes can underlie shifts in different contexts. The former could produce the CDP (Choice-Dilemma Problems) item's characteristic, individual distribution (often skewed in the direction of shift) while the latter acts to translate the former into the group distribution" (p.513).

A second general observation we would make here has to do with the lack of specificity of the concept of risk used in studies employing the Choice-Dilemma Problem scenarios. Although probabilities of a successful outcome are a parameter of the subjects' judgment, payoffs associated with the alternatives are always stated in qualitative terms. Thus, the expected values accruing to choices among alternatives cannot be evaluated. It seems, therefore, a moot point whether, in fact, group-inspired changes in individual behavior are in the risky direction. Conceivably they are shifting from a position of greater risk to one of lesser risk.

A final observation relates to the occurrences of cautious shifts. Prior to discovery of the risky shift phenomenon, it was commonly held that, because of the explicit and implicit conformal pressures that develop (Schachter, 1951), "...the use of teams and committees as decision-making bodies in organizations leads to rigidity, a lack of boldness and risk taking, and to the selection of conservative courses of action" (Belovicz & Finch, 1971, p.151). This was the "conventional wisdom," alluded to in the preamble to Cartwright's review, against which the allegedly counter-intuitive risky shift effects were assessed. Our question is, "Where do occurrences of cautious shifts (which, as findings to be presented seem to demonstrate, are not random events) fit in?" Do they provide support for the 'conventional wisdom' when they occur and support for the 'new wisdom' when they do not?

We turn now to the question of the generality of the findings from risky-shift research. To what extent do the results depend upon the subjects and the type of intragroup interaction, or the specifics of the decision problem.

Stoner's work was conducted with graduate students in industrial management and initially there was some question as to whether or not the particular population chosen accounted for observed shifts. Since that time, however, the shifts have been found not to be characteristic of any particular occupation, professional group or age (Kogan & Carlson, 1969; Marquis, 1962; Rim, 1965; Siegel & Zajonc, 1967) or nationality (Bateson, 1966; Lamm & Kogan, 1970).

As indicated earlier, there are a number of different forms that the group discussion aspect of a risky-shift experiment can take. The most common requires that the group reach a unanimous decision concerning lowest acceptable odds of success for the alternative it recommends. Typically, no constraints on group organization, mode or substance of discussion or time limit are imposed. Post-discussion individual estimates are then compared against the group position and against the pre-discussion estimate. Wallach, Kogan and Bem (1962) report that shifts obtained in this form remain remarkably stable over a period of from two to six weeks after conclusion of the experiment.

The second most common form requires only that subjects engage in discussion with other group members. No consensus is required and comparisons occur only with respect to pre-discussion and post-discussion scores. The typical finding here is that, though the magnitude of shifts is smaller than when consensus is required, they are still reliably positive.

Another variation eliminates face-to-face discussion completely, requiring that subjects observe or read a transcript of a discussion of the CDP's by independent discussants without contributing their own thoughts. This increase in "psychological distance" results in a shift of less magnitude than observed in the form discussed immediately above (Kogan & Wallach, 1967; Lamm, 1967).

Still another variation requires that the initial choices of subjects be made known to the group prior to the discussion. This form, employed in experiments on the "Risk-as-Value" by Bell and Jamieson (1970), Teger and Pruitt (1967) and Wallach and Kogan (1965) does not lead reliably to shifts as significant as those associated with free discussion with or without group discussion.

The final paradigm of interest is one used by Alker and Kogan (1968) and Lamm, Trommsdorff and Kogan (1970) in which group discussion of unrelated topics and choice dilemmas was substituted for discussion of the experimental problems. Here, no significant shifts were found, giving little support to the hypothesis that interpersonal interaction, per se, accounts for the risky shift effect.

Most risky shift research has been conducted using some subset of the original Kogan and Wallach choice-dilemma materials. The obvious question arises, whether there is something unique about the materials themselves that accounts for the results characteristically observed.

Efforts to check on this possibility have taken two forms. In one, the content of CDP's is altered slightly to make the problem less abstract. For example, Rabow, Fowler, Bradford, Hofeller, and Shibuya (1966) changed the names of central characters, "Mr. J" and "Mr. B," in two of the test items to "your brother" and "your father." Such modifications, designed to create "moral dilemmas," resulted in shifts in the conservative direction.

The second method substitutes an altogether different type of problem for the CDP - typically a betting paradigm in which money is won or lost (see, for example, Collins & Guetzkow, 1964). Shifts, either conservative or risky, do not typically appear in this kind of situation.

> As Cartwright (1971) points out, then, ...
> "Sometimes groups are riskier than individuals, sometimes more cautious, and sometimes there is no difference. A great deal more research is required before we can confidently generalize from the choice dilemmas to the kinds of decisions faced by natural groups" (p.374).

In summary, groups clearly perform differently than do individuals on decision problems. Moreover an individual's behavior is likely to be different when he is functioning as a member of a group than when on his own. The risky shift is one of the more significant and better researched of these group effects. (For more complete reviews of this topic, the reader is referred to Belovicz and Finch (1971), Cartwright (1971), Davis, Laughlin and Komorita (1976), Dion, Baron and Miller (1970). Not all shifts are in the direction of greater riskiness, however, and the determinants of shift direction and magnitude are complex. This type of group influence on decision making behavior seems particularly relevant to operations and is one that should be studied more thoroughly in closer-to-real-life decision situations.

Decision Making and Plans

There are various ways in which one can conceptualize the decision-making aspects of a commander's task. According to one conceptualization, he absorbs information, weighs it, uses it to evaluate possible courses of action, perhaps goes back for more information, and then, when one course of action emerges as preferable - and preferable to the gathering of more information - finally arrives at THE DECISION. Thus ends one act of the drama, and the curtain falls. The sequence is repeated for each decision that must be made.

Although this conceptualization is useful for many purposes, it is unsatisfactory in some respects. In particular, it is too discrete - too episodic. By chopping the task into a series of decisions, it fails to capture the continuous, on-going nature of the work that is really being done. An alternative way of conceptualizing the commander's task is to view it as the continual formulation, elaboration and execution of a plan. The plan, in the sense in which we use the word here, is a rather complicated cognitive structure, for four reasons:

a. It usually has conditional branches. That is, the overall plan usually contains subplans (perhaps they are even named "Plan A," "Plan B," and so on) that may be selected according to circumstances.

b. It is both a plan of action and a plan for acquiring information. For example, the plan usually includes provisions for checking now and then to make sure things are going "according to plan." To put it more generally, the commander's mental list of things to do at various points in the future includes both actions that will affect the situation (e.g., orders about the disposition of forces) and steps he will take to acquire information about the situation.

c. Psychologically - if not logically - the commander's plan for action is integrated with his picture of the situation - both the situation that now exists and the situation he wants to bring about. Indeed the plan and picture are so intertwined that they must, we expect, be treated as parts of one cognitive structure. For convenience we shall use the word "plan" to refer to the whole complex.

d. The plan is always incomplete. Not all the parts of the picture have been filled in, and not all the possible consequences of the plan of action have been thought through. An important part of the commander's work is filling in, elaborating, or otherwise modifying, those parts of the plan that turn out to be important as a situation develops.

The difference between these two views of the commander-decision maker is more a matter of attitude than of strict logic. We can perfectly well use the first view if we remember that the successive decisions that must be made are not unrelated: much of the information that is absorbed for the sake of one decision, will be relevant to another; earlier decisions constrain later ones; and the sequence of problems that arise is far from random. Studying a stream of behavior by breaking it into episodes that somehow seem susceptible to analysis is a reasonable tactic, and has proven to be a fruitful one. The considerable body of research on the way people assess and combine probabilities and utilities is at least in part a result of this approach. The difficulty is that dividing a stream into episodes tends to obscure those properties of the stream that transcend particular episodes. On the other hand, trying to view the stream as a stream has its disadvantages too: one tends to lose sight of characteristics the episodes have in common. We assume the one point of view is helpful and heuristic for some purposes and the other for others. We believe, however, that viewing the commander's task as one of developing, elaborating and following plans has some heuristic virtues that have not been fully exploited. In adopting this general view, one is prompted to focus on certain aspects of decision making that are not highlighted by the more traditional view. The role of rule-driven behavior (predetermined decisions) and the importance of pattern recognition are cases in point. Much of what passes for decision-making could equally well be described as rule-following. For example, when one of the writers made an abortive attempt to learn to play bridge, there were a number of rules he was exhorted to follow unless or until he became more expert - "length before strength", "second hand low, third hand high", and so on. These aphorisms are like canned decisions, packed away beforehand for use when needed. From the present point of view they are pre-packaged plans that relieve the cognitive strain of going through the decision making process in real time - i.e., the strain of thinking out the whole tree of possibilities - when under the stress of playing a hand. Moreover, they embody and transmit some of the wisdom of players who are presumably more expert than the beginner.

Other examples come to mind. Boy Scouts are told to "splint them where they lie," and are taught that when finding a person lying unconscious, one should recite "bleeding, breathing, poison, shock." The latter formula is particularly intriguing because one does not quite know whether to classify it as a plan of action or as a specification of the priorities for gathering information by which action will be determined. It illustrates the fact that the plan of action and the picture of the situation are so closely intertwined that it is hard to talk about either of them separately.

A military command generally has its Standard Operating Procedures. They are in part an aid to communication; they reduce the amount of communication that must flow between operators, and between the command post and other centers. But they embody predetermined decisions too; they are in part plans that have been worked out ahead of time by experts - the commander and his most knowledgeable subordinates - for use by less knowledgeable operators under the stress of action.

Much of what passes for decision-making could also be described as pattern recognition or concept identification. The week-end sailor looks at the clouds, says to himself, "It will pour rain soon," and heads for home. The motorist finds traffic creeping along unusually slowly, gets a glimpse of blue lights flashing up ahead, and turns aside to take another route. The player calling the defensive signals in a football game recognizes a passing situation and deploys his defenses accordingly. The commander of a ground force becomes aware of a particularly vulnerable point in his defense perimeter and orders reinforcements to increase the holding strength at that point. In all of these cases the course of action is obvious once the nature of the situation has been recognized, or at least recognition identifies a limited number of courses that will be considered.

Experienced planners know this - or so it seems to us. Much of their information-gathering activity is directed not toward the evaluation of particular courses of action, but to getting a better picture of the situation. This is because in many cases the actions that must be taken are predetermined, subject to the situation that prevails. The problem, in those cases, is to classify correctly the prevailing situation.

Nothing that we have said is meant to deny the importance of the kinds of activities usually associated with the term "decision-making," or of the results of experiments on those activities. We mean only to suggest that something of relevance to the design and operation of systems and to the understanding of commanders' behavior could be gained from studies directed at learning more about how people make, modify and follow plans.

A related research objective and one that would be especially germane to the problem of improving C^3 operations, could be that of exploring the feasibility of extracting rules or algorithms for situation-specific behavior from experts, or groups of experts. The production and evaluation of such rules would undoubtedly require a multistep process involving induction, and verification by application of tentative rules in simulated situations, because people may not, in general, be able to verbalize on demand the rules by which they operate.

Supposing, however, that one could get procedural information out of experts in such a form that it could be incorporated into computer programs. How might it be used? Three possibilities come to mind:

a. One might simply let the machine follow a procedure blindly. This idea is frightening in the general case, but there are particular cases in which it is not. If the decisions the procedure embodied were of little consequence in the overall scheme of things - like decisions about routes by which messages might best be sent - then one might be quite content to have them made according to a canned procedure. Another case is that in which the procedure is only a procedure for displaying to someone the information he will probably want, in the order in which he will probably want it. If the information can be obtained at little cost (perhaps because it is already stored in the computer), and if the operator can easily ignore information he does not need and ask for additional information he wants, then just letting the machine follow the display procedure blindly may be quite reasonable.

b. One might display the procedure to the operator and let him follow it. The procedure then becomes a sort of reminder - like a pilot's check-list or the "bleeding, breathing, poison, shock" in first aid. The operator can deviate from it if his better judgment tells him he should.

c. More ambitiously, one might have the machine show the prepared procedure to the operator, let him edit it if the circumstances warranted, and then have the machine execute the edited procedure. This may be very difficult to contrive in general, but there are particular cases where it can be done.

These are three possibilities that come to mind readily. We believe that this is an area in which there is much room for creativity and imagination, and in which there application could have a significant payoff.

CHAPTER 5

EFFECTS OF STRESS ON PERFORMANCE

Personnel in military command, control, and communication systems are subjected to different degrees of job-related stress as operating conditions vary. Under routine conditions, job stresses in C^3 systems are probably comparable to those of office workers in large civilian organizations. In times of emergency, however, job stresses in military C^3 systems can become overwhelming: long hours of duty without relief, constant decision making under conditions of great uncertainty and extreme risk, uncomfortable or hazardous work environments, severe emotional arousal, and so forth. Since adequate functioning of C^3 systems at times of crisis can be crucial to national security, design of C^3 systems from the perspective of minimizing stress effects on performance is an important goal.

Attainment of this goal is complicated, however, by lack of basic understanding of how stressors affect human performance, For design purposes, much of what is known about stress effects is either unhelpfully diffuse or unreasonably detailed and situation-specific. This chapter is intended to focus attention on areas in which research on stress effects may provide findings that might be applied to solutions of design problems in C^3 systems. We first present a discussion of the issues in stress research that affect applicability to C^3 systems, as a context for subsequent discussion. A limited review of portions of the stress literature relevant to C^3 systems follows; a conceptual model of the effects of stressors on performance is then presented; and a set of problem areas stated in terms of the model is offered last.

Issues Affecting Applicability of Stress Research to C^3 System Problems

One fundamental problem is that of selecting from among the abundant definitions of the term "stress" a definition that has pragmatic value in research on C^3 systems. In a medical sense, Selye (1956) claims that "stress is essentially the rate of wear and tear in the body." Gersten, Langner, Eisenberg and Orzeck (1974) claim that Cofer and Appley's (1964) definition of stress as "a state where the well being (or integrity) of an individual is endangered and he must devote all of his energies to its protection" is a typical psychological definition. Graveling and Brooke (1974) define stress as "the environmental load." The dictionary (Morris, 1970) gives a more colloquial definition of the term as a "mentally or emotionally disruptive or disquieting influence." In a similar vein, stress may be viewed as the force that drives homeostasis, the essence of being alive, or indeed, the spice of life. Objective definitions of stress may bear little resemblance to subjective definitions, since individuals who have learned strategies for coping with stressful situations may not regard them as stressful. Similarly, some situations devoid of objectively measurable stressors may be highly stressful for reasons quite remote from the immediate environment.

The term is sometimes used roughly synonymously with "arousal" in the psychological literature (Hebb, 1955; McGrath, 1960), or "activation" in the physiological literature (Duffy, 1962), especially when its effects are viewed as beneficial, but it is rarely defined with any rigor. For example, in his book "Decision and Stress" Broadbent (1971) offers no definition for the term. Welford (1974), in a chapter entitled "Stress and Performance" in a book called "Man Under Stress" also avoids defining the term.

Cameron (1974) makes a similar observation with regard to the term "fatigue," which he views merely as a "label for a generalized response to stress over a period of time, which has identifiable and measurable characteristics, but it has no explanatory value." Cameron states that "It is not legitimate to describe any change in the individual's behavior as 'due' to fatigue, since the term is no more than a general description of his personal state at the time such changes are noted." Much the same may be said of stress.

It is clear that "stress" is a term without precise meaning. Some of the major sources of difficulty in defining stress are discussed below.

Cause and Effect

Part of the problem of defining stress is that it may be seen as either a cause or an effect, or both, depending on circumstances. For example, human operators learning new skills, practicing complicated motor tasks, or making decisions at a high rate under unfavorable conditions may or may not exhibit performance decrements. If performance decrements occur, there would be a strong tendency to attribute causality to the stressful nature of circumstances surrounding the operator's task. If performance decrements do not occur, many researchers would be tempted to claim that maintenance of performance under adverse conditions produced stress (an effect), with which the operator had somehow coped. The circularity of such general conceptions of stress is apparent. In C^3 systems, however, stress is of concern primarily as a cause of performance decrements.

Non-Monotonicity of Effects

Another part of the definition problem is that the same variable may be regarded as stressful or non-stressful at various levels under various conditions. For instance, the absence of noise may lull the operator of a particular machine into drowsiness; exposure to moderate levels of noise may facilitate work in another task by maintaining a sufficient level of arousal; and excessive noise in yet another work environment may hinder task performance. Thus, performance is sometimes improved by noise and sometimes impaired (Hockey, 1970).*

The case of noise is hardly unique; Welford (1974), among others, refers to this general relationship as the "inverted-U hypothesis."

---

*Poulton is said (Green, 1976) to have concluded (on the basis of a reinterpretation of the noise effects literature) that except when noise actually masks useful acoustic cues, it rarely impedes performance.

Interaction of Stresses

Despite the undeniable fact that a multiplicity of stressors in ever-changing combinations affect most people throughout their lives, the interactions of multiple stressors are poorly understood. MacPherson (1974), discussing the condition once known as "Tropical Fatigue" in the context of thermal stress, states emphatically that "There can be no question but that stresses are additive and that heat stress adds materially to the burdens to be borne by those who live and work in a hot environment." His ensuing discussion of the causes of the vague symptoms of "Tropical Fatigue" (loss of appetite, sleep, memory, energy, and weight; diffuse pains, headache, nausea, etc.) is worth quotation for the insight it offers on combined effects of stresses.

> "We all live under some degree of physical and mental stress. We worry about our job, our families, the high cost of living and so on. The list is endless, but somehow we seem to get by, or most of us do: only the weaker brethren crack under the strain. However, when thermal discomfort is added to the general burden of stress, some of the less weak may yield and manifest symptoms. Because the weather is respectable and can be talked about--it is an acceptable opening to any polite conversation--the sufferer attributes his illness to the heat and fails to refer to other and perhaps more important causes which are socially less acceptable. Unjustifiable emphasis is thereby given to the role of heat in the production of the illness and its true nature is obscured. It would be quite wrong, however, to say that the heat did not play a part, an important even if non-specific part, in the sufferer's illness."

In the context of C^3 system research, it may thus be useful to consider not only the immediate, job related stresses, but also long term, non-job related stresses to which C^3 system personnel are subjected.

Lack of Understanding of Mechanisms

Perhaps the greatest confusion about the meaning of the term "stress" may be traced to poor understanding of the manner in which it affects people. The most commonly cited mechanisms are competition for cognitive or physical resources, and distraction of attention. Thus, in the case of motor tasks, coping with physical stressors expends energy that becomes unavailable for sustaining performance. In the case of emotional involvement or cognitive tasks, a finite stock of "psychic energy" may be depleted by stressors, and hence be unavailable for task performance.

Alternatively, stressors may be seen merely as irrelevant attentional demands that preclude full concentration on tasks of interest. In a slightly different context, Easterbrook (1959) holds that emotional arousal restricts "the range of cue utilization." Repeated interruptions that force attention to alternative chores are often cited as "explanations" for degraded performance. Welford (1974) sees in signal detection theory the basis for an explanation for stress effects; he posits that "as arousal increases, the cells of the brain become readier to fire." This, in turn, is assumed to produce a less stringent decision criterion, which increases both correct and false positive performance.

Such explanations are clearly inadequate, since they merely shift the problem of definition to another area. Terms such as "attention," "psychic energy," "cue utilization," etc. are hardly any better defined than "stress." An alternative mechanism for stress effects in the context of C^3 system performance is offered later in this chapter.

Individual Differences

There is considerable evidence that tolerance to stress effects varies greatly both within and among individuals. Within individuals, levels of extraneous stimulation that would seriously hamper performance at some times may go virtually unnoticed at others. Similarly, some people seem to thrive on stress and seek it out (daredevils, professional sports figures, test pilots, and the like), while others "fall apart" under small amounts of stress and may go to great lengths to avoid it. The degree to which physiological and personality factors predispose people to react to stress is poorly understood at present. If reliable means of discriminating among individuals on the basis of stress tolerance could be found, they could be of great interest for personnel selection in C^3 systems.

Lack of Uniformity in Measures

Despite fundamental ambiguities about the meaning and mode of influence of stressors, a bewildering variety of measures of stress may be found in the literature. Most of these dependent variables are relatively simple to measure but quite difficult to interpret. Certain investigators seem to favor certain measures, whether for reasons of convenience, tradition, or success in finding effects.

For those with a physiological orientation, dependent variables purportedly related to stress include cardiovascular activity (heart rate, blood pressure, pulse rate), respiratory activity (rate, volume, oxygen consumption), peripheral electrical activity (GSR, skin conductance, electromyograms), blood chemistry and endocrine activity (urinary and blood catecholamine levels), and a miscellany of other indicators: pupil diameter, skin temperature, potassium excretion, eosinophil counts, medical histories, and so forth.

Investigators who prefer performance measures have evolved a formidable array of ad hoc measures of stress effects in specific experimental paradigms. Such measures range from decrements in piece-rate production in highly contrived tasks, to increments in reaction time, to changes in compensatory tracking performance, to shifting emphasis on speed vs. accuracy, to critical flicker fusion frequency, and even to spiral after effects. There is a strong tendency to favor multiple measures, as in performance batteries (e.g., Bieber, Margroff, & Berkhout, 1973 or Smith et al., 1974). Rating scales, such as that of Berkun, Bialek, Kern, & Yagi, 1962 and Rahe (1974) have also found many applications.

Measures most useful in C^3 system applications are likely to be those directly related to quantitative changes in performance in particular tasks.

Uncertainty About Time Course

Although stress is a ubiquitous factor in task performance, little is known in a quantitative way about its time course. Even though it is often observed that people can tolerate extreme levels of stressors for brief periods of time without suffering obvious performance decrements, there is little basis for predicting how people react to lesser exposure to stressors for longer periods of time.

Studies of acute effects have been among the more common and perhaps more relevant to C^3 systems. One common experimental paradigm requires a single task for which base level performance can be reliably described. Changes in performance on this task are sought within minutes or hours of an experimental manipulation thought to be stressful. Cameron (1974) described another common design, an "interpolated task," which subjects must perform after exposure to prior stressful conditions, such as prolonged work on an initial task.

Studies of chronic stress are less common, but still potentially relevant to C^3 systems. A typical sort of study involves retrospective comparisons in health records (over a period of years) of similar people working in stressful and nonstressful jobs (Buzzard, 1973). Methodological criticisms of the latter sort of study are numerous.

Although these sorts of acute and chronic studies bridge time periods from minutes to years, it is unlikely that they are concerned with similar processes, and that inferences drawn from them may be readily reconciled. For example, one way of coping with short term stress is by mobilization of physiological mechanisms designed for this purpose. Over the long term, however, sustained high levels of adrenalin and noradrenalin may be injurious to health.

Furthermore, the great middle ground between hours and years is poorly explored: there is comparatively little known about stress in the intermediate term of days and weeks. Thus, for example, levels of stress that may be tolerable for fractions of a day may degrade performance unacceptably if maintained over a week. Ignorance in this area may be especially important to rectify from the viewpoint of emergency operation of C^3 systems.

Another area of little understanding is that of delayed effects and after effects of stress. It is common experience that recovery from certain stresses may require hours or days. It is not as clear whether cumulative effects of stress may occur relatively long after the application of stress has ceased. Delayed costs of coping, although of considerable intrinsic interest, are of relatively little interest for present purposes.

Adaptation

As MacPherson (1974) puts it, "application of non-lethal stress to a living organism usually results in some form of adaptation." Human beings, perhaps the most adaptable of all organisms, practice numerous means of coping with prolonged stress. Physiologically, they may be capable of acclimatization to physical stressors such as heat, cold, and hypoxia, and of acquiring immunity to disease producing agents. Behaviorally, they may be able to modify life habits in a manner that minimizes the ill effects of particular stressors. Cognitively, they can learn strategies of coping with stressors that can rapidly attenuate their effects. Socially, they may seek support from interpersonal relations within organizations or with specific individuals.

In many cases, adaptation to stress is achieved at some cost - whether in metabolic, psychological, or social terms.

Except in cases where there are obvious physical correlates of adaptation, such costs have proved extremely difficult to quantify. For example, performance is often readily maintained in the presence of aversive noise exposure if motivation to perform is sufficient. Many people forced to work under unusually noisy conditions feel that additional effort is required to maintain performance. Nonetheless, attempts to measure this additional effort in a systematic and quantitative manner have met only indifferent success (Glass & Singer, 1972).

Knowledge gained about adaptation to stress could be of considerable benefit to training of C^3 system personnel for performance in crisis.

Interaction with Motivation

Individual reactions to stressful situations are strongly conditioned not only by differential physiological tolerances and experience in coping with stress, but also by differential motivational sets. Willingness (if not ability) to endure discomfort or suffering to achieve goals perceived as valuable can often obscure even gross effects of stress. For example, Chiles (1955) inflicted days of continuous performance without rest upon test subjects in a Link trainer, after which they were asked to perform a secondary task on a pursuit rotor-like apparatus. As Cameron (1974) notes of this study,

"After periods of as long as 56 hours, and in a condition of such physical weakness that some had to be carried bodily from the trainer to the testing apparatus, subjects performed well within normal limits on the (secondary) task. To attribute this result to motivation appears something of an understatement; the performance of the subjects was little short of heroic."

If test subjects can be incited to such unstinting performance in a laboratory setting of little personal significance, it is little wonder that anecdotes of superhuman performance in true emergencies abound. In the short run, it seems that few stresses exert effects as prominent in highly risky, potentially emotionally charged crisis situations as might occur in C^3 systems.

Stress Findings Relevant to C^3 Systems

Research that could be reasonably termed stress-related has been undertaken over an impressive range of physical, physiological, psychological and social contexts, and with an equally catholic selection of stressful agents. Very little of this vast body of research is directly applicable to prediction of performance in C^3 systems. For example, there is not even a standard taxonomy of stressors around which a critical review of the literature pertinent to C^3 systems can be organized.

For current purposes, stressors are classified as physical, physiological, psychological, and social. The physical stressors include the thermal (heat, cold, humidity), mechanical (vibration, acceleration, fluid pressure), and sensory (noise, glare, odor, deprivation, etc.). Additionally, ingested or inhaled substances may be classified as physical stressors, including drugs, toxic substances and gasses (such as noxious fumes, oxygen deprivation, poisons, and incapacitating agents).

Physiological stressors include muscoloskeletal fatigue (from both hypokenesia and demanding physical workloads), sleep deprivation (drowsiness, jet lag, circadian rhythms), age, disease and illness. Psychological stressors include the cognitive (information under/overload and perceptual load) and the emotional (fear, anxiety, insecurity, frustration, etc.). Social stressors include occupational factors (such as career pressures), organizational structures, major life events, crowding, solitude, and so forth.

Physical Stressors

Research on thermal stress has focused on extremes of heat (Wyndham, 1974; Shvartz et al., 1974) and cold (Lockhart, 1968; Anderson, Hellstrom, & Eide, 1968) not likely to be encountered in the typical C^3 system operation. For example, a typical study of the effects of exposure to cold on alertness (Poulton, Hitchings, & Brooke, 1965) produced a slight relationship between decreased body temperature and response latency in a vigilance task. Poulton et al. had to expose lookouts on the open bridge of a ship at sea in winter to extreme temperatures for half an hour to achieve this finding. The implications of such findings for office work, or person-computer interactions in sheltered surroundings, are obviously minimal. Furthermore, much heat stress research has had as its goal establishment of criteria for permissible physical (rather than mental) workloads. Probably few C^3 system tasks are highly demanding physically.

Similarly, work on vibration and acceleration has concentrated on maximal allowable g forces for safety or performance considerations in aerospace vehicles (Griffin, 1973; Jacobson, Hyatt, & Sandler 1974). The levels of accelerations to which test subjects, infra-human primates, and cadavers have been subjected in centrifuges, aircraft, rocket sleds, and drop towers far exceed those likely to be encountered in most C^3 system operations.

There has been considerable research as well on motion sickness and disorientation (particularly under water). Of the experiments focused on vestigular effects (Kennedy, 1972) stimulus conditions have ranged from ships in heavy seas, to aerospace vehicles in zero-g flight regimes, to centrifuges and motion simulators, to motion picture films taken from moving ground vehicles. Much is known about conditions that give rise to motion sickness, and, in general, ability to perform while suffering motion sickness is clearly impaired, perhaps beyond the compensatory ability of strong motivation (Kennedy, 1976).

The literature on noise effects on performance is voluminous but inconclusive. Apart from speech interference and impairment of hearing following long term exposure to high noise levels, other putative effects are arguable. Dosage-effect relationships in particular are not yet known in any detail. It is not necessary to cite particular studies to illustrate contradictory conclusions: whole books (e.g., Glass & Singer, 1972) have been written summarizing the myriad conditions under which performance effects may or may not be found. It seems unlikely, however, that performance decrements would be significant in noise environments typical of most C^3 system operations. In many military high noise environments, noise *per se* is often little more than a minor discomfort compared to more immediate physical and psychological stressors.

Early studies on sensory deprivation tended to concentrate on extreme measures such as darkened, quiet, body temperature water baths. Perceptual disruptions of various sorts may occur within hours under such conditions. Later studies have investigated effects of prolonged exposure to homogeneous, monotonous or "meaningless" (random or informationless) stimulation. Prolonged sensory or perceptual deprivation can hardly be considered as major stressors in C^3 systems, however, since such tasks inherently entail communication and execution of action. Nonetheless, as Zubeck (1973) points out, social isolation and reduction of variety and meaningfulness of stimulation may affect people stationed in remote locations on long duty shifts.

Physiological Stressors

Little can be said in a quantitative manner about the effects of aging on tolerance to stress. Apart from the obvious physical infirmities associated with age, there is little detailed knowledge of mental mechanisms that might make the aged less able to cope with stress. Thus, platitudes are common in discussions of the effects of age on tolerance to stress. Poulton (1972b), for instance, notes that "old people should not be made to rush," and that "a healthy old person is likely to be at less of a disadvantage than a diseased old person." Similar comments could be made about the advantages of robust health, abstinence from drug abuse, and a good night's sleep for sustaining performance in C^3 systems.

The sleep deprivation literature, although extensive, is not as readily interpretable for current purposes as might be expected. Foulkes (1966), for example, claims that "loss of efficiency in mental and physical functioning, irritability, and tendencies toward perceptual distortion and ideational confusion" accompany prolonged sleep deprivation. While it is certainly possible to observe bizarre changes in personality and behavior in individuals deprived of sleep for scores of hours (Kleitman, 1963), such extreme effects could hardly be expected to occur in the normal course of events during C^3 system operation.

The more interesting area of sleep deprivation effects is that of chronic deprivation of relatively small amounts of sleep. Such effects may be more relevant to emergency operation of C^3 systems over a period of weeks. At present, it has not been possible to document dramatic effects of this type of sleep deprivation on task performance.

Psychological Stressors

Poulton's (1972a) discussion of emotional stressors (primarily personal threats such as failure, culpability, injury, or death) documents mixed effects. Experimentation on emotional stress can be highly controversial, since stress manipulations often involve deception of test subjects. Experiments conducted by Berkun et al., (1962) produced stress by telling Army personnel the aircraft in which they were riding was about to crash; that their lives were threatened by misaimed artillery shells, forest fires, and radioactivity; that errors supposedly committed by them had resulted in injury to others, and so on.

The results as well as the ethics of such research are ambiguous.* Excessive levels of arousal often produce performance decrements, but do not affect all people uniformly. Mild threats, in fact, may raise efficiency. Extreme threats may produce few discernable results. Without a reliable metric for emotional stress, however, it seems fruitless at present to attempt predictions of what levels of emotional arousal will impair or improve performance in C^3 systems.

A number of measures to alleviate the effects of psychological stressors have been studied. The most common concern has been the stress attributed to boredom, information underload, or tedium, as typified in a vigilance task. Among the measures studied have been stimulant drugs such as benzedrine (McGrath, Harabedian, & Buckner 1959), placebos (O'Hanlon, Schmidt, & Baker, 1964), electric shock (Sage & Bennett, 1973), artificial signals (Poulton, 1972b), incentives (Davies & Jones, 1975), and knowledge of results (Warm, & Epps, 1974). Other potential means of alleviating information underload are scheduling (including shift length, rest and exercise periods), and task restructuring.

By and large, it seems that stimulants and placebos can, in the short run at least, aid performance in tedious (vigilance) tasks. Electric shock and noise exposure can also aid

---

*It is difficult to obtain permission to conduct research that requires such gross deception of test subjects in the United States (including the military) today. Furthermore, it is not at all clear that this sort of research would yield information unobtainable by other means, such as field observation of groups exposed to extreme stresses for extra-scientific reasons.

performance, possibly by maintaining a suitable arousal level. Knowledge of results and incentives have also been shown to be helpful in highly structured situations, as has the use of artificial signals. The use of such measures in high rate of performance, information overload tasks in C^3 systems is unlikely to produce a meaningful decrease in stress effects, however.

Furthermore, although it may be possible in the short run to alleviate stress in C^3 systems by some of the above methods, all of them have associated disadvantages as well. Stimulant drugs may become addictive, placebos presumably cannot work forever, electric shocks are aversive, too many artificial signals may be tiresome, meaningful incentives may be difficult to provide, scheduling may be limited by availability of personnel, and so forth.

Social Stressors

Life events that have been considered stressful include not only those generally viewed as unfortunate (such as disease, death, divorce, losses of jobs and property, and social upheavals) but also those usually considered happy (marriage, birth of children, promotion, etc.). Few of these events are immediately relevant to task performance in C^3 systems, but all of them affect the individuals who work in these systems to some extent. There is some reason to believe (Caplan, 1976) that organizations may be structured to minimize the effects of social stressors on individuals. There has been no systematic study as yet of how appropriate such structures may be in C^3 systems.

Conceptual Model of Stress Effects on Performance

A common problem encountered in attempts to apply research findings from numerous studies conducted for different purposes and from different perspectives is the absence of a unifying framework of thought. Models proposed for limited purposes seldom account for findings of all studies conducted for similar purposes, let alone for those of studies conducted for unrelated purposes. A common framework would be especially useful in the present effort to utilize findings of physical, physiological, psychological, and social studies as sources of concrete solutions to design problems in C^3 systems.

A simple conceptual model of stress effects on performance is thus suggested here as a guide to further research applications in C^3 systems. This "reserve capacity" model is intentionally general, but nonetheless embodies principles applicable to design of specific studies of stress effects on performance. It is most directly presented in an hydraulic analogy, illustrated in Figure 5.

The fluid which flows through the reservoir in Figure 5 is intended to represent the ability to sustain task related work in C^3 systems. The work may be either mental or physical; it is assumed only that a finite supply is available to be expended on work of all sorts. In particular, coping with stressors is viewed as one form of work. The fluid is assumed to have multiple sources, shown as tanks representing physiological, psychological, and social origins. The fluid is collected in a reservoir, which serves as a buffer of reserve capacity. The reservoir is drained by the necessity of supporting performance in multiple tasks.

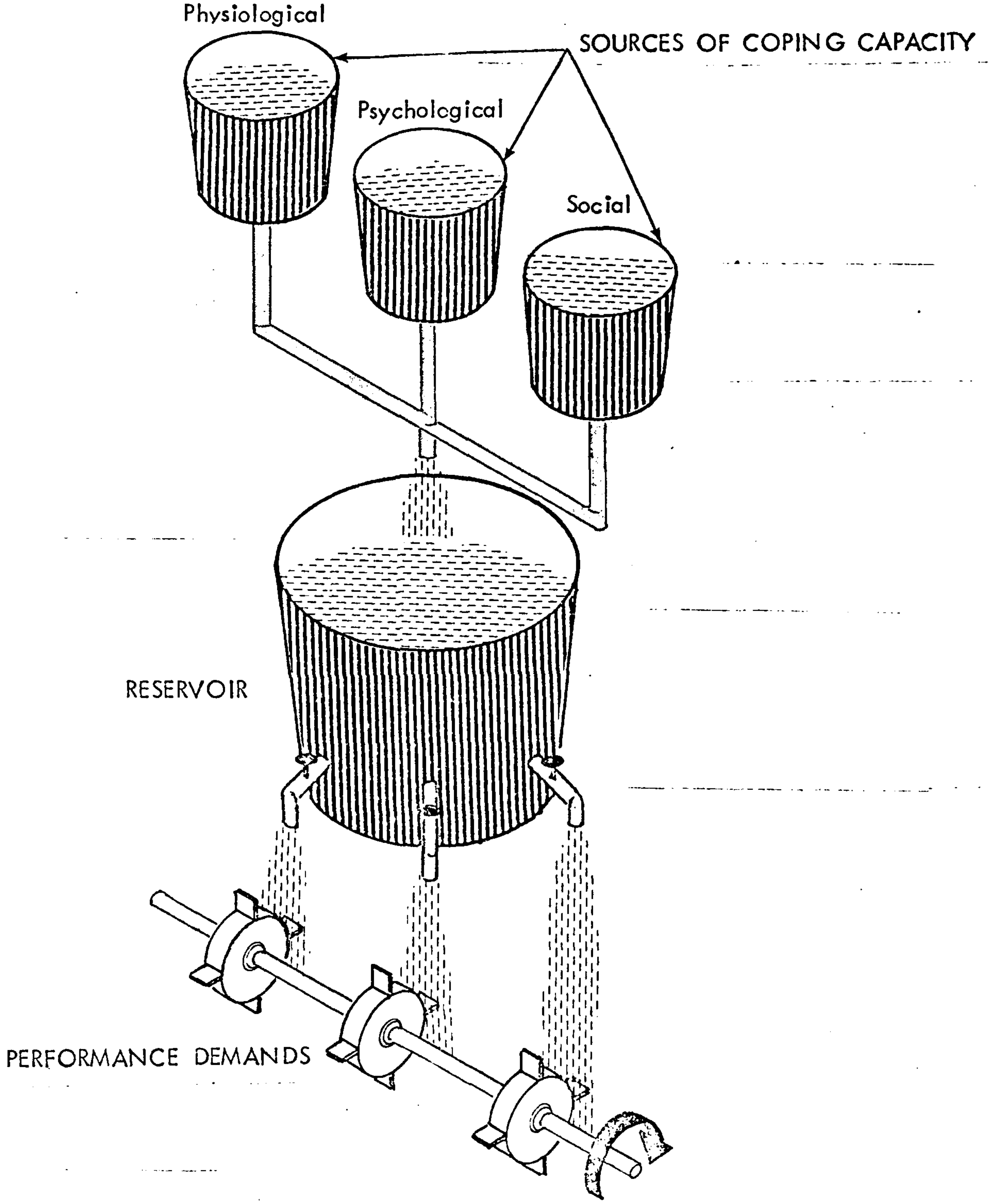


Fig. 5 HYDRAULIC ANALOGY FOR RESERVE CAPACITY MODEL

The key element of this reserve capacity model is its dynamics: performance can be maintained only as long as the rate of replenishment of the work supporting fluid can maintain the fluid level in the reservoir against various drains. The model strongly emphasizes that stress effects are not static over time, and that variability in human performance under stress is to be expected.

The reserve capacity model is consistent with most major findings discussed in previous sections, and affords parsimonious mechanisms for them as well. Specifically, the model does the following:

1) It provides a mechanism by which great amounts of stress are tolerable for short periods of time; high outflow rates can be achieved by draining the reservoir more rapidly than it can be replenished.

2) It is consistent with the finding that stress effects are additive; numerous small drains draw on the common reservoir as effectively as fewer large drains.

3) It can account for the "inverted-U hypothesis" of non-monotonicity of stress effects on performance; small amounts of stress can be accommodated by discharging the reservoir at a rate equal to the replenishment rate, but large amounts of stress may require discharge rates higher than those sustainable from the buffer or its sources.

4) It offers an alternative to the mere passage of time as an explanation for recovery from stress; the reservoir must be

replenished before it can be drawn upon at a high rate after it is emptied.

5) It can reconcile the "competition for resources" and "irrelevant attentional demands" explanations for performance decrements attributed to stress; both explanations can be cast in terms of multiple drains on the reservoir.

6) It accommodates individual differences in coping abilities; these may be attributed to varying reservoir capacities.

The general model lends itself readily to elaboration for particular purposes. For example, pumps, valves and plumbing can be added or rerouted, and the entire model can be expressed in a few differential equations. Mathematical simulations of model performance may be compared with performance data over a wide variety of research designs, ranging from vigilance decrement to information overload studies. Even in a non-quantitative way, however, the reserve capacity model may be useful as a source of hypotheses for applied research. Issues suggested by the reserve capacity model may be classified as relating to sources of capacity, buffering provided by the reservoir, and rates of expenditure.

Issues Related to Sources

Questions about the sources of reserve capacity include the following: What proportions of total capacity are generated by physiological, psychological, and social sources? Do production rates of the three sources interact? How is the physiological source affected by age, conditioning, and acclimatization? How is the psychological source affected by learning, motivation, and emotion? How is the social source affected by interpersonal relations and organizational structure? How precisely may rates of production of reserve capacity be measured or estimated? Are these rates related to measurable physiological, personality, or cognitive traits? Can individuals with unusually high or low production rates be identified? Can factors affecting replenishment rates be identified?

Issues Related to Buffering

Questions about the dynamics of buffering provided by the reservoir include the following: Can the size of the reservoir be measured by means independent of specific performance decrements? How much does the reservoir vary in size within and among individuals? Can its size be increased through training, exercise, or conditioning? Is reservoir size predictable from any other measurable individual traits?

Issues Related to Rates of Expenditure

Questions about rates of expenditure of reserve capacity include the following: Can C^3 system tasks be ordered by the rates of expenditure of reserve capacity needed to sustain performance? How effective are various measures for minimizing performance drains in the reservoir? Can optimal rates of expenditure be identified for different time spans?

Summary

The term stress, as applied to C^3 system performance, may be viewed as a general phenomenon, albeit a poorly defined one, that has many potential manifestations. Variability within and among individuals in tolerance to stress effects is considerable, and may be attributed to a large number of causes. Theoretical understanding of coping behavior and factors that mitigate and exacerbate stress induced performance decrements is poor. Systematic documentation of stress effects on performance in C^3 systems probably requires a common conceptual framework. One such framework, a "reserve capacity" model, is suggested here. Models such as the present one may be used as a source of testable hypotheses for applied research that may eventually provide practical solutions to stress problems in C^3 systems.

CHAPTER 6

PERSON-COMPUTER INTERACTION

Developments over the past two or three decades in computer and communications technology have far reaching implications for military systems, and for plans and programs relating to national defense. Michaelis (1975) has claimed that "within the past century only the introduction of the airplane, the submarine and nuclear weapons have provided similar challenges" (p. 28). The full potential of these developments will not be realized, however, without considerable research and development effort.

One of the major aspects of the problem of realizing the potential of computer technology is that of determining how to design systems that are optimally matched to their human users. Although the problem of "interfacing" people with machines is one that human factors scientists and engineers have concerned themselves with for many years, the research results that find their way into Human Engineering design guides has only limited applicability to the problem of computer system design. This is because the computer is a qualitatively different type of machine from most that people have used in the past. We usually think of machines as devices for changing energy from one form to another, and for accomplishing work in the process. In the case of computing machines energy transformation and work are incidental. The computer is designed to transform information structures not energy. It manipulates symbols not forces. In its behavior it is more analogous to a person thinking than to a person doing physical work. Consequently, while established human factors design principles apply to computer systems to the extent that these systems are similar to other systems that people use (e.g. principles pertaining to the layout of work spaces, the design of displays, etc.) there are few validated human factors principles that pertain to those aspects of computer systems that make them unique. Sime (1976) made a similar point by contending that the development of the digital computer has established the need for new discipline, which he has termed "cognitive ergonomics." Traditional ergonomics, he points out, has emphasized the need to design machines so as to be compatible with the perceptual and motor capabilities and limitations of the human being. However, the types of problems that must be solved in order to maximize the efficiency and effectiveness of the interaction between people and computers tend to be more cognitive in nature than has been true in the case of other machines.

There are, we believe, three basic needs that relate to the use of human factors principles in the design of interactive computer systems. The first need is to pay greater attention to those established human factors principles that do apply to person-computer systems as well as other systems. It is apparent even to the casual observer that basic human principles are often violated in the designs of person-computer systems. Whether the designers have wittingly ignored these human factors guidelines or simply have not known of their existence is seldom clear. The effect, of course, is the same in either case. The information to which we refer is readily accessible in codified form in several human engineering handbooks and design guides (VanCott & Kinkade, 1972; Woodson & Conover, 1964).

The second need is to do the research that is necessary to discover new design principles, and in particular those that relate to the more cognitive aspects of the operation of person-computer systems. A fundamental part of this problem is that of articulating the research issues that need to be addressed.

The third need, as we see it, is that of attempting to organize the knowledge that has been gained as a result of developing and operating many systems. While there has not been a concerted effort to develop a discipline such as cognitive ergonomics, many systems designers and developers have been sensitive to the need to design their systems with the user in mind. Consequently, they have often incorporated features in their systems for the express purpose of facilitating the use of those systems. While the effectiveness of these efforts have not been evaluated rigorously in most cases, and while the "knowledge" is often strictly opinion, there is, we believe, much to be gained by attempting to document the efforts and organize the folklore. The experience of system developers should be a rich source of insights and ideas that would not only be useful as tentative guidelines for designers of future systems but would represent points of departure for research.

The need for more attention to be given to human factors considerations in the design and operation of computer systems has been noted by numerous writers (Dunn, 1976; Kriloff, 1976; Nickerson, 1969; Shackle & Shipley, 1970). The fact that people have been able to use systems that have been developed in the absence of explicit guidelines is not inconsistent with this assertion. As Sime (1976) has pointed out, the need for good ergonomic design is often obscured by man's ability to adapt to poorly designed systems. This is a particularly insidious problem in the case of computer systems, he notes, "for man can bend his mind in more directions than he can bend his arm, and can adjust his cognitive behavior to cope with the most perverse of systems" (p. 1). Coping with the system is not necessarily using it to maximum advantage, of course. Moreover, as the community of potential users of computer systems becomes increasingly heterogeneous and includes more and more people who are not trained in computer science or computer-related disciplines, the need to optimize the designs of these systems from a human factors point of view also increases.

Although the need for such work has been recognized for some time, only in the very recent past have efforts been made in this direction. As recently as 1970 DeGreene (1970) could make the statement that "Nothing has been written on the human engineering of computer systems" (p. 281). DeGreene's statement was incorrect; several papers dealing with the subject had been published by 1970. Few of them reported empirical data of the type from which a discipline such as cognitive ergonomics could be developed, however. Since 1970, several empirical studies have been conducted on various aspects of person-computer systems and operations. As Sackman (1976) has recently pointed out, however, the experimental literature to date remains fragmented and disconnected, and largely void of integrating concepts and theories. Several review papers have been prepared (Frederiksen, 1975; Kriloff, 1976; Licklider, 1968; Rouse, 1975, Shackel & Shipley, 1970;); but these papers, for the most part, have been narrative summaries of work that has been done. No one has yet made a serious attempt to provide an integrating theoretical framework for this field.

It is our opinion that an attempt to develop such a framework is probably still premature. What is needed at the moment are more attempts to articulate the types of questions that should be addressed experimentally, to develop better methodologies for doing the experimental work, and finally, to apply those methodologies to the collecting of the kind of data that can provide a solid basis from which a useful theoretical framework might be built.

A Perspective Regarding Person-Computer Systems

The usual block diagram of a computer system shows the computer itself at the center, with a variety of peripheral devices attached to it. The user is typically represented in these diagrams as an I/O device that constitutes one of the system peripherals. In this view of the world, the computer is preeminent and the user is of secondary importance. An alternative representation, and the one that we prefer, would show the user as the central and most important element. The computer would be represented as one of many tools that he uses to accomplish his objectives.

We prefer the second representation over the first because it emphasizes the importance of designing the computer system so as to insure its usefulness and its useability. It also puts the system itself in proper perspective. The tools that it provides are not the only ones on which the user relies. Moreover, unless the computer-based tools are more effective than the alternatives to which the user has access, he probably will find other ways to get his task done.

This view also calls attention to the fact that no matter how sophisticated the tools are that the user has at his disposal, the degree of success with which these tools are applied to the solution of actual problems will depend in no small measure on the user's skill in exploiting them. Carbonell (1967) has suggested that as computer technology progresses human limitations will become the major constraints in the functioning of person-computer systems. His notion is that, as programs are developed to bear more and more of the burden of computation, bookkeeping, graph plotting, and other similar chores, the product of the person-computer interaction will be determined more and more by how capable the person is of taking full advantage of these tools. The tools represent potential problem-solving capability, but it is up to the user to see that they get applied effectively to the problems that need to be solved.

Emphasizing the role of the computer as that of one among many tools, and the role of the person as a tool user, is consistent with Sime's contention that there is a need for a new discipline that might be referred to as cognitive ergonomics. Many of the problems associated with how the user interfaces with computer-based tools have to do with such factors as communication, problem formulation, decision making, information manipulation, and the like. The computer-based tools that he will be using provide him with the opportunity to extend his cognitive abilities in much the same way that other types of machines have given him the opportunity to extend his sensory and motor capabilities. How to design computer systems and computer-based tools so as to maximize the chances that this extension is realized effectively is not yet known.

This view also reminds us of a danger inherent in the use of any tool, a danger that is probably the more grave, the more sophisticated the tool involved. That is the danger of problem solvers' "bending" problems so that the tools that happen to be at their disposal can be applied to them (Stewart, 1976). Sackman (1976) refers to this danger as that of "force fitting problems into available 'procrustian' software." Such bending or force-fitting can have serious consequences. One is that the thinking of the problem solver may be constrained inasmuch as he may be inclined to think about the problem in terms of the capabilities and limitations of the tools at his disposal rather than in terms of the problem itself. A second possible consequence is that by bending his problem the problem solver may qualitatively change it so that he ends up solving a problem that he is able to solve, given the tools at his disposal, but not the problem that was originally posed for solution. The view that we propose reminds us to focus on the user and his tasks as the reasons for which the tools exist. The tools themselves are means to an end, and should be evaluated in terms of the degree to which they in fact aid the user in realizing his objectives.

Finally, this view leads us to focus on the human being's capabilities and limitations as a proper starting place for determining what kinds of tools he needs to amplify his capabilities and to compensate for his limitations. These capabilities and limitations have, of course, been the focus of attention of psychological research for many years. Preceding chapters of this report have summarized some of this research.

Modes of Computer Use

One of the reasons that the study of person-computer interaction has proven to be difficult is the fact that the nature of the interaction has been changing more or less continuously since computers first made their appearance on the scene. The following taxonomy of usage patterns is only one of several that might be used, but it preserves, roughly, the order in which different usage possibilities were developed.

Dedicated Machines

In the beginning of the computer era, there were no such things as computer systems; there were only computers. Each machine operated independently of others, and was used by a single user at any given time. Machine time was scheduled in blocks of minutes or hours, and for the duration of his scheduled block a user had the machine to himself.

Programming initially was done in machine code; although assembly languages were developed very quickly.

Programs were loaded into the machine either manually through toggle switches at the console, or by means of punched cards or punched paper tape that had probably been prepared with the use of an off-line device.

The development of other modes of interaction did not supplant the hands on use of dedicated machines, but supplemented it. Some machines, in particular minicomputers that are owned and operated by individuals or small groups for specific applications may still be utilized in this way. The programming languages, helps, and tools that are available are, of course, much different from those that had to suffice for the early users.

Closed-Shop Batch

As demands for computing resources increased, computer "centers" evolved operating procedures that were designed to make efficient use of the machine's time. Barriers were constructed between the end user and the machine itself, and the common mode of operation was to deliver a program to the center, where it would be run by an operator when time permitted, and to pick up the results of the run perhaps a day, or if one were fortunate, a few hours, later.

A noteworthy characteristic or closed-shop facilities was the utilization of operating procedures that attached a greater importance to the machine's time than to that of the individual human user. Thus the user stood in line, as it were, for his turn to make use of the machine. The process of developing satisfactorily-operating programs typically required several iterations of coding at one's desk, submitting a developing program for compilation and trial execution, and debugging at one's desk with the help of diagnostics produced by the machine when the compilation and execution were attempted.

Programs for batch operations typically were written in a compiler language such as FORTRAN or COBOL. Often, as in many business applications, when they were intended to be used repeatedly, and minimization of running time was desirable, they would be recoded in machine language to eliminate some of the inefficiencies that constituted the price of convenience of higher level languages.

Remote-Entry Batch

In a remote-entry batch operating system, several I/O terminals, often small computers, communicate with a large central computer. An early example of such an operation was the Triangle University Computer Center. In this case three universities, Duke University, North Carolina University, and the University of North Carolina, shared one large (IBM 360/75) centrally-located facility. Each of the campuses had its own computing center on which was located a much smaller machine (e.g., 360/20, 360/30). These smaller centers transmitted programs to the central facility for compilation and execution and produced the compilation listing and results of program execution on output devices at their locations.

To the user, this type of remote entry batch operation is practically indistinguishable from what we referred to above as closed-shop batch. He still brings his program to a computer center, has it run by an operator, and picks up the results at a later time. The fact that the actual computing is done at another site is, from his point of view, irrelevant.

On-Line Time Sharing

The development of time sharing in the early 1960's made possible a qualitatively new mode of computer use. By allowing multiple users to have apparently simultaneous access to a single large computing system, each user can be made to feel that he has the machine's full resources at his disposal. This mode has been referred to by a variety of terms, among them "on-line", "immediate-access," "interactive" and "conversational." In this mode, there typically is a two-way exchange of commands, requests, and information that takes place on a time scale that is roughly comparable to that of a person-to-person conversation.

One of the major advantages of interactive systems, as they have been implemented, is the fact that they deliver computing power to where the user is instead of forcing him to come to where the computer is. The user typically works with a terminal that is located perhaps in his office, perhaps in his home, but in any case, not necessarily in the computer center. Portable terminals equipped with acoustic couplers make it possible for the user to obtain access to the computer via telephone lines from essentially any place where he has access to a phone.

When time sharing was initially introduced, its major significance was the fact that it made computing facilities more accessible to programmers, because of the increasingly sophisticated software that has been developed over the years, however, time sharing systems now often effectively place huge computing resources and data bases in the hands of individuals with little or no knowledge of how a computer works or how to write a program.

Stewart (1976) has criticized the tendency of some researchers who have attempted to identify the relative merits of on-line and batch-processing systems to split hairs in their definitions. He views as of little practical significance the distinction between "instant" remote batch-processing and on-line processing, both of which permit the user to interact with the computer through a remote terminal, and provide relatively quick turnaround. What matters, he contends, is whether the user can communicate with the computer through a remote terminal and, either immediately or within a few moments, obtain the results of typing a few commands, or whether the turnaround requires several hours or days and involves the utilization of other media, such as punched cards and line-printer output. In our view, Stewart's point is well taken. Splitting hairs as to whether a system should be considered "on-line" or "instant batch" probably serves no useful purpose. The point is to distinguish those operational characteristics that make for efficient and effective person-computer interaction from those that do not. A similar observation can be made with respect to the question of whether a particular interaction should be considered conversational in character. What is of greatest interest is not whether the interaction is conversational, but whether it is productive and satisfying (Nickerson, 1976).

Some Interactive Uses of Computers Relating to C^3 Operations

The applications of interactive systems to C^3 problems are numerous. We mention here only a few by way of examples. We have picked these examples specifically to represent the range of tasks that can be facilitated by computer involvement.

Modeling

Kroger and Christie (1963) point out that limited war situations lend themselves less well to control by choosing among preset moves than do those posed by strategic strike situations in a central war. The need for flexibility on the part of commanders is particularly great in the former case. Kroger and Christie express concern that the pressure to introduce automated information processing procedures in the interest of speed and volume could have deleterious effects on staff flexibility. What is needed rather than the ability to invoke preset procedures is the ability to explore the consequences of alternative courses of action before selecting among those alternatives. Computer models can provide this type of capability.

The use of computer modeling in a command and control situation is illustrated by a model that is used interactively to resolve conflicts in air traffic control situations (Whitfield, 1976; Whitfield & Stammers, 1976). In this case, the computer model plots present and future positions of aircraft in a particular airspace, and thereby detects potential conflicts. When such potential conflicts are detected, the computer informs the controller of the fact. The controller then may modify flight paths or change flight path parameters and test the consequences of these modifications or changes by exercising the model under the new conditions. The computer's tasks include projecting the paths of individual aircraft and predicting possible conflicts. The computer and controller jointly have the task of devising a solution to a potential conflict. It is the controller's job to see that the solution that is devised is implemented.

An example of a modeling system developed expressly for use by military command personnel is the Quick-Reacting General War Gaming System (QUICK). QUICK is a software system programmed in FORTRAN originally for the NMCSSC CDC 3800, and now being modified for the Honeywell Information System (HIS) 6080, in keeping with the WWMCCS standardization policy. On the basis of input data and user-supplied parameters, QUICK produces global strategic nuclear war plans, simulates the planned events and provides statistical summaries of the expected effects of these events. The system is composed of five major subsystems (data assembly, weapon/target identification, weapon allocation, sortie generation, and simulation), which are executed in series. The system can be run automatically through all steps; alternatively, component programs can be used in interactive mode from remote terminals (Blackledge, 1974). The program is maintained by the Strategic Forces Analysis Support Project at the NMCSSC and is used primarily for the support of the Studies. Analysis and Gaming Agency (SAGA), an element of the JCS.

The remote-access, interactive capability is a major component of the effort to modify QUICK for operation on the HIS 6080. The intent in providing this capability is to:

"a. Enable the nonprogrammer user to update and otherwise manipulate the QUICK data base; i.e., add, delete, or alter data as required through available remote-access devices.

b. Provide for a full batch-mode job submission capability for the execution of QUICK programs from a remote terminal, and to enable the user to monitor the status of his job, and direct the output mode.

c. Provide for the selective investigation of entire program outputs, or segments thereof, from a remote terminal, and provide for user control of the output mode" (Dambro, Page & Pelliciotto, 1974).

The remote-access user interacts with the QUICK system through a set of time-sharing programs which permit him to create and edit files, submit and run batch jobs, determine the status of batch jobs, scan files, obtain hard copy output, etc. He may work with a visual-display terminal and has available both temporary and permanent disk file storage. The typical mode of operation when updating the data base involves requesting a file from the permanent storage area, thus bringing it into the temporary storage area, modifying the file via other system software packages, and then returning it to the permanent storage area where it is available for utilization in a QUICK execution. The intent is that the user be required only to be aware of the availability of the files that he wants to use and of the operation of the time-sharing subsystems available to him, and not of the interaction of the various systems and subsystems. Examples of user-computer protocols for several of these subsystems are given in Dambro et al. (1974).

In general, the purpose of such modeling facilities is to permit the user to explore the implications of various actions that he might take in a specified situation. The ability of the model to make these implications explicit is limited, of course, by the accuracy and precision with which the model reflects the situation that is being modeled, and the validity with which it represents the interdependencies among the various factors that will determine the future course of events. The importance of these limitations on the fidelity of a model must not be underestimated. In a real sense every set of battlefield conditions is unique. Any model can only be an abstraction of that set of conditions and therefore is likely to be inaccurate in one or more respects. If a commander is to make use of such models, even for the exploration of alternatives, he must have a complete understanding of both those aspects of the model that are accurate and those that are not. To build a model of any degree of sophistication which can make these properties explicit probably requires extensive use of heuristics of intelligence that are only beginning to be understood. The utility of such models also is limited by the skill with which the users exercise them.

On-Demand Information Retrieval

One of the primary uses of computers has been the storage and retrieval of information. The most common way, and until recently the only way, that the information that is stored in a computer data base was distributed to users was by means of standardized, formatted reports. Managers in business operations, and personnel in various positions of responsibility in military situations, would periodically receive computer outputs containing a variety of information that is presumably relevant to the performance of their jobs. Such standard formatted managerial reports from batch-processing operations are bound to be inefficient because the assumption underlying their production is that all recipients need the same information, and this is not usually the case. A common complaint among recipients of such reports has been that the computer outputs they typically receive often contain much information that is irrelevant and of little use to them; further, the relevant information they receive is difficult to locate among the irrelevances, and is often presented in a format that makes it difficult to use (Eason, 1976).

An alternative to standard formatted reports that is being provided by managers with increasing frequency is a system that collects and stores data and provides the user with a means of interrogating the data base and retrieving from it only that information for which he has use. There is some evidence that this type of system is fast becoming preferred over formatted report procedures (Eason, 1976; Hedberg, 1970; Morton, 1967).

This approach has obvious advantages and will undoubtedly become increasingly widely used. It is not without its problems and pitfalls, however, First, it works only to the extent that the user knows what information will be useful to him and how to interact with the system well enough to obtain it. That is, such systems make demands on the user that automatic information dissemination procedures do not. Second, as has been pointed out by Ivergaard (1976), a risk that is inherent in any demand-driven information distribution system is that in exercising their ability to choose what information gets to them, the users of such systems could become increasingly conservative in their views. This danger is probably more real in general-purpose information systems than in problem-oriented systems, but it should not be overlooked even in the latter case. The evidence is fairly compelling that people often seek information to corroborate existing views rather than collecting it in an unbiased fashion. Even better than a demand-driven information distribution system would be a system that had sufficient intelligence so that when given an hypothesis by the user it could deliver information relevant to that hypothesis, irrespective of whether the information tended to corroborate or to disprove the hypothesis.

Document Preparation and Message Handling

Tools designed to facilitate the preparation of documents and the editing of text are among those that promise to be most widely used by nonprogrammers. The antecedents of these tools were the online program-editing capabilities incorporated in interactive programming languages designed to run on time-sharing systems. Several text-editing programs are currently in operation (Callahan & Grace, 1967; Englebart & English, 1968; Magnuson, 1966). Each of these programs provides for roughly the same basic set of text manipulation operations: copying, deleting, inserting character strings. They differ greatly, however, with respect to what operations they permit beyond this basic set, and the details of the command language in terms of which the operations must be specified. Considerable work needs to be done to determine what would constitute an optimal design of an interactive document preparation system.

Message preparation and distribution are also functions that are becoming increasingly widely used. Several computer-mediated communication systems are currently being developed by the Advanced Research Projects Agency. Each of these systems provides the user with tools to compose, edit, store, retrieve, copy, forward, modify, and otherwise operate on messages. This type of computer use seems likely to grow rapidly in both the commercial and military worlds. There are a number of human factors issues that remain unresolved regarding how best to design such systems. However, work is currently underway that should provide answers for many of these questions.

Decision Aiding

Use of interactive computer systems to aid decision makers in various aspects of decision-making tasks has received considerable attention from researchers.

Among the primary roles of the human being in the area of decision making are those of goal setting and value specification. While the computer can be given many decision-making tasks, identifying the goals toward which one should strive and specifying the values of alternative decision outcomes are uniquely human tasks. This is recognized in some existing decision-aiding systems. Among the major inputs to the QUICK system, for example, is a set of parameters reflecting the planner's views vis-a-vis the strategic objectives. These parameters are expressed, among other ways, in terms of the relative values of the targets being considered. The target values that are assigned, and the relationships among the values of various target classes, then become forcing functions for the plan that is generated by the system (Blackledge, 1974).

Interactive Situation Display Generation

In a study of potential applications of computer technology to tactical intelligence operations, Bowen and his colleagues (1971, 1975) concluded that one of the most promising applications is that of interactive situation-display generation. Situation displays are critical tools for military commanders, and, as Bowen et al. pointed out, for the task of generating such displays, the skills of people and the capabilities of computers complement each other in a particularly useful way. The human being can play the role of pattern recognizer and decision maker, while the computer plays that of an exceedingly fast clerk, data manipulator, and draftsman. A situation may be represented digitally within the computer in great detail. A situation display, then, may be thought of as a look at some aspect of the situation that is represented digitally within the computer's data base. In any complex situation, there will be many more details in the data base than can be represented effectively on any single display, so what one wants is a method of constructing a variety of displays, each of which presents to the viewer some salient aspect of the situation. One might, for example, wish to look at the deployment of sightings of all vehicles of some specified type or types within some geographical region and some specified period of time, or one might wish to look at all highways that are passable by vehicles of specified types, or one might wish to have all the bridges in some particular region shown on the map without having the display also cluttered by other types of information. Such displays are easily generated within the current state of the art. What is not yet known is how best to design a language that will permit the user to interact most effectively with the system in generating such displays.

The "User"

User Types

A frequently expressed objective of human engineering is that of designing systems with the user in mind. This seems an appropriate objective in the case of computer systems as well as in that of any other. It seems well to recognize, however, that a computer system is likely to have many different users, and not all of them will have the same needs.

Stewart (1974) distinguishes three types of users who may have different requirements even when using the same system: managers, specialists, and clerks. He makes the point that these groups will differ in terms of degree of flexibility and adaptability that they will demand of a system, the amount of effort that they are willing to expend to learn how to use the system, the degree to which they are willing to modify standard operating procedures, and the amount of information that they will need regarding details of system operation. Managers, according to several writers, are less likely than technical specialists to be willing to invest a large amount of time in learning to use computer-based tools (Damodaran, 1976; Eason, 1976; Stewart, 1976). Damodaran suggests that the managers' most common solution to this problem is to have someone else use the computer in his behalf. It seems quite possible that to the extent that managers are reluctant to make use of existing systems their reluctance may be based on the less-than-optimal design of those systems from their point of view. One can hardly blame a manager for failing to invest a large amount of time in learning to become a proficient system user if he does not have convincing evidence that such an investment would bring appreciable dividends in terms of making him more effective at his job. The assumption that managers are not among the most heavy users of computer systems should not be used as an argument that systems should therefore not be designed for their use. Better design could be the catalyst that would stimulate greater use.

Stewart (1976) defines a specialist as one who operates primarily within a particular discipline or problem area, and has at his disposal a body of specialized knowledge and a set of techniques that are applicable to the problems with which he deals. A specialist is distinguished from a manager who is likely to have a superficial understanding of many specialities, but may be expert with respect to none of them. Stewart further distinguishes three types of tasks that specialists are likely to perform with computers: (a) data input for predefined outputs, (b) instructions for data processing, and (c) high-level language programming. These task types represent a progression in terms of the degrees of freedom afforded the computer user and the amount of flexibility and creativity required of him. The data-input-for-predefined-output mode of operation, for example, requires only that the specialist provide a program with specific data according to prescribed input formats. The instructions-for-data-processing mode requires that he interact with the program, at least to the extent of selecting among a variety of processing options. The high-level-language-programming mode requires that he write his own programs. Specialists, unlike managers, Stewart contends, tend to be more than willing to devote time and effort to the learning of how to use the tools, providing that the tools represent genuinely better ways of solving the problems with which they have to deal.

Clerks seldom have the option as to whether or not to learn how to use a computer system. If their job requires that they learn, they do; if not, they do not.

Each of these user types is found within the military. There are clerks who will use whatever computer system that they are required to use because it is their job to do so. There are specialists who will be willing to invest considerable time and effort to learn how to use systems on the chance that such knowledge will make them perform better at their speciality. There are managers or commanders who will be reluctant to make such investments unless the resulting increase in their effectiveness is a surety; and even then they may sometimes prefer to use the computer through one or more intermediary.

There is another dimension in terms of which computer users may be classified, and that is the dimension of skill with respect to the system(s) they are using. Ideally, a system should be able to accommodate users representing all levels of expertise. This is not an easy objective to realize, however, because features of an operating system that would be helpful for a novice often would be unwanted by, and perhaps irritating to, an expert user. The novice tends to want much instruction and hand-holding from the system. The expert, on the other hand, wants nothing from the machine that is not essential to his work session. He wants the messages that the machine gives him to be as terse as possible; he becomes frustrated whenever he is forced to watch his terminal outputting messages that provide him with information for which he has no need.

One approach that has been taken to this problem is to develop systems that can operate in either of two modes (expert or novice). The user can then identify the level at which he wishes to operate, and the system will treat him accordingly. The fallacy with this approach is the fact that a dichotomous distinction between experts and novices is an oversimplification. Users represent all levels of expertness. Moreover, a given user may be expert with respect to some aspects of a system, while being a novice with respect to others. The challenge remains to design systems in such a way that they satisfy the needs and preferences of users at all skill levels. Nickerson and Pew (1971) have suggested several features that could be incorporated within a system to help on this problem (user-interrupt capability, brief coded computer-to-user messages backed up with full explanations in response to requests from the user, prompts and helps designed to upgrade the skill level of the novice), but a general solution to the problem is still wanting.

A closely associated problem is that of building a system in such a way that the novice naturally becomes more and more expert through normal use of it. It should not be assumed that they will invariably do so. It is possible for an individual to interact with some systems for a very long time and never discover what he needs to know to become truly expert in their use. What is required is that the sorts of helps, hints, and hand-holding that the system provides for the user who is functioning at a given level of competence will give him the information that is necessary to bring him along to increasingly higher levels.

The Commander as a User

Although, as we have noted, any computer system is likely to have many users and those users are likely to have different requirements, there is a sense in which the ultimate user of a C^3 system is an individual with command authority. His is the final responsibility for actions taken by his command, and in a sense the purpose for the resources at his disposal is to support his decision making and to effect the decisions he makes.

Morton (1962) interviewed four retired Air Force and Navy senior officers (three of four-star and one of two-star rank) in an effort to obtain a better understanding of how commanders perceive their jobs and in particular, how they relate to electronic command and control systems. The results of the interviews are reported informally and are not easy to summarize. Also, in interpreting these results it is necessary to bear in mind the small sample of opinions that they represent.

It appears from the responses of Morton's interviewees that commanders vary considerably in how they view, and perform, their jobs. There was general agreement that it is the proper function of a commander to establish policy, and that his is the final responsibility for the results obtained by his command. Commanders differ greatly, however, in the degree to which they delegate authority, and the specific ways in which they look to staff and subordinates for assistance. Functions that may be performed by staff include: the filtering and interpreting of incoming data, the generation of ideas and candidate battle plans, the critiquing and evaluation of possible courses of action, and the translation of command decisions into operational procedures.

The interviewees were agreed that the computer could be a valuable--perhaps indispensable--tool for command and control, but most were adamant about limiting its decision-making functions. "Commanders are willing to use the machine as they now use the staff--as a command adjunct whose usual function is to filter and advise and decide subsidiary details, but which does not issue orders on large issues except in the most dire of emergencies when the commander cannot be available in time to make the decision and issue the order. ...All of our interviewees made it quite plain that they believe the final measure of decision should rest with the commander and not with the computer" (p. 19). What a commander might want from a system includes: timely information displayed in a useable way, situation simulation, some screening and evaluation of alternative courses of action, assistance in the analysis of "what do you do if..." problems, and assistance in translating high-level decisions into detailed operational plans. Some of the interviewees could imagine the computer making tentative decisions, providing the commander retained the capability of overriding any decision that it made.

While all of the interviewees expressed positive attitudes toward technological innovation in general, they volunteered several caveats. Some skepticism was expressed regarding the ability of operating forces to maintain and utilize some of the equipment found in electronic command and control systems. The problem of backup was noted: after building up a dependence on a sophisticated system, what happens if the system fails and it is necessary to fall back on manual procedures? Lack of credence in system outputs was also mentioned, but this was perceived to be a problem that could be counteracted, at least in part, by providing commanders with more adequate instruction and training with respect to the systems they are expected to use. A related, but more complex problem, is that of the fear of erosion of authority. One interviewee pointed out that because commanders tend to be generalists, they may lack the specialized technical knowledge to understand fully the algorithms on which a computer-based system's operation depends, and thus feel powerless to influence the rules for system operation. "The commander of an electronic command and control system may therefore exercise considerably less autonomy over how his command is run than does the chief of a non-computerized command, because so much of what happens within the command is outside the control of the former. For example, contrast this situation to the kind of pre-programming which the commander can institute with his staff, and the responsiveness of the staff to the wishes of the commander" (p. 15). This type of problem makes the case for the development of person-computer dialogue techniques that will permit a commander to interact with a system in ways that do not require a technical training to understand what is going on.

A major implication that Morton saw in the results of his study was the need for further investigation of the importance of individual stylistic differences among commanders in how they organize their staffs and run their commands, and of the implications of such differences as exist for the design and operation of command and control systems. In particular, one would like to know, he contends, how important it is that command and control systems be flexible enough to accommodate these differences. Will systems that are adjustable to the styles of their users lead to better command decisions than those that are relatively fixed, and force all users into the same patterns of interaction? So far as we know, this question is still unanswered.

The Role of Users in System Development

Perhaps the only way to assure that a system will be both useful and useable is to have intended users intimately involved in its development from the earliest stages. To introduce users to a system only after it has been fully developed is to invite disaster. A caveat is in order, however. Often users who have worked closely with developers have been very special types of users, namely those who have a great deal of knowledge about the workings of the system at various levels. The danger is that such users will be able to interact effectively with the system whether or not it is optimally designed. Moreover, their intimate knowledge of the workings of the system could turn out to be a necessary condition of their effective utilization of its capabilities. And it is not necessarily the case that such users will be fully aware of how important their inside knowledge is. If a system is intended to be used eventually by people who are not intimately involved with the details of its design, _this_ class of users should be involved early in the developmental process. Such persons should not be expected to understand the system's software organization and bit manipulation; they should, however, have an understanding of the nature of the interaction between the user and the computer as it occurs at the level of the user terminal.

A corollary to this idea that "real" users should be involved in the early stages of a system's development is the assumption that systems that are intended to help people perform ill-defined tasks probably cannot be pre-designed; instead, they have to be evolved. And the evolution can best be guided by the insights that are gained by the attempt of operational personnel to use it in their work (Licklider, 1968; Nickerson & Stevens, 1973; Stevens, 1968).

Approaches to the Study of Person-Computer Interaction

Studies of person-computer interaction that have been done to date can be grouped, at least roughly, in terms of which of the three following methods has been employed: (1) observation, (2) simulation, and (3) controlled experimentation.

Observation

Here the approach is to take measurements and gather statistics on real systems with real users working on real problems. An advantage of the approach is the fact that objective measures often can be made unobtrusively by the system itself. This is important inasmuch as these systems tend to be very costly and typically the motivation is high to keep them productively busy.

A disadvantage is the fact that one has little or no control over the parameters whose effects one may wish to study.

Some investigators have looked for behavioral invariants and have attempted to find functional relationships that would characterize person-computer interaction in general ways. The number of observational measurements that have been made by investigators of person-computer interaction is large. They include the number of commands executed per unit time, interaction cycle time, task turnaround time, output/computer-time ratio, user and system response times, work session duration, console-time/CPU-time ratio, statement interpretation rate, overhead rate, number of user input lines per second, rate of user requests, (Bryan, 1967; Carbonell, Elkind & Nickerson, 1968; Grignetti & Miller, 1970; Grignetti, Miller, Nickerson & Pew, 1973; Scherr, 1965). Perhaps the most reliable finding to date is the tremendous amount of user and system variability. It is not uncommon to find individual differences on a given performance measure in ratios of 10:1 or 20:1 even when individuals are working in, presumably, identical situations.

Simulation

The advantage of the simulation approach is economic. Because of the great expense of large computer systems, studying the effects of manipulation of very many parameters would be prohibitively costly. If, however, one can simulate a system with respect to at least those aspects of it that are relevant to the question of interest, then one can explore the implications of various parameter changes in a more economical fashion. (Although one should not fall into the trap of assuming that all simulations are cheap!)

A good example of the use of simulation to study human performance in command and control systems is provided by the work of Topmiller, Mills, Chubb and their colleagues in the Systems Effectiveness Branch of the Aerospace Medical Divisions Human Engineering Division. They have developed a simulation facility with 5 person-computer display consoles, and have used it to study the human factors design considerations of the BUIC System, and the AWACS early warning system. They recently have implemented an RPU command and control system having four enroute controller positions and a terminal control position. Many system variables have been studied including the number of vehicles per operator, the RPU flight-control algorithm, and the frequency of actual path up-date. In each case person-machine system performance data have been collected on experienced teams of controllers, thus providing important information for future design trade-off studies.

The main difficulty in the use of simulation is that of validating the simulation model. Any predictions that are derived or conclusions that are drawn from a simulation study are only as good as the model is valid. Moreover, in order to test a model fully, one must build a system so that the predictions of the model can be checked against the behavior of the real thing. Because of this fact, the economies that are realized by simulation research are realized at some risk. That is to say, one checks the validity of one's model in certain respects, and if it passes the tests to which it is put, then one may be inclined to proceed on the assumption that the model is also valid with respect to those aspects that were not put to the test. In many cases the gains that have been realized by using simulation models have proved to be worth the risks involved. Without doubt, one can find examples in which this has not been true, as well.

Controlled Experimentation

A few investigators have attempted to run controlled experiments on person-computer interaction. This approach has the advantage that one can manipulate system parameters and study the effects in a systematic fashion. It also has its problems, however. Perhaps foremost among them is the fact that the systems of interest tend to be very large and costly, which means that unless the experimenter is in an unusual situation, he may find it difficult to have access to a system for this type of use. Moreover, often because of the fact that computer systems are so costly and the developers are vulnerable, there may be strong vested interest in obtaining a particular result. Finally, even if these sorts of problems do not exist, there is the problem of the generality of one's findings. Too often results from experiments cannot be generalized beyond the particular systems, users and situations with which they were obtained. In studying the relative merits of different programming languages, for example, one must take care that the results are not contingent upon the particular programming problems that were selected or on the specific subjects who wrote the programs. To get around the problem of programmers bringing different types of programming experience to the task, some investigators have used subjects who lack any programming experience (Adams & Cohen, 1969; Gold, 1967; Smith, 1967), but this approach has its limitations also. What proves to be best for novices will not necessarily prove to be best for experienced programmers. Similarly, what proves to be most advantageous with respect to one programming problem may not be so with respect to other problems.

Studies of User Behavior

Research on the behavior of users of computer systems is still in a very early stage. Investigators are groping for effective approaches and methodologies. There are several problems associated with doing such research: the cost of computer systems and operations which tend to make them relatively inaccessible to researchers, the difficulty of gaining adequate control of experimental situations, the tremendous variability in most performance measures that have been taken, the lack of generality of findings, and the rapid development of computer technology (which can make results obsolete before they are reported). The picture is not hopeless, however. Promising observational and experimental techniques are being developed; suggestive results are being obtained, a few of which may prove to be generally valid. Moreover, the need for research is becoming more apparent as computer facilities become accessible to an increasingly heterogeneous set of potential users.

In this section, we review several experimental investigations of user behavior. All of these studies can be considered pioneering efforts, in the sense that they represent early attempts to explore a new field of inquiry. The review is highly selective. The intent is to provide examples of studies of various aspects of user behavior, not to review the field exhaustively.

Studies of Programming

Weinberg (1971) reports the following experiment designed to investigate the influence of programming goals on performance.

Four programmers were asked to work on a problem that the experimenter estimated should take about one-fifth of each subject's time over a period of ten weeks. The problem specification was identical for all four programmers, but for two of them the last page of the instructions read:

> Your objective on this project should be to get a fully debugged program which is as efficient as possible. You may use as much core storage as needed, up to 128k. Although you are not working for fast completion of the project, you should plan to reach the final test by the end of the ninth week.

For the other two, the last page read:

> Your objective on this project should be to get a fully debugged program in as short a time as possible without considering the efficiency in speed or space of the program insofar as those factors will slow down the completion of the program. However, you should not spend any more time on the project than the normal work load allotted (one-fifth time). You must, however, keep the program size less than 128k.

The programmers who were instructed to complete the program as quickly as possible required less than half as many computer runs and took only about one-third as much programming time as did those who were instructed to produce an efficient program. The final programs of the "fast programming" pair took about ten times as long to run, however, as did those produced by the other programmers. The second experiment was done with a type of program for which "the possibilities for spectacular savings were not so great." Although the differences between groups were not so large in this case, they were in the same directions as those obtained in the first experiment.

Weinberg drew two conclusions from these results: (1) "The gains to be had from striving for efficiency depend on the type of problem as much as anything else," and (2) "a large proportion of the variance between programmers on any job can be attributed to a different conception of what is to be done" (p. 129).

Weinberg had also asked each of the programmers to estimate the time that would be required to complete the task. Apparently the estimates that the programmers made were influenced considerably by the instructions they had been given. Specifically, those who had been instructed to finish as quickly as possible gave more conservative estimates than those who had been instructed to emphasize efficiency of the program. And, unlike the other group, they performed better than their estimates. Weinberg's conclusion: "If a goal is set explicitly, there are two effects: programmers work toward that goal at the possible expense of another goal, and programmers will be far more conservative (or accurate) in estimating how well they will meet the goal. Estimates on goals not emphasized will probably be completely unreliable, both because they are not made carefully and because they are not important enough to resist being sacrificed to other goals" (p. 131).

The results of these experiments must be assessed in the light of the very small sizes of the samples used.

Studies of Programming Languages

There are two reasons why one might want to study programming languages from a human factors point of view: (1) to compare the strengths and weaknesses of existing languages for the purpose of choosing among them for specific applications, and (2) formulation of principles of language design. The first of these reasons relates to short-range objectives, the second to long-range goals. Existing languages are bound to be replaced by better ones so comparative studies have a very transitory utility. What one really wants to determine is the features of a language that make it easy to learn, easy to interpret, and powerful in application.

An approach that several investigators have taken in keeping with the long-range goal is that of studying the effects of manipulating specific language features (Gannon, 1976; Ledgard, 1971; and Sime, 1976). Ledgard (1971) invented several "miniature languages" and used them to illustrate several principles relative to the design of programming languages. Sime (1976) has reported a study in which such miniature languages were used in an empirical investigation of the effects of certain language features on program writing. Three mini-languages were compared, two of which allowed the nesting of conditional statements, and the third of which concatenated conditionals with a sequence of GOTO statements. One of the languages using nesting also printed the statements in such a way that the depth of nesting was apparent from the spatial structure of the program listing. Sime had nonprogrammers learn to write programs in these languages. He also had experienced programmers trace program outcomes, given the truth value of the predicates, or try to determine what the truth values of relevant predicates would have to be to produce specified outcomes. The results demonstrated the superiority of the language that used nesting and made the depth of nesting obvious in its spatial structure. The main conclusion drawn by Sime from this study was "that the syntax in spatial structure of programmed text should be designed to allow two-way tracing of the control structure. This means that the information required for such tracing should be overt and obvious, and not requiring undue visual search" (p. 21).

Studies of Debugging

Boehm (1973) has estimated that the process of debugging accounts for from 25% to 50% of the time required to develop a new computer program. Studies of debugging, therefore, have considerable practical value vis-a-vis the goal of finding ways to facilitate program development. Debugging is also an interesting activity as a prototype of problem solving in general. It is a bit surprising that psychologists have not made greater use of debugging tasks to study problem solving behavior.

Gould and Drongowski (1974) have studied program debugging as a paper and pencil (noninteractive) task. They planted bugs in four of IBM's FORTRAN statistical library programs and had experienced programmers try to find them under several different conditions. Two variables of interest were: (1) the type of bug that had been planted, and (2) the type of debugging aids provided to the programmers.

Three types of bugs were used: (1) array bugs (statements that referred to subscripted variables); (2) iteration bugs (DO statements); and (3) assignment bugs. All bugs were nonsyntactic, and the bugged programs were syntactically correct. Each bug was contained in a single line, and any given program had only a single bug.

Subjects were given one of several debugging aids: The program listing only (control group), the input-output of the bugged program, the I/O of the bugged program plus the I/O that would have resulted had the program run correctly, the class of bug that the program contained, or the line of code which contained the bug. The last-mentioned group was given the task of fixing the bug. The other groups had only to find it.

Programs that were used varied in length from 29 to 59 executable statements. Given four programs and three types of bugs for each program, there were 12 bugged listings. Each programmer debugged all 12 (3 within a given session). With 30 programmers (6 per group) the authors had 360 debugging sessions, which provided a reasonable sample of data.

The programmers were surprisingly fast. Several factors may have contributed to their speed: (1) They knew there was only one error and that it was nonsyntactic. (2) They worked with each program three times. The bug was different each time, but there may have been positive transfer from one session to the next anyway (the authors present data that bear out this possibility). (3) Some of the programmers had information that would considerably narrow the search. (4) There is the possibility of collusion among subjects (however, because of the next point, this seems improbably). (5) Apparently the subjects were highly motivated (the experimenters note that each of them had claimed to be an experienced FORTRAN programmer and took the task very seriously, and was highly frustrated if he could not find the bug in the alloted time of 45 minutes; reputation was apparently at stake.)

Bugs in assignment statements took three or four times as long to find as bugs in array or iteration statements. Individual differences were large, as usual, but not so devastating as in many other studies.

A surprising result was the finding that debugging aids apparently were little or no help in reducing debugging time--they were possibly a hindrance in one case. Errors were higher in the no-aid case than in others, but not significantly so. Other investigators have also reported somewhat counter-intuitive results of a similar kind. Yasukama (1974), [reported in Shneiderman & McKay (1976)], for example, found that "high level" comments on a FORTRAN program were of no help to subjects attempting to locate a bug. Shneiderman, Mayer, McKay and Heller (1975) found flowcharts to be of little or no help in a similar study.

Gould and Drongowski suggest that the programmers in all groups followed a "simple hierarchical debugging strategy." Debugging at one level does not require understanding of what the program is supposed to do: detection of endless loops, of arrays that exceed dimension limits, of incorrect iterations. Detection of some bugs, however, does require such understanding. The experimenters note evidence that the programmers debugged at the top level before trying to understand a program. If the bug was of a type that could be detected here, it was found quickly. If the bug was not found at this level, the programmer began getting into the substance of the program and concentrated on assignment statements, short ones first and then longer ones. Digging into more complex statements tended to be put off until that was the only thing left to do.

Other investigators who have studied either types of errors that programmers make, or the debugging process per se, include Miller (1973), Young (1974), Gould (1975), Litecky and Davis (1976), and Shneiderman and McKay (1976).

On-line Versus Off-line Operation

A question that has motivated several studies is that of the relative merits of on-line versus off-line systems. Typically, an experimenter has had two groups of subjects work on similar programming problems, one group working in an on-line conversational environment, and the other with a closed-shop batch system. The results of these studies have been mixed, and it is difficult to draw any general conclusions from them. Sackman and Citrebaum (1972) summarize ten of these studies (see their Table 1-5, p. 29), comparing the results with respect to four measures: man hours, computer time, costs, and user preference. The studies did not all agree with respect to any one of these measures. Users tended to prefer the on-line mode of interaction, which required fewer man hours than did batch mode.

In retrospect, the question of on-line versus batch mode of operation is of questionable meaningfulness. It seems unlikely that either mode is better than the other in any general sense. Each undoubtedly has its place, and which will prove to be most effective in any given instance will depend upon many things, but most particularly on the nature of the task being performed.

Effects of System Dynamics on Problem Solving Performance

Perhaps the first study of the effects of system response time on user performance was conducted at the Lincoln Laboratory by Morefield, Weisen, Grossberg, and Yntema (1969); (see also Grossberg, Weisen & Yntema, 1976). The experiments involved the Lincoln Reckoner, a system that was designed for use by scientists who are not necessarily programmers. It provided the user with a set of tools for executing numerical computations on arrays of data. The user interacted with the system via keyboard, a printer, and a cathode-ray-tube display. He could also obtain a hardcopy Xerox record of any material displayed on the CRT.

Subjects (the experimenters themselves) worked on a variety of problem-solving tasks. The nominal delay of the system in responding to commands resulting in output was varied over the following values: 1, 3, 10, 30, or 100 seconds. (The actual delay varied ± 10% of the nominal delay in all cases but the 1-second nominal delay; in this case, the system responded as quickly as possible.)

It should be noted that not all commands involved output; in general, output occurred as a result of a request for a display or as an error message. As a consequence of this feature, the effects of delays in output were presumably quite different from what they would have been if the system had insisted that after each command the user wait for a ready message before beginning to type the next command. The user was free to enter commands as rapidly as he wished, until he finally entered one that requested a display; until then, he had no need to wait for a response.

Because of the significance of this study as an early attempt to investigate person-computer problem solving in a controlled way, we will consider the tasks that were used and the results that were obtained in some detail.

Task 1: "Railroad Tracks." The first task that was used required a large number of interactions with the computer. The problem was to modify iteratively a horizontal line until it fell between a pair of "railroad tracks" which had the shapes of third-degree polynomials. The subject performed the task by making use of a command that added to the line a Gaussian bell-shaped bump, whose height, width, and horizontal location he specified. The command was entered by typing the letter B followed by numerical values for the three parameters. The effects of successive bump commands were cumulative. The display was automatically updated following each command, and was always visible. The independent variable in this case was the time between the issuance of the bump command and the updating of the display.

Each subject did 25 problems, 5 for each delay, a different pair of railroad tracks being used in each case. (In this experiment every command produced an output, either an error message or a new display.)

Mean gross time to complete the task increased, and the output rate (ratio of mean number of outputs to mean completion time) decreased, as system delay increased. Net completion time (time to complete the task with the delays subtracted out) decreased as system delay increased. The subjects commented that they could use the long delays to plan and type the next command. (One should bear in mind in interpreting this result that the longest delays were less than 2 minutes.) Finally, increasing delay resulted in a slight decrease in the number of commands the subject used to perform the task.

Task 2: "Black Box." The second task was to discover which of several possible transformations of two linear inputs would produce a specified output. Initially, a display contained two straight lines which represented the linear inputs, and one curved line which represented the desired nonlinear output. The subject could examine the result of any transformation or combining operation, and could also undo a result (essentially back up) if he wished.

About a dozen commands were available with which the subject was to identify the transformation that he wished to make on each of the inputs and the rules for combining the inputs. Each subject worked on 50 problems, 10 at each delay, with different inputs, transforms, and combining operations.

In two respects, the results were very similar to those obtained with the railroad-track task; gross completion time increased, and output rate decreased, as system delay increased. However, the net completion time increased slightly with delay, whereas the number of outputs did not change with this variable. In this case, therefore, it appears that the delay not only made the subject wait but degraded his performance during the time he was not waiting.

The black box task required only about half as many commands as did the railroad-track task, which suggests that it was the easier of the two problems to solve. The difference in the effect of system delay on performance in these tasks is consistent with the idea that the more the task requires thinking on the part of the subject between computer outputs, the less detrimental a forced delay. It may prove to be difficult to substantiate this idea quantitatively, because the "thinking demands" that a problem imposes on a user will be a function not only of the nature of the problem on which he is working, but on the type of help he is getting from the computer as well.

Task 3: "Scatter Shot." The third task used by Morefield et al. represented an attempt to study situations that are more representative of those in which scientists and engineers actually use the Lincoln Reckoner. A diverse set of problems was developed, hence the term "scatter shot". Some constraints were used, e.g., each task was to take a moderate amount of time, and was not to require special training.

Sixteen tasks were defined. Subjects had full resources of the Reckoner system when working on them. All performance measures were adjusted by means of a calibration procedure to correct for differences and difficulties across tasks. (Tasks apparently did vary considerably in difficulty.)

Again the results were similar in some respects to those obtained with the other tasks: gross completion time increased, and output rate decreased, with increasing delay. Net completion time increased (a result similar to that of Experiment 2). The mean number of outputs varied irregularly and uninterpretably with delay.

Conclusions from the three tasks Morefield et al. concluded that their results "demonstrate the feasibility of obtaining functional, quantitative relations in human factors experiments on man-computer interaction, specifically, in experiments on the delay in the machine's response" (p. 46).

The gross completion time results give some indication of how the amount of time the user must spend at the console depends on system delay. Unsurprisingly, these variables increase together. Output rate gives a rough indication of the <u>instantaneous</u> load that the user puts on a system. Again, unsurprisingly, this tends to go down with increasing delay. (We should note that reducing instantaneous load is not necessarily a desirable thing, even from the system engineer's point of view, if the total (integrated) load imposed by the entire interaction is increased.)

The net completion time results obtained by Morefield et al. suggest that how detrimental delays are may depend on the characteristics of the task that is being performed, although the nature of the relationship is not clear. There is not much evidence in these results (perhaps with the exception of those from the first task) that subjects adjust the number of output requests to compensate for the delays that they encounter.

<u>A study of "Lock-Out" effects</u>. Boehm, Seven, and Watson (1971) have reported an experiment, the purpose of which was to test a hypothesis of Gold's "that restricting one's access to the computer for a period of time <u>after</u> the presentation of current results ('lock-out') might improve performance by inducing the user to concentrate more on problem-solving strategy than on tactics" (p. 205).

The subject was given a map showing streets, freeways, and accident frequencies at different intersections. His task was to choose the locations for three emergency hospitals and to specify a set of decision rules regarding under what conditions freeways, instead of secondary roads, should be used for access to these hospitals. The objective was to minimize the average waiting time per emergency for the area.

The computer had been programmed to provide information on demand regarding the effectiveness of any assignment of hospital locations and decision rules. For example, in answer to one request it would type out a matrix showing the shortest response time from every intersection to the most accessible hospital, as well as the average expected transit time per emergency, taking into account the frequency of emergencies in the different regions and the accessibility of the various hospitals to the different intersections. Thus, the man's role was to suggest problem solutions, and the computer's role was that of quantifying their implications.

The experimental variable was the time that the user was locked out of the system (prevented from typing in a new input) after each set of results had been presented to him. Five conditions were used: no lock-out, 5-minute lock-out, 8-minute lock-out, a variable lock-out (5-minute mean), and a choice lock-out (in which subjects were asked to lock themselves out as much as possible, but otherwise had free access to the console).

Inter-subject variability was very large, forcing any conclusions to be tentative ones. The group with the 5-minute lock-out did as well, or perhaps better, than any other group, and did considerably better than the group with the 8-minute lock-out. Subjects with free access averaged twice as much computer usage as did subjects with restricted access. Dissatisfaction with restricted access was expressed by subjects even in groups with relatively high performance.

Commenting on the very large individual differences, the investigators note that "it is difficult to imagine anyone ever formulating a single model of man-computer problem solving that would fit even our small group of subjects, which included some whose performances were so irregular that they had to be dropped from the analysis" (p. 208). Noting the dissatisfaction of the users concerning the restricted access, even on the part of those who seemed to profit from it, the investigators suggest that the results cast doubt on the validity of user acceptance as an index of system effectiveness.

Some Examples of Human Factors Considerations in Existing and Planned Systems

While there is, as yet, no theory or even organized body of data to provide human-factors guidelines for developers of interactive computer systems, many system developers and designers have attempted to be sensitive to user needs and preferences. In this section we give some examples of features of several systems that were motivated by such considerations.

An Agricultural-Loan Management System

The Agricultural Stabilization and Conservation Service of the U.S. Department of Agriculture has worked on the design of a computer-based system to facilitate the processing of small-loan applications in keeping with government land conservation and reconstitution policies. Plans called for a central computer and data base located in Kansas city, field office terminals in approximately 2000 state, county, and district centers, and data concentrators between the field office terminals and the computer. The field office terminals would be keyboard input devices, but would include video output and about 4k of local storage capacity. The system was to be used in a real-time interactive mode by ASCS field office personnel who know little about computers but much about ASCS programs.

Several features have been specified that relate to the way in which the users would interact with this system (Pew and Rollins, 1975). Among them are the following:

a. Frame Orientation: The user-computer interaction is to take place through a series of "frames". A frame is a formatted display. It may be thought of as a page, essentially what will fit onto the video display at a given time. Three fundamental kinds of frames are defined in the system: menu frames, input frames, and output frames. Only input and menu frames require input from the user.

b. Documentation pointers: Every nonmenu frame is to contain an abbreviated reference to an operator´s manual by giving the number of the paragraph where an explanation of the procedure involved in that particular frame may be found. Documentation pointers are to be located in the same spot on each frame on which they occur.

c. Split-screen capability: Background fixed fields and foreground variable fields are to be distinguished on the video display by brightness level. The background fields are protected from keyboard control.

d. Default key: When struck, the default key is to produce the single most likely (as prespecified) character or sequence of characters in the current character position (at which a cursor is pointing) in the variable field. Inasmuch as the predictability of user inputs is expected to be generally high in this system, this feature should considerably increase the user's input rate.

e. Help key: The striking of the help key will elicit help messages that relate to the current field entry, or will refer the user to such messages in a manual.

f. Escape key: One press of this key will back the user up to the first frame of the currently active subsystem; two presses will get him back to the starting menu.

g. Suspense files: The intent of this feature is to permit the user to store incompleted files so they can be retrieved and completed in subsequent work sessions.

A Management Information System

An example of a system built for a special class of users who are not computer specialists is the BBN Management Information System (Ebeling, 1974). One purpose of this system is to provide management with useful information on demand. The system contains a set of master files of descriptive and summary transaction data, and makes these data available to managers via interactive inquiry of the data base and on-demand report generation. The system operates in both batch and interactive modes, and makes use of both commercially available software packages and custom-built programs. The major system software modules include a set of foreground system routines (control system, report system, retrieval system, purge system, etc.) and a set of foreground transaction routines (time-sheet entry system, requisition system, accounts payable system, manpower loading system, etc.).

Perhaps the module of greatest interest, from the point of view of user-computer dialogue is the report system. The purpose of this module is to provide the manager with the means of obtaining financial and other managerial reports on request. The dialogue in this case is best characterized as an interrogation of the user by the computer. The computer asks a series of questions of the user to determine what it should do by way of report preparation and distribution. Examples of these questions include: "Report series?", "Report No.?", "Priority?", "Distribution?". The user answers these questions with brief, and usually numerically coded, inputs. Each question has a set of admissible answers, and this set is displayed for the user upon request.

The system has some features assigned to protect the user from incurring unnecessary costs. The user is forced, for example, to confirm a report selection by answering the question, "Is this the report you desire?" before the specified report is produced. Also the system gives the user an estimate of what the production of a report will cost before it is produced. Finally, it provides him with a set of options for report production that permits him to select among several speed-cost tradeoffs the one that best suits his needs. The four options are IMMEDIATE, EXPRESS, DEFERRED, and OVERNIGHT. He selects the one that he wishes by typing the first three letters of its name in response to the question, "Output option?' The cost of IMMEDIATE, DEFERRED, and OVERNIGHT are, respectively, 1.1, 0.6, and 0.4 of the cost of EXPRESS. EXPRESS is the default option and represents the response that one would get if the program were run with neither a special priority nor a handicap.

A Physician's Aid System

Many writers have emphasized the desirability (and in some cases, the necessity) of involving potential users of a computer-based system in the design and development of the system. This involvement may be justified on more than one basis.

(1) To the extent that a computer system is being developed for the purpose of facilitating the performance of some task, it seems likely that the people who normally perform that task would have useful thoughts to contribute to the question of how the task that they perform might best be facilitated.

(2) A sine qua non of a useful system is that it be used, and there is some evidence to indicate that potential users are more likely to accept a system and give it a fair trial if they have been involved in its design and development than they will if they have not been so involved. This point is illustrated by the experience of Castleman and his colleagues in developing the CAPO (Computer Aids in the Physicians's Office) system (Castleman, Whitehead, Sher, Hantman & Massey 1974). The intent of the designers of this system was to provide aids to physicians in five areas: medical history taking, patient education, continuing physician education, medical consultation, and accounts receivable. The software was run on a centralized time-shared facility and accessed through terminals located in physicians' offices. The user terminals were composed of a cathode ray tube display and a keyboard input device with a 30-character per second output capability.

The developers of this system have reported many insights that should be of interest to anyone who plans to attempt to introduce a computer-based system into a person-oriented operational environment. For our purposes, however, the point of interest is the effect of the physician's participation in the design phase of the system, and his subsequent willingness to make use of it. Castleman et al. point out, for example, that physicians were generally reluctant to use the automated medical history taking program at first. "In every practice, once the physician had modified the original questionnaire in some way, often quite minor, it was much more highly regarded by the physician and was used with greater frequency" (p. 17). In the same vein, the writers point out the need they discovered for customizing programs for specific users. It was difficult, they found, to convince one individual to use programs that had been developed for another. "The practices considered it a <u>necessity</u> to modify any such application program in some way, however slight, to customize it to their own particular circumstances, before the program was found really acceptable" (p. 39).

Another, somewhat related, observation made by Castleman et al. was that the most readily accepted programs were those that could be integrated smoothly into the existing office routine. Presumably programs requiring drastic modifications in the standard operating procedures would be likely to meet with considerable resistance. This is not to say that such programs would never be accepted or could not prove to be preferred in the long run; however, it does suggest that, other things being equal, the less disruptive the introduction a new way of doing things is, the more likely is it to be accepted without a struggle.

A Technical Information Handling System

The PROPHET system is a computer-based information-handling system for pharmacologists, medical chemists, and other scientists who study drugs together with their chemical and biological interrelationships. The system is intended to facilitate research on the relationship between molecular structure and biological activity, and to foster collaboration among scientists working on such problems. The system itself is described in several reports but most completely in Castleman, Russell, Webb, Hollister, Siegel, Zdonik and Fram (1974).

Recently, Rubin and Risley (1976) have reported some conclusions based on the experience gained from introducing this system to a fairly large number of scientists over a period of five years. The major conclusions that they draw are the following:

a. "Beginning users have a lot to assimilate. The system shouldn´t overtax their memories." This suggests the desirability of incorporating memory aids within the dialogue procedures. One means to this end is that of applying an English-like command language, i.e., a language in which the commands mean approximately what a naive user would interpret them to mean on first seeing them. Another is that of using common-sense default values for variables (e.g., for variables whose values he may fail to specify). Another way of minimizing the memory load imposed on the user is that of providing display menus whenever the user must choose the operation that he wishes to specify from among several alternatives.

b. "The system should provide the resources for users to gradually increase the complexity and sophistication of their interactions with the computer." This means providing the user with the option of eliminating on demand some of the features that are provided for the benefit of novices. It also means giving the user the capability of developing increasingly sophisticated interactive procedures that are specially tailored to his own needs.

c. "Users should have a major role in how the system develops."

d. "Traditional reference manuals are nearly useless for beginning users." Rubin and Risley do not condemn traditional user manuals in a wholesale fashion. They acknowledge their value as a means by which experienced users may discover new system features or review old ones. Their point is that they are not useful as a means of introducing novice users to a system's capabilities. What the naive user needs, they contend, "is a functional introduction to the system, a document which allows him or her to accomplish a complete interaction with the system and to achieve some goal." They advocate, for this purpose, the use of "scripts," recorded interactions between PROPHET and experienced users which permit the beginning user, in effect, to observe how an experienced user goes about working with the system toward the solution of a particular problem.

e. "'The personal touch' is an important component of success in such a system." Interaction between the developers of the PROPHET system and its users is seen as especially important. This interaction is provided for in two ways: periodic visits to user groups and 24-hours-a-day access to a PROPHET expert via telephone. Periodic visits (about every six weeks) are considered an important way to assure that PROPHET users are learning what is necessary to become more and more expert in their utilization of the system.

f. "User groups need a local 'protagonist' to keep things running smoothly." The role of the protagonist is to serve as the resident expert for the user group, helping new users to get on the system for the first time, keeping infrequent users posted with respect to new system features or capabilities, and being a point of contact for the PROPHET staff.

The Cambridge Project and the Notion of a Consistent System

A purpose of the Cambridge Project was to make powerful computing resources available to nonprogrammer behavioral scientists, thus providing them with sophisticated tools for research without forcing them to become computer specialists themselves. The project has generated a number of relatively independent computing tools, most, but not all, of which are separate computer programs. Each group of tools has been developed by a different set of scientists, and each was designed with some specific application in mind. Typically, the developers themselves have been the major users of the tools they have been developing, although the intent of the project was to produce some that will be useful by other scientists who are not involved in the developmental process.

A long-range goal of the project was to assemble the tools that were developed into a "consistent system." The idea of a consistent system was a central one to the project. It is of some interest to note that this term replaced the term "integrated system," which was originally used in the project's documentation. This distinction is an important one. Whereas "integrated system" suggests a system for which all the components interact in some way, "consistent system" carries no such connotation. The latter term does suggest, however, a system for which the procedures involved in using one system component are compatible with those involved in using others. A user wants not to be forced to learn about aspects of the system other than those directly relevant to his use of it. By definition, a consistent system would have this property. On the other hand, a user also wants to be sure that whatever he learns in order to use some capability of a system for one task will not interfere with his learning what is necessary to use other capabilities for other tasks. He would strongly prefer that what he learns to use Component A will in fact facilitate his learning of what is necessary in order to use Component B. He would be especially pleased if it turned out that he already knew much of what he needed to know in order to use Component B by virtue of the fact that he had once learned to use Component A. In attempting to realize this kind of consistency, the developers of this system set two explicit objectives: (1) that the data that are output by one program be usable as inputs to any other program - insofar as this makes sense, and (2) that all programs be invoked in a consistent fashion (from the user's point of view), thus making it convenient for a non-programmer user to construct "agents" to run programs for him.

The effort to combine programs written by dozens of individuals, working in varying degrees of isolation from each other, into a system with the kind of consistency that a user needs in order to be able to move freely from one component to another is a very ambitious undertaking. It is also a worthwhile one, perhaps an essential one, if the system is in fact going to be used by people other than its developers. Some of the steps that the Cambridge Project staff took in this direction are instructive.

Organization of data files. Consider first the way in which the problem of data files was handled. This is a key problem, and two extreme ways of handling it are fairly obvious. On the one hand, one might insist on rigidly prescribed formats, in which case users would have no flexibility and, for that reason, might be inclined not to use the system. At the other extreme, one might permit every user to organize his own files in any way he wishes, in which case the consistency across programs would be nil, and it would be very difficult for users to share the results of their individual efforts. The Cambridge group struck a compromise: its strategy was to try to acquire a sufficient degree of standardization to make the sharing of files feasible, without giving up much in the way of flexibility and freedom of innovation for the user. A file has three components: (1) data, (2) a machine-readable description of the format of data, and (3) a 8-character code that identifies the format, not of the data themselves but of the description of their format. What the Cambridge Project group standardized, therefore, is not the way data must be organized, but, rather, a procedure for specifying or discovering their organization, and a procedure for "publishing" a new organization and making it part of the conventions of the system.

An important aspect of the approach is the fact that if a user does not like any of the data formats that are available within the system, he can invent a new one. Of course, if he does so, he must also describe it (and, presumably, write a program to decipher the description). This approach provides for standardization, which is necessary, but in doing so it does not prevent the individual user from organizing his files to suit himself.

Command language. The convention used by the command language for accessing programs is similar in principle to the method used by many programming languages for calling subroutines. The name of the program is given, followed by a string of program parameters, the interpretation of which is the task of the program that is called. The burden is on the user of a program to know what the appropriate calling sequence of that program is. That is to say, he must list the program parameters in precisely the correct order and format.

Programs as agents. Programs in the consistent system are viewed as users' agents. This idea is analogous to that of subroutines, or procedures, at another level of person-computer discourse. The principle of subroutinization has been considered by many to be one of the more powerful principles that has emerged from the experience of trying to tell computers what to do. The use of subroutines accomplishes several things for the programmers. It simplifies the problem of thinking and constructing complex procedures by permitting one to break complicated tasks into simpler subtasks, develop procedures for performing the subtasks, and then to use those procedures as components in the solution of the larger problem. It also accomplishes a significant economizing of computer storage inasmuch as the code for any procedure need exist within the machine only once, rather than appearing embedded within the calling program once for each time that that particular task must be performed.

Perhaps the greatest advantage that comes with the use of subroutines, however, is the possibility of building a library of procedures, that is specifically tailored to one's computational needs. Once a subroutine has been properly written, it need not be written again (at least as long as one works with the same computer system). The user needs to remember only what the subroutine does and how it is called into action. Consequently, by enlarging his library of subroutines, a programmer acquires a more and more powerful bag of tools. Each time he writes a new applications program he will have more preprogrammed procedures to call into play than he had before. In some cases, writing a new program may mean little more than producing a sequence of "calls" to existing subroutines. In part, what the Cambridge Project group is suggesting by the notion of programs as users' agents is that a consistent system should be designed in such a way that its various component packages can be treated in a fashion that is analogous to how subroutines are treated within an individual program.

System maintenance. The term "maintenance" has two important connotations when applied to a system that is intended to provide services to users while in the process of being further developed. The more familiar connotation is that of keeping the system in good repair. No less important from a user's point of view is the connotation of maintaining the system's stability. The problem in this case is how to assure that a developing system will be sufficiently stable in the eyes of the nondeveloper user so that he will be able to work with it over extended periods of time without continually having to learn new operating procedures. One might refer to this as the problem of "longitudinal consistence."

To solve this problem, the Cambridge Project attempted to do two things: (1) localize all contacts with the underlying system (MULTICS) in the "substrate" - a pseudo-operating system amounting to only about 50,000 words of object code - which will be "maintained" to compensate for the inevitable changes in MULTICS, and (2) design everything (data structures, command conventions, etc.) so that improvements are always made by adding new software, never by changing existing software.

Summary

The systems described above have been developed (or planned) for a variety of applications. They have in common the fact that the intended users include individuals who are not trained in computer science and have no programming experience. Collectively, they incorporate a fairly broad spectrum of features, that were motivated by the desire to make the systems useful to and useable by such people. These features are not, for the most part, based on empirically validated principles, but on intuitions and opinions gained from experience working with such systems. We believe that an effort to organize these intuitions and opinions and to articulate them in a form that would permit them to be tested experimentally would advance the state-of-the-art of person-computer interaction.

Alternative Dialogue Forms

Each of the systems discussed in the preceding section operates in an interactive mode. That is to say, the user and computer engage in a process of giving and taking information that, for want of a better term, we may refer to as a dialogue. In this section we consider explicitly the notion of person-computer dialogue and distinguish among several types of dialogues that can occur. These types differ considerably with respect to the nature of the demands they impose upon the user.

Menus

With a menu structure a set of alternatives is presented and the user is provided with a way to select one or more of the alternatives displayed. The response mode may involve the moving of a cursor, the typing of a symbol or pointing with a light pen, stylus or finger. Frequently a hierarchical sequence of menus is provided to elaborate a tree structure of possible choices. A user who is naive with respect to computers is still likely to understand a menu procedure. The only requirement is that the terminology used to describe the choices be understandable. Whenever more than two or three choices are possible and the designer is not willing to assume that the user knows what the alternatives are, then a menu is an appropriate structure to use.

Formatted Inputs

A form-filling frame is a computer translation of a printed form onto a CRT display. The extent and complexity of the form may vary from one designed to obtain a value for a single parameter or variable to one intended to support a complex data-entry operation. Generally, the fixed or background part of the form is a protected field, and is displayed at an intensity different from that of the data fields. The data field may be of fixed or variable length and usually requires a terminator. The data fields may be edited, but the background fields may not be. Again, only knowledge about the terminology in the background fields and understanding of rudimentary cursor control is required of the user to interact with such a form.

Question and Answer Inputs

Printing terminals are not suitable to form-filling frames, but an abbreviated question and answer format or prompting format can serve the same purpose. In these cases a sequence of fixed or background fields is presented, one at a time. The user fills in the appropriate response and then is presented with the next background item. If the background item is phrased as a question it is called a question and answer frame. If it is more abbreviated it might be called a prompted entry. Again in either case the desired input may be a fixed length code or a variable length code, perhaps even an extended piece of free-form text that will be entered, but not interpreted.

In a typical message-system application package designed for printing terminals, a prompted message composition sequence calls for entries in a series of header fields and then for the entry of the body of the message as a text field. In a CRT-based terminal the same operations would be displayed as a form-filling frame. Note that with any of the preceding methods there is no ambiguity concerning what is required. These techniques are particularly well adapted to inexperienced users: note that the computer remains in control of the interaction. In contrast with these techniques are those in which control is transferred to the user at points where what is to be done next has been intentionally left incompletely specified. In this case, the user has the opportunity to initiate commands rather than simply to select a response from among those suggested, as described next.

Limited-Syntax Command Languages

These procedures are direct descendants of interactive computer programming languages in which commands may be either stored for future execution or executed immediately upon being specified. We will refer to them as limited-syntax command languages. Such languages represent a broad class of dialogue forms. Use of these languages requires considerably more knowledge on the part of the user than does use of the dialogue types mentioned above. The acceptable commands at any point in a transaction must be recallable by the user, not just recognizable as in the previous cases.

By use of the term "limited syntax" we mean to suggest that the structures in which the commands must be formulated are predefined. Typically a command line begins with a verb such as "edit", "compute", "locate", etc. The verb is followed by one or more arguments that successively bracket the domain and specify the values of parameters to be addressed. When the command is parsed the system expects to find one of a pre-defined set of possible arguments at each point in the sequence. Thus both the structure and vocabulary must be learned by the user. The advantage of limited-syntax structures is their conciseness and efficiency. If a particular activity is frequently repeated, it would be tedious and unacceptable to an experienced user to be forced through a series of menus or form-filling frames to define a desired activity.

In systems in which considerable training can be justified, command languages can be made still more concise by introducing abbreviations and brief codes for both command verbs and arguments, as has been done in commercial airlines reservation systems. This renders the transcript of a transaction relatively unintelligible to the uninitiated, but provides for substantially more rapid command entry. It should be clear, however that successful design of command languages requires study and care, if their formulation is to be consistent with the way in which the user is likely to think about the operations to be performed.

Some command languages provide a middle ground between terseness and the ease-of-use of form-filling or prompted input. At each point at which new argument is to be introduced the user has the option, in these systems, to depress a special key. Depression of the key causes the system to provide the user with a brief description of the class of argument that it is expecting at that point in command formulation. For example, the terse command:

File Typescript 1 C^3 File

might be presented in fully prompted form as,

File (input file) Typescript 1 in (output file) C^3 File.

It should also be noted that increasing the versatility of command structures in such ways requires much more extensive front-end software to make the dialogue flow smoothly.

Special-Purpose Function Keyboards

Another way to promote conciseness is to replace the requirement for typing each command verb with use of a labelled special-purpose function keyboard that identifies each command with a single key press. Such a keyboard has the advantage of providing a reminder concerning what commands are available at any point without the concomitant constraints of requiring the user to step through a lengthy menu or form-filling frame. Sometimes the special keys are integrated into a standard alphanumeric keyboard and sometimes they are designed as a separate block of keys for the designated purpose. This approach has the disadvantage of being relatively inflexible if the terminal is to be used for many different purposes. Some special purpose keyboards are provided with labelled overlays that permit the user to change the assignment of keys according to the requirements of different software packages, but to change assignments within a particular application can be confusing and conducive to high error rates. As a practical matter overlays also have a tendency to get lost or misplaced, placing still heavier demands on the user's memory of key assignments.

An example of the effective application of a special function keyboard is provided in the system simulation of a remotely-piloted vehicle (RPV) command and control system developed by the Air Force Human Engineering Division. This system employs IBM 2260 terminals dedicated to monitoring and controlling a flight of RPV's. Each of four CRT terminals displays flight path information and vehicle status information. It has a light pen, an alphanumeric keyboard and a separate function keyboard to activate frequently-used commands. These commands specify such operations as changing display scale factor, changing RPV speed or altitude, and modifying flight path. Many of the commands require an argument that is provided either through a light pen designation of the vehicle whose status is to be changed or a description of the revised flight path. The alphanumeric keyboard is used to enter new parameter values such as velocity or altitude. The special purpose keyboard is particularly effective because the operators sometimes work under severe time constraints and typically need to issue well-defined and unchanging commands.

Natural Language

There has been considerable interest in developing the capability to implement person-computer dialogues in relatively free-form natural language. This capability has been pioneered in the development of computer-assisted instruction systems in which the users are dealing with subject matter that is rich in vocabulary and for which the unnatural constraint of menus or question and answer formats communicates the image of an instructor as a robot with a relatively low IQ. The problems involved in implementing natural language dialogues are legion. Such a capability requires a large stored vocabulary with an associative network of synonyms and related concepts, a parser of natural language grammar, a sentence-constructing algorithm for synthesizing answers to questions that arrive in unpredictable formats, a historical record of previous sentences and their analysis in order to interpret the context sensitive features of human discourse, and so forth.

Considerable progress has been made on these problems. Intelligent programs such as SOPHIE (Brown, Burton, Bell, Bobrow, 1974) represent the state-of-the-art today. The following sample of a SOPHIE protocol from an electronics trouble-shooting training session illustrates its power but also the limitations in what it has been possible to achieve. The user (U) has been given a circuit diagram of a power supply that has a fault in the circuit. He is trying to identify the fault by asking the computer (C) a series of questions about its performance.

U: What is the output voltage?

C: The output voltage is 25.3 volts.

C: If you are not sure of whether a value is correct, you can ask "is that value correct?"

U: Is that correct?

C: In a working circuit the output voltage is 30.1 volts. That value seems to be incorrect.

U: What is the output current?

C: =output.

C: The output current is 25 milliamps.

U: Is Q5 conducting?

C: I don´t know the word conducting. Maybe you could rephrase your question.

U: Could Q5 be shorted?

C: The shorted modes of a transister I handle are: All junctions shorted or one junction shorted with the opposite terminal modelled open. Should Q5 have fault mode EB/EC/BC=SH/SH/SH.

U: Y

C: OK, I´ll use EB/EC/BC=SH/SH/SH.
Hmmm, let me think about that for a minute. (pause)

C: When the settings were:
CC=1.0 CR=High load=1000
VC=1.0, VR=High
That fault does not explain the following symptom: You observed that the voltage at N11 was 25.3 volts. If Q5 had fault mode EB/EC/BC=SH/SH/SH it would be -4.1 millivolts. In a working circuit it is 30.1 volts.
etc.

Natural language seems appropriate and sufficiently effective to be used in such non-critical applications as instructional systems, however in the command and control context natural language interfaces based on keyboard entry seems unnecessarily verbose and inefficient for highly trained users who can be expected to learn to use more efficient command languages for interacting with a computer.

On the other hand the development of natural language interpreters will be an important component of future developments in computer-based information and retrieval operations in which data storage and retrieval algorithms will be based on the semantic content of natural language messages rather than simple scans of key words or prestored phrases.

A researchable question of considerable importance for both dialogue development and information processing operations is whether individuals can learn to communicate using a constrained subset of natural language grammar and vocabulary that would reduce both the verbosity of natural language dialogue and the difficulty of computer interpretation of human input. Such a dialogue mode may be thought of as intermediate between limited-syntax command languages and natural languages; it should provide further useful information for improved structuring of command languages for ease of human use. Relatively little empirical work has been done on the effectiveness of such languages. One study of the effects of restrictions on vocabulary size was reported recently by Kelly (1975). Two people were placed in separate rooms and had to communicate by teletype. They worked on two tasks requiring interactive problem solving. The investigator compared performance with (1) an unlimited vocabulary, (2) a 500 word vocabulary consisting of 425 function words and 75 task-related words and (3) a 300 word vocabulary consisting of 225 function words and 75 task-related words. In each case the subjects were given training with the admissible vocabulary. Time to solve the problem and total number of communication exchanges proved to be independent of vocabulary size in this study. Much remains to be done, however, to explore minimum vocabulary size, restrictions on grammatical constructions and the generality of this finding to other classes of tasks.

Human Speech

Speech as computer input is usually thought of in connection with natural language capability. In fact it is a nearly orthogonal possibility. One could use speech as an input mode (e.g., isolated words, small vocabulary) without a natural language capability; and one can have a natural language capability without speech.

Practical speech recognition systems for isolated words separated by 250 msec have been developed and several groups are working on the problem of computer understanding of connected discourse. Because these developments are relatively new and the recognition of connected discourse has not yet reached the state of being operationally useful, little attention has been devoted to the human factors design of dialogues that exploit human speech. While it is acknowledged that speech can produce substantial improvements in speed of input, several questions must be answered before we know the circumstances under which such systems will be practical and before recommendations for interactive speech dialogues can be made. Can a user easily adapt his or her speech to the requirement for artificial word segmentation? Can an individual sustain voice input in this mode for long periods of time? What are the most effective means of providing feedback concerning the correctness of encoding of the speech input? What backtracking modes are needed to deal with recognition errors? For what classes of data input is it most effective: numerical data, alphanumeric codes, text? If text is feasible, how does one integrate formatting and punctuation instructions with the text itself? What are the conditions under which voice output is a useful mode to be preferred to CRT or hardcopy printout? These are uncharted areas in need of substantive research.

Interactive Graphics

Alternative dialogue procedures for application to interactive graphics have not yet received as much attention as have other dialogue types. Many of the same procedures are applicable however. Menu selection is a commonly used procedure for selecting frames to be displayed. The menus are shown as an integral part of the frame and the choice of menu items is most frequently accomplished by directly pointing with a light pen or stylus to the item in question. Manipulation of the graphical entities on the display - positioning a cursor, modifying scale factors, rotating or positioning displayed objects, designating selected attributes - may be accomplished by keyboard operations, but it is frequently more compatible to display "light buttons" directly and use a light pen or stylus to activate changes in these parameters. Other alternatives involve the full range of graphical input devices, including the joy stick, bowling ball control, the mouse or finger position sensors.

Some Tentative General Principles of Interactive-System Design

As was suggested earlier, the initiative with respect to the front-end design of interactive systems has rested with systems designers. Their approach has typically been to put a system together, try it out, acquire a feeling for how it behaves, and carry through many iterative steps to achieve a final design. Over the last several years designers and others representing the user community have begun to step back and reflect on systems in the field. Several papers have attempted to abstract principles of dialogue design that promote effective, interactive sessions. In the material that follows we attempt to bring together the information derived from these several sources into a tentative integrative set of principles. Virtually none of the principles has the force of empirical evidence behind it. But, in the absence of empirical data, they represent the only existing human factors design guidelines for such systems. Moreover, in combination, they constitute a set of working assumptions and hypotheses and as such, they invite attempts at empirical verification or refutation.

Not every principle that we encountered is incorporated in what follows. Our criteria were informal but in general the suggestion had to come from two independent sources and it had to appeal to our own intuitions.

As we discuss the principles no attempt will be made to attribute them to their specific sources because different authors express them in different ways and at different levels of specificity. Rather we provide here the sources we have used to develop this integrative summary: Engel and Granda (1975); Foley and Wallace (1974); Kennedy (1974); Martin (1973); Nickerson and Pew (1971); Pew and Rollins (1975); and Smith (1974).

The discussion is organized in terms of eight "topical" notions. As stated, these notions are too general to be useful. The intent in stating them was simply to delimit broadly the types of issues or general objectives to which the principles are addressed.

Understanding the User and His Task

A cardinal tenet of interactive system design is: "Know the user population." The precise dimensions of information about which such knowledge is useful and important will vary from application to application. Some of these dimensions are characteristics of the users themselves and some relate to the user's job or activity description; it is not always possible to make a clear delineation between the two. Among these dimensions are the following:

(1) Knowledge or expertise as a computer programmer, systems analyst, computer operator, or other computer-related specialist.

(2) Knowledge concerning the specifics required for carrying out the job. What can be assumed concerning the user's understanding of how a particular application is to be carried out?

(3) The extent to which user's job or activity will focus on interactive terminal usage. Will he be using a terminal on a dedicated or casual basis? Will the terminal serve as an information source or as the basis for regular work performance?

(4) Level of decision-making authority and responsibility.

(5) Educational background.

(6) Availability of special skills or aptitudes such as clerical skills, managerial skills, mathematical skills.

(7) Expected duration of stay in particular job; employee turnover.

(8) Sources of job-related motivation. Is the user intrinsically motivated or must the interactive tasks be designed to promote such motivation.

(9) Extent to which terminal usage will be an option versus a job requirement.

(10) Attitudes toward computer technology and its introduction in the work setting.

These characteristics obviously are not isomorphic with dialogue principles, but they provide the background information that is needed to make choices among alternative dialogue forms and to assign relative importance to the various design considerations that will be discussed next. It is only with this kind of information in hand that one can seriously hope to design a system to be maximally compatible with the characteristics of its users.

Making the System Conceptually Simple

Many of the desirable characteristics of interactive dialogue may be classified under the general rubric of promoting simplicity from the user's perspective.

Taylor, as early as 1967 said,

"When a man uses a computer to aid him in solving a problem, that problem must have a least two extreme forms of representation and perhaps many intermediate forms.One of these forms should give the problem a structure which is amenable to the man. The other form gives the problem a structure amenable to the innermost processes of the computer. The processes within these internal structures should be invisible to the user while they are working for him. Conceiving, designing, and developing a smoother transition between these two forms of representation is the crux of the man-computer input-output problem" (Taylor, 1967, p. 2).

One idea that has emerged is that the designer should formulate and keep in mind a perspective or model of how the system should be perceived by the user. The more complex and powerful the system actually is, the more important it is to have such a model well-defined.

The model must obviously be application-specific. In a personal message handling system, for example, one might want to create the image of a series of user-defined file folders, classified by subject, together with a free-form working space in which messages may be drafted and edited. Alternatively the model might be one in which the user initiates a message draft by calling up one of a series of prestored forms appropriate to the class of message that is to be prepared. In this case the storage space might be structured according to message classes with message subject matter as subcategories.

In the case of a data base of geographically-distributed resources one might create the model of the geography with the ability to focus in on particular sites and probe for greater and greater detail concerning the specific characteristics of the resources available at those sites. Alternatively one might structure the model on the basis of a taxonomy of resource characteristics and provide for retrieval of the geographical distribution of particular classes as one of a number of possible data base inquiries. The critical point is to formulate the model in a conceptually simple way that is as compatible as possible with the way the user thinks about his job and the tasks assigned to him. Quite different dialogue forms may result from different structural models.

Given that an appropriate and simple structure has been provided, it is also important to impose a logical structure and segmentation on the activities the user is to perform. A hierarchical structure is often the most suitable and a useful principle is to break tasks and subtasks into logical modules, just as an experienced programmer organizes his code. No matter how complex the interactive task, it can probably be broken down into a series of simple steps having logically well-defined boundaries. Menu selection activities lend themselves nicely to a hierarchical sequence of choices. In a graphical interactive task such subtasks as indicating the initial and terminal points of a line with a light pen and issuing the command to connect them is representative of a well-defined module.

In the development of a limited-syntax command language the same concepts apply. A typical issue is whether to have a single level language with all possible commands available at the same level or instead to organize the system into a series of subcommand modules in each of which the set of available commands is limited. While there are advantages to a single level language, when the complexity imposes the requirement to formulate many distinctive command labels for similar but distinct functions or requires every command to be qualified by a long string of arguments, the goal of simple structure is probably better achieved by a shift to a multiple-level hierarchically-organized command structure.

A further dimension of clarity and simplicity is represented by the concept of minimizing the requirement for the user to refer back to previous actions taken within the same transaction sequence, or to recode, compute, or interpret presented information in order to add a new entry. The system should provide its input in a directly useable form and provide prompts or reminders of the current state of transaction development. If the user has just completed a lengthy menu sequence to get to a particular frame, a label or capsule summary header should communicate the parameter values or attributes that characterize that frame. In a lengthy sequence it may be appropriate to recap periodically what has been established and ask for a confirmation. Provision of such summary headers or reviews are especially important when default values are employed or in more intelligent systems when the user is allowed more degrees of freedom in expressing a request. The user should always be informed concerning how his request was interpreted.

It should never be necessary to enter information already available in the system, but it may be appropriate to display such information at a time when it is relevant again to the current activity. When it is displayed it should be in the form needed at that point even if the format is different from that provided in the data base or when it was originally entered. For example, in a payroll or cost-accounting system salaries may be stored in hourly rates, but if the current activity requires monthly or yearly rates, the computer should make the required transformation and display accordingly. Similarly, a search for salary that is initiated on the basis of social security number might appropriately return both the employee's name as well as salary so that the user is reassured that the search code was correctly entered and interpreted.

Provision of a command verification or confirmation requirement carries with it the responsibility to make it easy for the user to modify a request that is revealed to be inconsistent with his intent. In particular, it should be possible to move backwards through a dialogue sequence with the provision to change an entry, a parameter setting or perhaps a whole transaction sequence. Introducing such changes should not require re-entry of all the correctly entered material even if it is within the same frame line.

One final principle belongs under the general rubric of simplicity -- minimizing the complexity and clutter of visually displayed information. Displays should be tailored to the activity at hand and minimize the presentation of irrelevant information. One author advocates one "idea" per frame, but that suggestion leaves open the question of what constitutes an "idea" and provides little guidance concerning the allowable complexity of individual frames. The designer must also make judgments concerning the trade-off between the generation of unique frames for every new context in which a given set of information is required and economies involved in standardized frames for certain kinds of information that are used repeatedly throughout a transaction or application package. Minimizing complexity in presentation and interpretation is a desirable goal in any case.

Maintaining Consistency from the User's Perspective

Whereas applications programs tend to be written by different teams of programmers, they often tend to be activated by the same population of users at different times. For this reason, consistency and some degree of standardization are very important at all levels: from frame to frame, from dialogue to dialogue, and from program to program within the same system. Format consistency exploits a user's ability to learn where to expect to find particular items of information. Consistency of command syntax similarly permits the development of relatively "automatic" or preprogrammed actions on the part of a user and permits him to extrapolate principles learned for the execution of well-practiced commands to the acquisition of procedures for learning new ones.

With respect to information in a data base, the user has a right to expect that if he has modified an item at one point in time in a particular file that that change has been introduced in every file location where the change is relevant. Similarly, in programs that permit user definition of terms or symbols or the introduction of user-defined synonyms for existing parameters, the system must permit use of those synonyms at any point in the program where they are relevant.

The development and application of a conceptual model of user interaction in a particular system provides a means for carrying the concept of consistency to the deepest levels of program implementation. The user can develop expectancies or rules of how the system ought to operate. If the principle of consistency has been adhered to carefully in developing the system, he is then rewarded by discovering that a new procedure is in fact implemented according to his expectations.

Saving the User's Time

Regardless of the level of expertise, the user's time represents a valuable resource that should be conserved. While many of the recommendations given in the name of simplicity and consistency can make important contributions to efficiency as well, there are others related to specific user control and response activity.

It is generally found that a small proportion of available actions or commands accounts for a large proportion of the activities that a user undertakes. While the programmer must provide alternative responses for all possible user actions, it is desirable to analyze the relative frequencies of possible actions, and make those with the highest probability of occurrence the easiest to accomplish.

Operator actions can be made simple by assigning unique function keys to them, by the judicious use of default settings of parameters most frequently used, by exploiting the potential for positioning a CRT cursor under program control to the place where the user would be most likely to move it next and by minimizing the need to shift from one response device to another. The latter suggestion is of particular concern in graphics systems in which it is very convenient to provide a light pen, stylus or joystick control of cursor position, but where these activities may be mixed with keyboard data entry or parameter value specification. "Light buttons" or menu items selected directly on the display may be provided for control of some of the desired functions but there is a trade-off between the number and specificity of such parameter specification operations that are available and the ease of accomplishing the same specifications via function keys or the alphanumeric keyboard itself. In the extreme case a "number pad" could be provided directly on the display to be used to designate numeric parameter settings by stylus, but the appropriateness of this choice will depend on the size of the display surface, the frequency with which such settings must be made and several other variables relating to efficient use of the display surface. If the complexity of the display requirements are such that a separate frame is needed to provide such a "light button" array, then it seems likely that shifting to a keyboard mode would be preferred in order to retain the active graphics on the display.

Another dimension of efficiency is the relative verbosity of the messages and command forms used. Here again there are no hard and fast rules. Two critical characteristics of the user enter into the decision. We need to know whether he is a casual or dedicated user, that is whether the interactive activities are a substantial part of his daily job and the same systems are used repeatedly. We also need to know the user's level of experience with the system. Since this latter characteristic is transient we need some means for coping with the range of levels of experience that will be encountered.

There are several means available for dealing with the issue of command or message verbosity. In a command-oriented system one procedure is to design the system to accept full command names or abbreviations with the choice between them left to the the user. A switch might be provided so that when the user enters the system he can set it appropriately, or, in a more sophisticated system the setting might be chosen automatically on the basis of performance data collected on the user for the purpose of determining his level of expertise.

A more adaptable alternative defines for each command a pair of synonyms, one concise and one verbose, each of which is acceptable at any time. This mode accommodates the case of the typical user who progresses in experience differentially with different parts of the system. As the abbreviations are learned they may be used. Some systems incorporate a command scanner that permits typing a full command or the minimal set of initial characters necessary to distinguish uniquely that command from other possible commands at the same level.

There are, of course, some applications, such as airlines or auto reservation systems, in which the frequency of use and demand for rapid interaction with a client justify considerable investment in training with a concise command set directly.

Paralleling the need for ways of dealing with users of different levels of experience with commands, is the need for similar adaptability in the presentation of computer-to-user messages, particularly messages signalling error conditions. Again choice of concise or verbose versions might be provided through a user profile switch or alternatively both (or multiple) levels of detail always could be made available. The user would first see the concise or abbreviated version but could, by a simple command, request a greater level of detail. With either scheme, careful design of the messages themselves is very important so that some critical point of interest is not lost in the verbosity of a detailed explanation.

Whatever means are considered for promoting user efficiency, the designer must recognize that the potential importance of efficiency interacts with many other system parameters. Thus, conciseness of presentation is more important for relatively slow printing terminals than for high bandwidth CRT devices and the efficacy of the approaches suggested is dependent on the anticipated latency from user request to computer response.

Keeping the User Informed

It has sometimes been stated as a conjecture that friendly interactive systems are those in which the user feels in control of the interaction, or at least in "touch" with the system, at all times. It is not clear, especially in the typical interactive environment in which the varying system load and unpredictable system crashes are beyond the user's control, how to foster this perception. An important component, however, is the provision of suitable feedback so that the user is continually informed about current system status and the options available at each point in time.

For each user action some response from the system should be provided. In many cases the result of the action is obvious from the changing display that results. In others it is not. An important component of the designer's art concerns the judgment about the nature, extent and complexity of the feedback that is appropriate to various classes of actions. Activities that cause superficial changes from the user's point of view require only simple cues that status has changed. These may be nothing more than repositioning a cursor or the change of a displayed parameter value. Activities that cause fundamental changes in a data base may require an indication of the ramifications of the requested change. It is important that feedback from these remote changes reflect that the requested action has actually been taken (when it has) and not simply a preprogrammed cue originating at the level of the dialogue control program. In any case the less obtrusive these acknowledgments can be made, the better. Some cases involving a long response time to carry out a complex calculation or chain of data base changes may require an immediate acknowledgment that the request has been processed as well as later feedback that the requested action has been accomplished. Also in cases in which the user's terminal is dormant awaiting completion of an action, periodic feedback informing the user that the system is still working on his problem is especially helpful. In a highly reliable system in which failure is infrequent it may be sufficient to provide the user with an interrupt key that provides information regarding current system status on demand without jeopardizing the system's progress on the task in question.

It is interesting to note the subtle interaction between user attentiveness to these feedback cues and user attitudes about system reliability. Novice users are usually suspicious of what the machine can do for them regardless of its real dependability. As experience accumulates they may become less dependent on such cues. However, for systems with heavy loads, long response times or frequent failures, the efficacy of such cues will persist regardless of the user's level of expertise.

Closely related to the need for a reaction to every action is the desirability of providing the user with a sensible next step at every point in the development of a transaction. A most frustrating state for the user to find himself in is the dead-end at which no possibility of further action is perceived. Certain actions should always be possible such as aborting the transaction and beginning again, interrupting a lengthy printout, deleting an ill-formed request, or editing a partially-entered data item. In addition there are many opportunities for prompting the user concerning possible useful next steps. Sometimes these prompts may take the form of a menu of options, but they may also be communicated by selectively illuminating the set of function keys that represent the current set of alternatives, by highlighting appropriate light buttons, by calling attention to items on the CRT through color, blinking, differential intensity or inverse video (black on white), by displaying current options in a special window dedicated to alerting functions or simply by repositioning a cursor to a new location that suggests the next class of actions that is available. These latter alternatives have in common the fact that they serve the purpose without cluttering the display by adding to the material already there. Also they can be accomplished without transitioning to a new frame of information.

Improving User Accuracy and Facilitating Error Handling

One of the primary advantages of introducing an interactive system is the potential gain in the quality of the data base that can result. Providing users with the opportunity to correct their own errors both at the time of initial data entry and through consistency checks based on their knowledge of the transaction in progress can produce better quality data and can promote the user's personal responsibility for accurate performance.

Error control involves both prevention and correction. Wherever possible the designer should anticipate possible points of confusion and provide feedback concerning them. Points of confusion cannot always be defined at the time of design, but a well-designed system may provide for the accumulation of statistics on errors committed and a retrospective analysis will provide the data needed to add alerting feedback concerning potential error conditions. Suitable mechanisms for providing such feedback might include on-line prompting regarding the format or syntax for an entry, or a help feature which permits the user to obtain an explanation of the most efficient way to achieve a desired result. Sometimes it may require nothing more than a change in wording or format to eliminate a source of confusion. A further way to eliminate errors as they occur is to provide a means of recovery from commands that require a considerable effort to undo. Critical actions should not be dependent on a single keystroke nor should they be one-key variants of other frequently used commands. Some systems require confirmation of commands that modify the data base or produce a lengthy printout. This procedure may not be very effective for dedicated users, however, because the confirmation can become automatic. A better means is to provide commands that abort a printout without disturbing prior work, undelete a deleted file, or undo a complete command sequence and restore the processing activity to an earlier state. Short of such sophisticated techniques, provision of a means for backtracking one step at a time through a completed interactive sequence, would seem to be a minimum requirement.

The other half of error control is the facilitation of diagnosis and correction of errors that inevitably will occur in spite of all efforts to prevent them. Perhaps the major gains over existing practices are to be made in better communication to the user of error conditions that can be detected by the computer system. Many users have seen and complained about the obscure and ambiguous error messages that programmers tolerate and that frequently filter down to users of production programs. The best error messages communicate three things: (1) the location of the error, (2) the nature of the error, and (3) some suggestion concerning how to correct it or how to find out how to correct it.

The location of an error in frame-oriented dialogues may be signalled by positioning the cursor to the display element in error, by blinking or highlighting the string in error, or in the case of printing terminals retyping the string in question.

With respect to error identification, the objective should be to express the error in terms understandable to the user and that do not require knowledge of the data structure or internal string manipulation commands. Error messages should be clear and concise and expressed in positive terms wherever possible.

"The entry should have four characters"
is preferred to
"You entered the wrong number of characters."

Error message should be as specific as possible. Among the alternatives:

(1) "Search for this entry produced no output."
(2) "Search for this requisitior produced no output."
(3) "Search for requisition 3723 produced no output."

The greater specificity of the third alternative is to be preferred whenever such specificity is feasible.

Avoid entertainment and personification. Alternative (3) above is preferred to, "Try again, I cannot find that requisition in my files." While this message seems friendly on first viewing, after a few repetitions it becomes tedious.

Recommending suitable means for correcting errors is more difficult, short of an intelligent system that understands the user's goals and intent at each point. However, in systems in which many of the dialogue sequences are predefined and common errors can be observed on the basis of pretesting or collection of error statistics as an integral part of the system design, error-correction messages may be formulated to cover the high frequency cases relatively easily. The message may suggest a direct recovery procedure, branch to a tutorial routine that will provide the needed information or, as a last resort, refer the user to a page or paragraph reference in a manual that is likely to provide the answer. The value of providing users with the resources to correct their own errors cannot be overestimated in terms of payoff in increased user motivation and acceptance of the system.

Integrating Training with Operation

The potential for using interactive systems in a training mode has been recognized since the earliest efforts in computer-assisted instruction. There is currently much innovative work underway to enhance the integration of interactive production programs with interactive training capabilities. Here we are not so concerned with the use of separate tutorial systems to teach system use, but rather with on-line user aids that facilitate recall of command terminology or syntax, and that provide assistance when the user would otherwise need to seek outside help. At a minimum a user should have the capability of displaying at anytime the options available at a given point in an interactive sequence. Typing of a question mark might call forth a listing of the command options available at that point.

Many systems incorporate a help or describe feature which, upon typing, "Describe X," will solicit a brief explanation of the command and provide an example of its arguments and syntax. The difficulty with such procedures is that if the user needs help he is likely not to know the name of the command that he needs. The general solution to this problem lies in a much more sophisticated system which incorporates a model of the user's context and goal structure to provide suggestions for the procedure he is undertaking. The nature of the desired information also depends in a complex way on the user's previous experience with the system and the number of times that particular procedure has been initiated in the past. The desired information could range from a detailed explanation of the command capability to a simple description of possible command arguments. Identification of the most effective on-line help procedures is a researchable problem. It has much in common with the problem of specifying effective error recovery procedures.

CHAPTER 7

PROBLEMS FOR FUTURE RESEARCH

In this chapter we list several research problems that appear, from the foregoing review, to be potential areas for research relevant to the design and operation of C^3 systems. This list is not intended to be exhaustive. In particular, no effort has been made to duplicate the many suggestions for research that came out of our workshops and are recorded in the workshop notes in Volume II.

*In spite of considerable research on the design of displays, it is still not known how information should be organized and sequenced during presentation in order to maximize assimilation and retention.

*The factors that determine where a person will tend to look on a complex visual display are not well understood, nor are those that determine what aspects of a display he will remember after having scanned it. More experimental work is needed to understand better both attention-getting and retention-assuring properties of visual displays.

*Inasmuch as motion is known to attract a viewer's attention to a part of a display where the motion occurs and the eye is particularly sensitive to motion even in its peripheral field, ways of exploiting these facts in the design of dynamic computer-driven visual displays should be investigated.

*The experimentation that has been done relating eye fixations to higher mental processes should be extended. The monitoring of eye movements, especially in response to "planted" stimuli, could prove to be a useful source of information regarding the observer´s level of alertness. Also given a greater understanding of their cognitive determinants, eye fixations could provide useful clues to cognitive activity.

*In general, the relationship between eye movements and cognitive activity deserves much more careful investigation. The finding, for example, that the frequency of eye movements seems to be correlated with the general level of cognitive activity is of considerable theoretical and practical interest. Also, the possibility that eye-movement monitoring could yield some insights into decision making and choice behavior deserves further investigation.

*A process that is not well understood is that by which human beings integrate into one coherent scene information that is acquired by what amounts to be a series of discrete visual snapshots. A better understanding of this process is necessary to insure an optimal design of computer-based display systems that permit a user to, in effect, explore large displays through a window of variable size. The problem is how to facilitate the maintenance of a fixed frame of reference for the viewer, and to assure that the information that is presented to him can be integrated into an accurate representation of the overall scene.

*Some studies should be done to investigate the relative merits of large-scale visual displays as compared with small display consoles located at individual work stations. Given the existence of computer-based displays, more attention should be given to the possibility of making the format and dynamics of a display contingent on the user´s task. (e.g., searching, browsing, monitoring, reading for comprehension).

*An attempt should be made to determine the feasibility of using eye fixations as inputs to computer-driven displays, i.e., to use the point at which the user is fixating as control information when he requests additional data regarding something on the display.

*A much better understanding is needed of how people assimilate, retain and use the kind of topological and geographical information that is typically contained in maps.

*The advent of computer-driven dynamic displays opens up many new possibilities for encoding and formatting information. Human factors guidelines for display designs have limited applicability to these displays. There is a need for parametric studies aimed at producing a new set of guidelines that take into account the degree of flexibility that computer-based systems provide.

*Some research should be addressed to the problem of associating specific types of errors and distortions of memory with methods for presenting information.

*There is a need for systematic study of the relative advantages and disadvantages of oral and nonoral methods of interperson communication. The need for such study is accentuated by the advent of computer-mediated message systems.

*A correlative need is for an investigation of how the existence of new methods for communication, such as computer-mediated message systems, are likely to effect the way in which operational groups are organized and interact.

*Both for purposes of encryption and bandwidth compression, many military communication systems of the future will make use of digitized voice. Some work on the evaluation of the quality of digitized speech is currently underway. What is not known is how good the quality has to be to be acceptable. Some work should be done to try to determine, for example, how important it will be for participants in a teleconferencing session to be able to recognize each other´s voices, and to receive the types of extra linguistic cues regarding such things as emotion and stress that are normally carried by the speech signal.

*More research should be done on the potential effectiveness of nonvoice methods of communication. In particular, the work of Chapanis and his colleagues in comparing speech and typewriting as communication modes should be extended to investigate the effectiveness of performance after extensive practice in communicating via typewriting.

*In a similar vein, studies should be undertaken to determine how the mode of communication may affect communicating strategies and habits over extended periods of time.

*A study that could provide some insights into how to design better interactive languages would be one in which people were forced to communicate via typewriter over a significant period of time and were permitted to evolve their own conventions regarding syntax, abbreviations, coded messages, and so on.

*Studies of the utility of instrumentally-speeded speech for specific application such as monitoring tape recordings for specific content, reviewing recorded briefing sessions, presentation of material for "refresher" purposes should be conducted.

*The utility of using rapidly presented sequences of displays for purposes of monitoring for specific content or for memory refreshing should be explored.

*Another goal for research should be the development of more effective computer-based algorithms for time-compressing speech.

*There is a need for more adequate models in terms of which to quantify the "costs of observing" operational displays and the "value of the information" obtained from them. There is also a need to extend theories of vigilance and monitoring that have been developed in the context of vigilance and monitoring that have been developed in the context of signal-detection tasks to the domain of more cognitively demanding tasks in which the requirement is to detect significant changes in complex, multivariate situations.

*In spite of the considerable amount of research on divided attention, relatively little is yet known about how an individual allocates his attention to competing demands in real-life situations. A critical question in the context of C^3 operations is that of how the individual copes with situations in which the demands exceed his capacity so that he must ignore some of them if he is to continue functioning at all.

*Little is known about the distinctive features that commanders use to classify military situations in terms of conventional tactical categories. Some experimentation with multidimensional techniques might provide some useful insights into the process of pattern recognition and classification in this context.

*The question remains whether people can be trained to be better decision makers in general, or whether they can only be trained to function better as decision makers in specific contexts.

*Numerous decision-aiding systems have been developed, at least on an experimental basis. It appears, however, that relatively few of them are being used in operational situations. An attempt to determine why some systems are used and others are not should provide some useful clues to effective decision-aiding system design.

*The question of how accurately people can judge their own knowledge and whether they can be taught to improve their accuracy in this respect is worth continuing attention.

*Critical incident analyses of operational problems that have arisen in various types of C^3 systems and a cataloging of the results would be useful.

*An effort should be made to develop a methodology for studying goal and reward structures at various levels within systems. What is particularly crucial is to be able to determine whether goals and incentives that operate at one level within the system are consistent with those at higher levels.

*The need for better performance evaluation methods at all levels of system operation is acute.

*The problem of information overload appears to remain a critical one in many C^3 situations. The need for better filtering, organizing, and sequencing techniques is apparent.

*The problem of facilitating the explication of decision makers' value spaces remains a critical one.

*The role of the system operator is becoming increasingly one of monitor and supervisor. The problem of maintaining his skills so that he can perform effectively in case of system failure becomes increasingly difficult. The development of techniques for maintaining proficiency should be a key objective of training research.

*There is need for better techniques to permit the user of a data management system to browse through a data base.

*There is a need for better interactive techniques to help individuals formulate the problems on which they are working. The need in this case is not so much to provide help in solving problems but help in conceptualizing them and identifying the dimensions of the problem space.

*The existing evidence that suggests that decision makers may perform less effectively when losing than when winning in a conflict situation is particularly relevant to C^3 operations. Further research on this issue should have a high priority.

*Training techniques should be sought that would better prepare decision makers to deal with unanticipated one-of-a-kind events.

*Computer-mediated communication should be explored with respect to its potential for facilitating conflict resolution. Some thought should be given to the development of some computer-based remediational techniques for experimentation with some of the classical conflict paradigms.

*The problem of credibility is a particularly important one and it has many dimensions. One of the more important aspects of this problem is that of finding ways for adversaries in a conflict to be able to be credible to each other when telling the truth.

*A variety of biases has been identified in the way human beings function as intuitive statisticians and logicians. There is a need for research aimed at developing training procedures and/or decision aids that would correct for such biases.

*Relatively little is known about the determinants of human belief states. There is no theory from which one could predict, for example, the degree of plausibility that an individual would attach to any given assertion. Inasmuch as people presumably act on their beliefs about the state of the world and the potential consequences of their decisions, such knowledge is badly needed.

*There is a continuing need for better reasoning aids. Included among the type of assistance that such aids might provide for the user are the following: helping make the tacit assumptions in a reasoning process explicit, providing information that is germane to the problem on which the reasoning is being done, helping to discover logical inconsistencies or inadequacies in an argument, keeping track of unevaluated hypotheses and tentative conclusions.

*There is also a continuing need for better predictive models that will help a decision maker explore the probable or potential consequences of alternative courses of action that he may wish to consider taking.

*A major problem of many decision makers, problem solvers, and policy makers is that of maintaining their own personal data base in a form that maximizes the availability of critical pieces of information when they are needed. Studies aimed at determining how best to organize and manage personal data bases could have significant implication for the performance of people in command positions.

*There is a need to understand better the effects of acute and prolonged stress on human performance and decision making. Unfortunately, the existing literature on stress is fragmented and difficult to interpret. Moreover, it is very difficult, if not impossible, to simulate in a laboratory the type of stressful situations that are likely to be encountered in C^3 systems.

*Techniques for monitoring the degree of stress under which a given individual is operating at a particular time are also needed.

*The problem of predicting how well any given individual will perform under stressful situations remains an important challenge.

*There is a need for more research on how various forms of chronic stress (including low-level stress) affect performance and on how these effects interact with those of acute stressors.

*There is a need for further research aimed at identifying effective stress alleviators, both those that can be applied during a stressful situation and those that can be applied following such situations.

*A major problem is that of fostering within the user of an interactive computer system an accurate appreciation for the capabilities and limitations of the tool that he is using. Users and would-be users of computer systems tend to err in both directions in their assessments of what computer systems can and cannot do.

*Apparently, one of the difficulties in obtaining acceptance of computer-based systems by military commanders is lack of confidence in the reliability of their outputs. To the extent that this lack of confidence is unjustified, research on how best to allay the potential users' apprehensions would be useful.

*More research should be done on the ability of people to communicate (both with each other as well as with the computer) with quasi natural but constrained languages.

*The potential applications of computer graphics and touch terminals as input devices in C^3 systems seems great and worth considerable experimentation.

*Some research is needed on the question of how an individual's cognitive representation of a computer system affects his utilization of it.

*Developers of interactive computer systems have incorporated many features for the purposes of making their systems easier to use and less susceptible to human error. A cataloging of such features, and a program of research to test their effectiveness would advance the state of the art of person-computer interaction.

*There continues to be a need to develop computer input terminals that are better suited to users' natural (not necessarily verbal) communication skills.

*An attempt should be made to design a work space, or work environment for an individual who makes extensive use of computer based tools (displays, information systems, decision aids, computer-based message systems, document-preparation tools) in his day to day activities. The layout of such a work space might differ radically from conventional work spaces.

ACKNOWLEDGMENTS

Several people contributed significantly to this project other than the authors of this report. The authors wish to express their appreciation, in particular, to: Charlene Long and Jessie Kurzon for tracking down much of the literature that was reviewed; Elsie Leavitt for handling the logistics of the workshops; Florence Maurer for organizing references and overseeing the typing of the final report; and Anne Kerwin, Gail Rushton, and Mildred Webster for typing various parts of the final report and the several drafts that preceded it.

REFERENCES

Abrams, M. H., Goffard, S. J., Kryter, K. D., Miller, G. A., Sanford, J., & Sanford, F. H. Speech in noise: A study of factors determining its intelligibility. Harvard University, Psychoacoustics Laboratory, 1944. OSRD No. 4023, PB 19805.

Adams, J. A. (Ed.). Preface to Special Issue on "Vigilance." Human Factors, 1965, 7(2), 91.

Adams, J. & Cohen, E. Time-sharing vs. instant batch processing: An experiment in programming training. Computers and Automation, March 1969, 30-34.

Adams, S. K. Decision making in quality control: Some perceptual and behavioral considerations. In C. G. Drury & J. G. Fox (Eds.), Human reliability in quality control. New York: Halsted Press, 1975, 55-69.

Alker, H. A. & Kogan, N. Effects of norm-oriented group discussion on individual verbal risk-taking and conservatism. Human Relations, 1968, 21, 393-405.

Allen, R. W., Schwartz, S. H., & Jex, H. R. Driver decision-making research in a laboratory simulation. Presented at NATO Conference on Monitoring Behavior and Supervisory Control, Berchtesgaden (F.R.) Germany, March 8-12, 1976, Preprints pp. 49-59.

Allport, D. A The rate of assimilation of visual information. Psychonomic Science, 1968, 12, 231-232.

Allport, D. A., Antonis, B., & Reynolds, P. On the division of attention: A disproof of the single channel hypothesis. Quarterly Journal of Experimental Psychology, 1972, 24, 225-235.

Alluisi, E. A., Coates, G. D., & Morgan, B. B. Effects of temporal stressors on vigilance and information processing. Paper presented at NATO Conference on Vigilance II: Relationships among Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy, August 3-6, 1976. To be published in a volume with approximately that title, R. R. Mackie (Ed.), New York: Plenum (in press).

Alpern, M. Effector mechanisms in vision. In J. W. Kling and L. A. Riggs (Eds.), Woodworth & Schlossberg's experimental psychology. New York: Holt, Rinehart, & Winston, 1971.

Amadeo, M. & Gomez, E. Eye movements, attention and dreaming in subjects with life-long blindness. Canadian Psychiatric Association Journal, 1966, 11, 501-507.

Anderson, L. L., Hellstrom, B., & Eide, R. Strenuous muscular exertion in the polar climate. Ergonomics, 1968, 11(3), 261-273.

Antes, J. R. The time course of picture viewing. Journal of Experimental Psychology, 1974, 103, 62-70.

Antrobus, J. S., Antrobus, J. S., & Singer, J. L. Eye movements accompanying daydreaming, visual imagery, and thought suppression. Journal of Abnormal and Social Psychology, 1964, 69, 244-252.

Atkinson, R. C. & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), Advances in the psychology of learning and motivation (Vol. 2). New York: Academic Press, 1968.

Atkinson, R. C. & Shiffrin, R. M. The control of short-term memory. Scientific American, 1971, 224, 82-90.

Attneave, F. Physical determinants of the judged complexity of shapes. Journal of Experimental Psychology, 1957, 53, 221-227.

Atwood, G. An experimental study of visual imagination and memory. Cognitive Psychology, 1971, 2, 290-299.

Averbach, E. & Coriell, A. S. Short-term memory in vision. Bell System Technical Journal, 1961, 40, 309-328.

Baddeley, A. D. Short term memory for word sequences as a function of acoustic, semantic, and formal similarities. Quarterly Journal of Experimental Psychology, 1966, 78, 362-365.

Baddeley, A. D. Estimating the short-term component in free recall. British Journal of Psychology, 1970a, 61, 13-15.

Baddeley, A. D. Effects of acoustic and semantic similarities on short-term paired associate learning. British Journal of Psychology, 1970b, 61, 335-343.

Bateson, N. Familiarization, group discussion, and risk taking. Journal of Experimental Social Psychology, 1966, 2, 119-129.

Beasley, D. S., Schwimmer, S., & Rintelmann, W. F. Intelligibility of time-compressed CNC monosyllables. Journal of Speech and Hearing Research, 1972, 15, 340-350.

Becker, G. M. & McClintock, G. G. Value: Behavioral decision theory. Annual Review of Psychology, 1967, 18, 239-286.

Bell, P. R. & Jamieson, B. D. Publicity of initial decisions and the risky shift phenomenon. Journal of Experimental Social Psychology, 1970, 6, 329-345.

Belovicz, M. W. & Finch, F. E. A critical analysis of the "risky shift" phenomenon. Organizational Behavior and Human Performance, 1971, 6, 150-168.

Bencomo, A. A. & Daniel, T. C. Recognition latency for pictures and words as a function of encoded-feature similarity. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 119-125.

Berkun, M. M., Bialek, H. M., Kern, R. P., & Yagi, K. Experimental studies of psychological stress in man. Psychological Monographs, 1962, 76(15), No. 534.

Berlyne, D. E. The influence of complexity and novelty in visual figures on orienting responses. Journal of Experimental Psychology, 1958, 55, 289-296.

Berlyne, D. E. Conflict, arousal and curiosity. New York: McGraw-Hill, 1960.

Berlyne, D. E. Curiosity and exploration. Science, 1966, 153, 25-33.

Bieber, R. A., Margroff, D. H., & Berkhout, J. The measurement of drivers' performance in stressful driving situations. Final Report, November 1973, No. PB-236 439/6GA.

Biederman, I. Perceiving real-world scenes. Science, 1972, 177, 77-79.

Biederman, I., Glass, A. L., & Stacy, E. W. Searching for objects in real-world scenes. Journal of Experimental Psychology, 1973, 97, 22-27.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W., Jr. On the information extracted from a glance at a scene. Journal of Experimental Psychology, 1974, 103, 597-600.

Birdsall, T. G. & Roberts, R. A. On the theory of signal detectability: An optimum nonsequential observation-decision procedure. IEEE Transactions on Information Theory, 1965a, IT-11, (2), 195-204.

Birdsall, T. G. & Roberts, R. A. Theory of signal detectability: Deferred-decision theory. Journal of the Acoustical Society of America, 1965b, 37, 1064-1074.

Bizzi, E. The coordination of eye-head movements. Scientific American, 1974, 231, 100-106.

Blackledge, M. A. Consideration of additional nuclear effects and fratricide avoidance in QUICK. NMCSSC, System Planning Manual SPM FD 89-74, May 1974.

Blount, P., Holmes, J., Rodger, J., & Coltheart, M. On the ability to discriminate original from mirror-image reproductions of works of art. Perception, 1975, 4, 385-389.

Boehm, B. W., Seven, M. J., & Watson, R. A. Interactive problem-solving - an experimental study of "lockout" effects. Proceedings Spring Joint Computer Conference, Interactive Problem Solving, 1971, 206-210.

Bourne, L. E., Ekstrand, B. R., & Dominowski, R. L. The psychology of thinking. Englewood Cliffs, N.J.: Prentice Hall, 1971.

Bowen, R. J., Feehrer, C. E., Nickerson, R. S., Spooner, R. L., & Triggs, T. J. Army tactical intelligence: An examination of computer-based displays and a structured decision procedure as aids to the intelligence analyst. BBN Report 2073, under Contract DAHC-19-69-C-0020, U.S. Army Behavior and Systems Research Lab., 1971.

Bowen, R. J., Feehrer, C. E., Nickerson, R. S., & Triggs, T. J. Computer-based displays as aids in the production of army tactical intelligence. USA Research Institute for the Behavioral and Social Sciences Technical Paper 258, February 1975.

Bower, G. H. Analysis of a mnemonic device. American Scientist, 1970, 58, 496-510.

Bower, G. H. Mental imagery and associative learning. In L. W. Gregg (Ed.), Cognition in learning and memory. New York: John Wiley & Sons, 1972.

Brand, J. Classification without identification in visual search. Quarterly Journal of Experimental Psychology, 1971, 23, 178-186.

Bregman, A. S. & Campbell, J. Primary auditory stream segregation and perception of order in rapid sequence of tones. Journal of Experimental Psychology, 1971, 89, 244-249.

Briscoe, H. R., Jr. Audio information rate changing technology: An accelerated briefing and training technique for senior officers. NTIS National Technical Information Service AD-778-878, 20 Oct. 1973.

Broadbent, D. E. Decision and stress. New York: Academic Press, 1971.

Brooks, L. R. The suppression of visualization by reading. Quarterly Journal of Psychology, 1967, 19, 289-299.

Brooks, L. R. Spatial and verbal components of the act of recall. Canadian Journal of Psychology, 1968, 22, 349-368.

Brown, J. S., Burton, R. R., Bell, A. G., & Bobrow, R. J. SOPHIE: A sophisticated instructional environment. Air Force Human Resources Laboratory Technical Report TR-74-93, December 1974.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. A study of thinking. New York: John Wiley & Sons, 1956.

Bryan, G. E. JOSS: 20,000 hours at a console - a statistical summary. AFIPS Conference Proceedings, Vol. 31, 1967 Fall Joint Computer Conference, pp. 679-777.

Buck, J. R. Dynamic visual inspection: Task factors, theory and economics. In C. G. Drury & J. G. Fox (Eds.), Human reliability in quality control. New York: Halsted Press, 1975, 165-186.

Buswell, G. T. How people look at pictures. Chicago: University of Chicago Press, 1935.

Buzzard, R. B. A practical look at industrial stress. Occupational Psychology, 1973, 47(1-2), 51-61.

Caldwell, E. C. & Hall, V. C. Distinctive-features versus prototype learning reexamined. Journal of Experimental Psychology, 1970, 83, 7-12.

Callahan, N. D. & Grace, G. L. AUTODOC: Computer-based assistance for document production. Proceedings of the Association for Computing Machinery National Meeting, 1967, 117-185.

Cameron, C. A theory of fatigue. In A. T. Welford (Ed.), Man under stress, 1974, 67-82.

Caplan, R. Personal communication. ARPA Conference on Stress Effects in C3 Systems, Los Angeles, August 1976.

Carbonell, J. A queueing model of many-instrument visual sampling. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7, (4), 157-164.

Carbonell, J. R. On man-computer interaction: A model and some related issues. BBN Report No. 1593, USAFCRL-68-0053, September 1967.

Carbonell, J. R., Elkind, J. L., & Nickerson, R. S. On the psychological importance of time in a time-sharing system. Human Factors, 1968, 10, 135-142.

Carpenter, P. A. & Just, M. A. Reading comprehension as eyes see it. To appear in M. A. Just and P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, in press.

Carterette, E. D., Friedman, M. P., & Cosmides, R. Reaction-time distributions in the detection of weak signals in noise. Journal of the Acoustical Society of America, 1965, 38, 531-542.

Cartwright, D. Risk taking by individuals and groups: An assessment of research employing choice dilemmas. Journal of Personality & Social Psychology 1971, 20, 361-378.

Castleman, P. A., Russell, C. H., Webb, F. N., Hollister, C. A., Siegel, J. R., Zdonik, S. R., & Fram, D. M. The implementation of the PROPHET system. AFIPS Conference Proceedings, 1974, 43, 457-468.

Castleman, P. A., Whitehead, S. F., Sher, L. D., Hantman, L. M., & Massey, L. D., Jr. An assessment of the utility of computer aids in the physician's office. BBN Report No. 3096, 1974.

Cattell, J. The inertia of the eye and brain. Brain, 1885, 8, 295-312. In A. T. Poffenberger, James McKeen Cattell: Man of Science. York, PA: Science Press, 1947.

Chapanis, A. Prelude to 2001: Exploration in Human Communication. American Psychologist, 1971, 26, 949-961.

Chapanis, A. The communication of factual information through various channels. Reprinted from Information Storage and Retrieval, 1973, 9, 215-231. Great Britain: Pergamon Press.

Chapanis, A. Interactive human communication. Scientific American, 1975, 232, 36-42.

Chapanis, A., Ochsman, R. B., Parrish, R. N., & Weeks, G. D. Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem solving. Human Factors, 1972, 14, 487-509.

Chapanis, A. & Overbey, C. M. Studies in interactive communication: III. Effects of similar and dissimilar communication channels and two interchange options on team problem solving. Perceptual and Motor Skills Monograph Supplement 2-V38, 1974 (whole) 344-374.

Chapanis, A., Parrish, R. N., Ochsman, R. B., & Weeks, G. D. Studies in communication: II. The effects of four communication modes on the linguistic performance of teams during cooperative problem solving. Human Factors, in press.

Chapman, D. E. & Sinclair, M. A. Ergonomics in inspection tasks in the food industry. In C. G. Drury & J. G. Fox (Eds.), Human reliability in quality control. New York: Halsted Press, 1975, 231-251.

Chase, W. G. & Calfee, R. C. Modality and similarity effects in short-term recognition memory. Journal of Experimental Psychology, 1969, 81, 510-514.

Chase, W. G. & Simon, H. A. The mind's eye in chess. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973.

Chernoff, H. The use of faces to represent points in n-dimensional space graphically. Journal of the American Statistical Association, 1973, 68, 361-368.

Chi, M. R. H. & Chase, W. G. Effects of modality and similarity on context recall. Journal of Experimental Psychology, 1972, 96, 219-222.

Chiles, W. D. Experimental studies of prolonged wakefulness. Technical Report 55-395, Wright Air Development Center, Dayton, Ohio, 1955.

Cimbalo, R. S. & Laughery, K. R. Short-term memory: Effects of auditory and visual similarity. Psychonomic Science, 1967, 8, 57-58.

Clark, S. E. Retrieval of color information from the preperceptual storage system. Journal of Experimental Psychology, 1969, 82, 263-266.

Cofer, C. N. & Appley, M. H. Motivation: Theory and research. New York: John Wiley & Sons, 1964.

Collins, B. E. & Guetzkow, H. A social psychology of group processes for decision making. New York: John Wiley & Sons, 1964.

Colquhoun, W. P. Simultaneous monitoring of a number of auditory sonar outputs. Presented at NATO Conference on Vigilance II: Relationships among Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy, August 3-6, 1976. To be published in a volume with approximately that title, R. R. Mackie (Ed.), New York: Plenum (in press).

Conrad, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-84.

Conrad, R., Freeman, P. R., & Hull, A. J. Acoustic factors versus language factors in short-term memory. Psychonomic Science, 1965, 3, 57-58.

Conrad, R. & Hull, A. J. Information, acoustic confusion and memory span. British Journal of Psychology, 1964, 55, 429-432.

Cooper, R. M. The control of eye fixation by meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory and language processing. Cognitive Psychology, 1974, 6, 84-107.

Craig, A. & Colquhoun, W. P. Vigilance effects in complex inspection. Presented at NATO Conference on Vigilance II:

Relationships among Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy, August 3-6, 1976. To be published in a volume with approximately that title, R. R. Mackie (Ed.), New York: Plenum (in press).

Craik, F. M. Modality effects in short-term store. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 658-664.

Cramer, H. L. & Talambiras, R. P. The braided speech method of time compressed speech. In E. Foulke (Ed.), Proceedings of the Second Louisville Conference on Rate and/or Frequency Controlled Speech. Louisville, KY: Center for Rate Controlled Recordings, Perceptual Alternatives Laboratory, 1971.

Dailey, K. A. Compressed speech: Potential application for Air Force technical training. AF Human Resources Laboratory, April 1974. Final Report AFHRL-TR-74-27, pp. 20.

Dainoff, M. J. Time course of visual and auditory encoding. Journal of Experimental Psychology, 1970, 86, 214-224.

Dallet, K., Wilcox, S. C., & D'Andrea, L. Picture memory experiments. Journal of Experimental Psychology, 1968, 76, 312-320.

Dambro, M. R., Page, R. L., & Pellicciotto, A. M. QUICK/HIS 6080 remote terminal utilization: Functional description. National Military Command System Support Center/System Sciences, Inc. METIS Corp. Distributed by National Technical Information Service, U.S. Dept. of Commerce, 1974.

Damodaran, L. The role of user support--actual and potential. Paper presented at the Advanced Study Institute on Man-Computer Interaction, Mati, Greece, September 1976.

Daniel, T. C. Nature of the effect of verbal labels on recognition memory for form. Journal of Experimental Psychology, 1972, 96, 152-157.

Daniloff, R., Shriner, T., & Zemlin, W. Intelligibility of vowels altered in duration and frequency. Journal of the Acoustical Society of America, 1968, 44, 700-707.

Davies, D. R. & Jones, D. M. The effect of noise and incentives upon attention in short-term memory. British Journal of Psychology, 1975, 66(1), 61-68.

Davies, D. R. & Parasuraman, R. Cortical evoked potentials and vigilance: A decision theory analysis. Presented at NATO Conference on Vigilance II: Relationships among Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy, August 3-6, 1976. To be published in a volume with approximately that title, R. R. Mackie (Ed.), New York: Plenum (in press).

Davies, D. R. & Tune, C. S. Human vigilance performance. New York: American Elsevier, 1969.

Davis, J. H., Laughlin, P. R., & Komorita, S. The social psychology of small groups: Cooperative and mixed-motive interaction. Annual Review of Psychology, 1976, 27, 501-541.

DeGreene, K. B. (Ed.). Systems psychology. New York: McGraw-Hill, 1970.

DeGroot, A. D. Thought and choice in chess. The Hague: Mouton, 1965.

Dewey, J. How we think. Boston: Heath, 1910.

Diehl, C. F., White, R. C., & Burk, K. Rate and communication. Speech Monograph, 1959, 26, 229-232.

Dion, K. L., Baron, R. S., & Miller, N. Why do groups make riskier decisions than individuals? In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 5). New York: Academic Press, 1970.

Dirlam, D. K. Most efficient chunk sizes. Cognitive Psychology, 1972, 3, 355-360.

Dodson, J. D. Simulation system design for a TEAS simulation research facility. Planning Research Corp., Los Angeles, Nov. 15, 1961.

Doost, R. & Turvey, M. T. Iconic memory and central processing capacity. Perception and Psychophysics, 1971, 9, 269-274.

Dornbush, R. L. Input variables in bisensory memory. Perception & Psychophysics, 1968, 4, 41-44.

Dornbush, R. L. Attention in bisensory simultaneous short-term memory. Perception & Psychophysics, 1970, 7, 244-246.

Dornbush, R. L. Delay of auditory input in "simultaneous" auditory and visual short-term memory. Perception & Psychophysics, 1971, 9, 97-98.

Drury, C. G. The effect of speed of working on industrial inspection accuracy. Applied Ergonomics, 1973, 4(1), 2-7.

Drury, C. G. Human decision making in quality control. In C. G. Drury & J. G. Fox (Eds.), Human reliability in quality control. New York: Halsted Press, 1975, 45-53.

Drury, C. G. & Addison, J. L. An industrial study of the effects of feedback and fault density on inspection performance. Ergonomics, 1973, 16(2), 159-169.

Duffy, E. Activation and behavior. New York: John Wiley & Sons, 1962.

Duncker, K. On problem solving. Psychological Monographs, 1945, 270.

Dunn, R. M. Built-in system ego-states and the transactions they induce. Position Paper ACM/SIGGRAPH workshop on User-Oriented Design of Interactive Graphics Systems, October 1976.

Eason, K. D. A task-tool analysis of manager-computer interaction. Proceedings of a NATO Advanced Study Institute, Mati, Greece, September 1976 (in preparation).

Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 1959, 66, 183-201.

Ebeling, H. W. The CHIMP manual. MIS Department, Bolt Beranek and Newman Inc., May 1974.

Edwards, D. C. & Goolkasian, P. A. Peripheral vision location and kinds of processing. Journal of Experimental Psychology, 1974, 102, 244-249.

Edwards, W. Probabilistic information processing system for diagnosis and action selection. In J. Spiegel & D. Walker (Eds.), Information system sciences, Proceedings of the Second Congress. Washington, D.C.: Spartan Books, Inc., 1965, 141-155.

Egan, J. P., Greenberg, C. Z., & Schulman, A. I. Operating characteristics, signal detectability, and the method of free response. Journal of the Acoustical Society of America, 1961, 33, 993-1007.

Eisenstadt, M. & Kareev, Y. Aspects of human problem solving: The use of internal representations. In D. A. Norman and D. E. Rumelhart (Eds.), Explorations in cognition. San Francisco: W. H. Freeman, 1975.

Embrey, D. E. Training the inspector's sensitivity and response strategy. In C. G. Drury & J. G. Fox (Eds.), Human reliability in quality control. New York: Halsted Press, 1975, 123-131.

Emmerich, D. S., Gray, J. L., Watson, C. S., & Tanis, D. C. Response latency, confidence, and ROC's in auditory signal detection. Perception & Psychophysics, 1972, 11, 65-72.

Engel, F. L. Visual conspicuity, directed attention, and retinal locus. Vision Research, 1971, 11, 563-576.

Engel, S. E. & Granda, R. E. Guidelines for man/display interfaces. IBM Technical Report TR00.2720. Poughkeepsie Laboratory, December 1975.

Englebart, D. C. & English, W. K. A research center for augmenting human intellect. American Federation of Information Processing Societies Conference Proceedings, 1968, 33, 395-410.

Erdelyi, M. H. & Becker, J. Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. Cognitive Psychology, 1974, 6, 159-171.

Estes, W. K. The locus of inferential and perceptual processes in letter identification. Journal of Experimental Psychology, 1975, 104, 122-145.

Estes, W. K. & Taylor, H. A. Visual detection in relation to display size and redundancy of critical elements. Perception and Psychophysics, 1966, 1, 9-16.

Fairbanks, G., Everitt, W. L., & Jaeger, R. P. Method for time or frequency compression-expansion of speech. Transactions of the Institute of Radio Engineers Professional Group on Audio, 1954, AU-2, 1954, 7-12.

Fairbanks, G., Guttman, N., & Miron, M. Effects of time compression upon the comprehension of connected speech. Journal of Speech and Hearing Disorders, 1957, 22, 10-19.

Fairbanks, G. & Kodman, F. Word intelligibility as a function of time compression. Journal of the Acoustical Society of America, 1957, 29, 636-644.

Finkel, D. L. A developmental comparison of the processing of two types of visual information. Journal of Experimental Child Psychology, 1973, 16, 250-266.

Fischhoff, B. Hindsight: Thinking backward? Oregon Research Institute Research Monograph, 1974a, 14(1).

Fischhoff, B. Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. Oregon Research Institute Research Monograph, 1974b, 14(3).

Fischhoff, B. & Beyth, R. "I knew it would happen" - Remembered probabilities of once-future things. ONR Technical Report No. 197- 026. Also, Oregon Research Institute Research Bulletin, 1974, 4(14).

Fletcher, H. Speech and hearing. New York: Van Nostrand Co., 1929, 293-294.

Foley, J. D. & Wallace, V. L. The art of natural graphic man-machine conversation. Proceedings of the IEEE, 1974, 62, 462-471.

Foulke, E. Listening comprehension as a function of a word rate. The Journal of Communication, 1968, 18, 198-206.

Foulke, E. The perception of time compressed speech. In D. Horton & J. Jenkins (Eds.), The perception of language. Columbus, Ohio: Merrill, 1971.

Foulke, E., Amster, C. H., Nolan, C. Y., & Bixler, R. H. The comprehension of rapid speech by the blind. Exceptional Children, 1962, 29, 134-141.

Foulke, E. & Sticht, T. G. Listening rate preferences of college students for literary material of moderate difficulty. Journal of Speech and Hearing Research, 1966, 6, 397-401.

Foulkes, D. The psychology of sleep. New York: Charles Scribner's Sons, 1966.

Frederiksen, J. R. Survey of the state of the art in human factors in computers. Consulting report prepared for ARPA/IPTO Program in Intelligent Terminals, July 1975.

French, N. R., Carter, C. W., & Koenig, W. The words and sounds of telephone conversations. Bell Telephone Technical Journal, 1930, 9, 290-324.

Friedman, L., Orr, D. B., Freedle, R. O., & Norris, C. M. Further research on speeded speech as an educational medium. Progress Report No. 2. Grant No. 7-48-7670-267. U.S. Office of Education, American Institutes for Research, Washington Offices Communication Research Program, (AIR-E-50-1/66-TR(2)), January 1966.

Funaro, J. F. An empirical analysis of five descriptive models for cascaded inference. Rice University, Department of Psychology, Research Report No. 74-2, May 1974.

Gardner, G. T. Evidence for independent parallel channels in tachistoscopic perception. Cognitive Psychology, 1973, 4, 130-155.

Garvey, W. D. The intelligibility of speeded speech. Journal of Experimental Psychology, 1953, 45, 102-108.

Garvey, W. D. & Henneman, R. H. Practical limits of speeded speech. United States Air Force Technical Report No. 5917. Dayton: Wright-Patterson Air Force Base, 1950.

Gerber, S. E. & Scott, R. J. Dichotic speech-time compression. Chapter V. in E. Foulke (Ed.), Proceedings of the second Louisville conference on rate and/or frequency-controlled speech, February 1971.

Gersten, J. C., Langner, T. S., Eisenberg, J. G., & Orzeck, L. Child behavior and life events: Undesirable change or change per se? Chapter 10 in B. S. & B. P. Dohrenwend (Eds.), Stressful life events: Their nature and effects. New York: John Wiley & Sons, 1974.

Gettys, C. F., Kelly, C. W. & Peterson, C. R. Best guess hypothesis in multi-stage inference. Organizational Behavior and Human Performance, 1973, 10, 364-373.

Gettys, C. F. & Wilke, T. A. The application of Bayes's theorem when the true data state is uncertain. Organizational Behavior and Human Performance, 1969, 4, 125-141.

Gibson, J. J. The perception of the visual world. Boston: Houghton-Mifflin, 1950.

Gibson, J. J. The senses considered as perceptual systems. Boston: Houghton-Mifflin, 1966.

Gibson, R. S. & Nicol, E. H. The modifiability of decisions made in a changing environment. ESD-TR-64-657, December 1964.

Glanville, A. D. & Dallenbach, K. M. The range of attention. American Journal of Psychology, 1929, 41, 207-236.

Glanzer, M. & Clark, W. H. Accuracy of perceptual recall: An analysis of organization. Journal of Verbal Learning and Verbal Behavior, 1963, 1, 289-299.

Glanzer, M. & Cunitz, A. R. Two storage mechanisms in free recall. Journal of Verbal Learning and Verbal Behavior, 1966, 1, 289-299.

Glass, D. C. & Singer, J. E. Urban stress. New York: Academic Press, 1972.

Gold, M. Methodology for evaluating time-shared computer usage. Ph.D. dissertation, Massachusetts Institute of Technology, Alfred P. Sloan School of Management, 1967.

Goldstein, H. Reading and listening comprehension at various controlled rates. Teachers College, Columbia University Contributions to Education, No. 821. New York, Bureau of Publications, Teachers College, 1940.

Goldstein, H. Reading and listening comprehension at various controlled rates. Teachers College contributions to education No. 821. New York: Bureau of Publications, Columbia University, Teachers College, 1940.

Gopher, D. Patterns of eye movement in auditory tasks of selective attention. Unpublished Ph.D. thesis, Hebrew University, Jerusalem, 1971. Cited in Kahneman, 1973.

Gould, J. D. Pattern recognition and eye-movement parameters. Perception and Psychophysics, 1967, 9, 399-407.

Gould, J. D. Eye movements during visual search. IBM Research Report RC-2680 (IBM Research Center, Yorktown Heights, N.Y.), July 1969.

Gould, J. D. Some psychological evidence on how people debug computer programs. International Journal of Man-Machine Studies, 1975, 7(2).

Gould, J. D. Looking at pictures. In R. A. Monty & J. W. Senders (Eds.), Eye movements and psychological processes. Hillsdale, N.J.: Lawrence Erlbaum, 1976.

Gould, J. D. & Dill, A. B. Eye-movement parameters and pattern discrimination. Perception and Psychophysics, 1969, 6, 311-320.

Gould, J. D. & Drongowski, P. An exploratory study of computer program debugging. Human Factors, 1974, 16(3).

Gould, J. D. & Peeples, D. R. Eye movements during visual search and discrimination of meaningless, symbol, and object patterns. Journal of Experimental Psychology, 1970, 85, 51-55.

Gould, J. D. & Schaffer, A. Eye-movement patterns in scanning numeric displays. Perceptual and Motor Skills, 1965, 20, 521-535.

Gould, J. D. & Schaffer, A. Eye movement parameters in pattern recognition. Journal of Experimental Psychology, 1967, 74, 225-229.

Graveling, R. A. & Brooke, J. D. An analysis of stress and the characteristics of the human response to stress. In J. D. Brooke (Ed.), British proceedings of sports psychology. British Society of Sports Psychology, University of Salford, 1974, 203-211.

Green, D. M. Temporal auditory acuity. Psychological Review, 1971, 78, 540-551.

Green, D. M. Personal communication. ARPA Conference on Stress Effects in C3 Systems, Los Angeles, August, 1976.

Green, D. M. "Unpublished work," 1976.

Green, D. M. & Birdsall, T. G. The effect of vocabulary on articulation score. In J. A. Swets (Ed.), Signal detection and recognition by human observers. New York: John Wiley & Sons, Inc. 1964, 609-619.

Green, D. M. & Henning, G. B. Audition. Annual Review of Psychology, 1969, 20, 105-128.

Green, D. M. & Swets, J. A. Signal detection theory and psychophysics. New York: John Wiley & Sons, 1966. Reprinted by Robert E. Krieger Publishing Co., Inc., Huntington, New York, 1974.

Gregory, D. S. Compressed speech--the state of the art. IEEE Transactions on Engineering, Writing and Speech, 1969, EWS-12, 12-17.

Griffin, M. J. Vibration, pilot vision and helicopter accidents. University of Southampton, Institute of Sound and Vibration Research, Human Factors Research Unit, Operational Acoustics and Audiology Group, Contract Report No. CR 73/9, June 1973.

Grignetti, M. C. & Miller, D. Modifying computer response characteristics to influence command choice. Proceedings of IEE Conference on Man-Computer Interaction, London, UK, September 1970, 201-205.

Grignetti, M. C., Miller, D., Nickerson, R. S., & Pew, R. W. Information processing models and computer aids for human performance. Final Technical Report--Section 2, Task 2: Human-Computer Interaction Models. BBN Report No. 2352, March 1972.

Grindley, G. C. & Townsend, V. Voluntary attention in peripheral vision and its effects on acuity and differential thresholds. Quarterly Journal of Experimental Psychology, 1968, 20, 11-19.

Grossberg, M., Wiesen, R. A., & Yntema, D. B. An experiment on problem solving with delayed computer responses. IEEE Transactions on Systems, Man, and Cybernetics, March 1976, 219-222.

Haber, R. N. How we remember what we see. Scientific American, 1970, 222, 104-112.

Haber, R. N. & Erdelyi, M. H. Emergence and recovery of initially unavailable perceptual material. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 618-628.

Haith, M. M. Developmental changes in visual information processing and short-term memory. Human Development, 1971, 14, 249-261.

Harwood, K. A. Listenability and rate of presentation. Speech Monographs, 1955, 22, 57-59.

Hebb, D. O. Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), Brain mechanisms and learning. London and New York: Oxford University Press, 1961, 37-46.

Hebb, D. O. Organization of behavior. New York: John Wiley & Sons, 1949.

Hebb, D. O. Drives and the conceptual nervous system. Psychological Review, 1955, 62, 243-253.

Hebb, D. O. Concerning imagery. Psychological Review, 1968, 75, 466-477.

Hedberg, B. On man-computer interaction in organizational decision-making: A behavioural approach. Gothenburg University, 1970.

Hirsh, I. J. Influence of interaural phase on interaural summation and inhibition. Journal of the Acoustical Society of America, 1948, 20, 536.

Hockey, G. R. J. The effect of loud noise on attentional selectivity. Quarterly Journal of Experimental Psychology, 1970, 22, 28-36.

Homa, D., Haver, B., & Schwartz, T. Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. Memory & Cognition, 1976, 4, 176-185.

Howell, W. C., Johnston, W. A., & Goldstein, I. L., Complex monitoring and its relation to the classical problem of vigilance. Organizational Behavior and Human Performance, 1966, 1, 129-150.

Hoyt, M. F. & Centers, R. Temporal situs of the effects of anticipated publicity upon commitment and resistance to counter-communication. Journal of Personality and Social Psychology, 1972, 22, 1-7.

Hudgins, C. V. & Stetson, R. H. Relative speed of articulatory movements. Archives Netherlands Phonetic Experimental, 1937, 13, 85-94.

Ikeda, M. & Takeuchi, T. Influences of foveal load on the functional visual field. Perception and Psychophysics, 1975, 18, 255-260.

Ingersoll, G. M. & DiVesta, F. J. Effects of modality preferences on performance on a bisensory missing-units task. Journal of Experimental Psychology, 1972, 93, 386-391.

Ingling, N. W. Categorization: A mechanism for rapid information processing. Journal of Experimental Psychology, 1972, 94, 239-244.

Ivergard, T. B. K. Man-Computer interaction in public systems. Paper presented at the Advanced Study Institute on Man-Computer Interaction, Mati, Greece, September 1976.

Jacobson, L. B., Hyatt, K. H., & Sandler, H. Effects of simulated weightlessness on responses of untrained men to +Gz acceleration. Journal of Applied Physiology, 1974, 36(6), 745-752.

Jerison, H. J. Vigilance: Biology, psychology, theory, and practice. Presented at NATO Conference on Vigilance II: Relationships among Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy, August 3-6, 1976. To be published in a volume with approximately that title, R. R. Mackie (Ed.), New York: Plenum (in press).

Jerison, H. J. & Pickett, R. M. Vigilance: A review and re-evaluation. Human Factors, 1963, 5(3), 211-238.

Jerison, H. J., Pickett, R. M., & Stenson, H. H. The elicited observing rate and decision processes in vigilance. Human Factors, 1965, 7, 107-128.

Johnson, D. M. A systematic introduction to the psychology of thinking. New York: Harper & Row, 1972.

Johnson, D. M. & Jennings, J. W. Serial analysis of three problem-solving processes. Journal of Psychology, 1963, 56, 43-52.

Johnson, N. F. Chunking and organization in the process of recall. In G. H. Bower (Ed.), The psychology of learning and motivation. Vol. 4. New York: Academic Press, 1970.

Johnston, W. A., Howell, W. C., & Williges, R. C. The components of complex monitoring. Organizational Behavior and Human Performance, 1969, 4, 112-124.

Just, M. A. & Carpenter, P. A. Eye fixations and cognitive processes. Cognitive Psychology, 1976, 8, 441-480.

Kahneman, D. Attention and effort. Englewood Cliffs, N.J.: Prentice Hall, 1973.

Kahneman, D. & Lass, N. Eye position in tasks of association and memory. Unpublished manuscript, Hebrew University, Jerusalem, 1971. Cited in Kahneman, 1973.

Kahneman, D. & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3, 430-454.

Kahneman, D. & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.

Kalsbeek, J. W. H. & Sykes, R. N. Objective measurement of mental load. Acta Psychologica, 1967, 27, 253-261.

Kantowitz, B. H. & Knight, J. L., Jr. On experimenter-limited processes. Psychological Review, 1976, 83, 502-507.

Kelley, H. H. & Thibaut, J. W. Group problem-solving. In G. Lindzey & Aronson (Eds.), Handbook of social psychology. Cambridge: Addison Wesley, 1968, 1-104.

Kellicutt, M. H., Parks, T. E., Kroll, N. E., & Salzberg, P. M. Visual memory as indicated by latency of recognition for normal and reversed letters. Journal of Experimental Psychology, 1973, 97, 387-390.

Kelly, M. J. Studies in interactive communication: Limited vocabulary natural language dialogue. John Hopkins University, Department of Psychology, Technical Report No. 3, December 1975.

Kennedy, R. S. A bibliography of the role of the vestibular apparatus under water and pressure: Content-oriented and annotated. Naval Medical Research Institute Report No. 1, M4306.03.5000BAK9, August 1972.

Kennedy, R. S. Personal communication. ARPA Conference on Stress Effects in C3 Systems, Los Angeles, August 1976.

Kennedy, R. S. The relationship between vigilance and eye movements induced by vestibular stimulation. Presented at NATO Conference on Vigilance II: Relationships among Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy, August 3-6, 1976. To be published in a volume with approximately that title, R. R. Mackie (Ed.), New York: Plenum (in press).

Kennedy, T. C. S. The design of interactive procedures for man-machine communications. International Journal Man-Machine Studies, 1974, 6, 309-334.

Kibler, A. W. The relevance of vigilance research to aerospace monitoring tasks. Human Factors, 1965, 93-99.

Kinchla, R. A. An attention operating characteristic in vision. Technical Report 29, Department of Psychology, McMaster University, Hamilton, Ontario, 1969.

Kintsch, W. Learning, memory, and conceptual processes. New York: John Wiley & Sons, 1970a.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), Models of human memory. New York: Academic Press, 1970b.

Kirsner, K. & Craik, F. I. Naming and decision processes in short-term recognition memory. Journal of Experimental Psychology, 1971, 88, 149-157.

Kleitman, N. Sleep and wakefulness. Chicago: University of Chicago Press, 1963. Second Edition. Also Chapter 1 in D. Foulke (Ed.), The psychology of sleep. New York: Charles Scribner's Sons, 1966.

Klump, R. G. & Webster, J. C. Intelligibility of time-compressed speech. Journal of the Acoustical Society of America, 1961, 31, 265-267.

Kogan, N. & Carlson, J. Group risk-taking under competitive and non-competitive conditions in adults and children. Journal of Educational Psychology, 1969, 60, 158-167.

Kogan, N. & Wallach, M. A. The risky shift phenomenon in small decision making groups: A test of the information exchange hypothesis. Journal of Experimental Social Psychology, 1967, 3, 75-85.

Kriloff, H. Z. Human factor considerations for interactive display systems. Position Paper ACM/SIGGRAPH Workshop, on User-Oriented Design of Interactive Graphics Systems, October 14, 1976.

Kroger, M. G. & Christie, L. S. Command and control in limited war: Execution and monitoring. North American Aviation Inc. and System Development Corp., 1963.

Kroll, N. E., Parkinson, S. R., & Parks, T. E. Sensory and active storage of compound visual and auditory stimuli. Journal of Experimental Psychology, 1972, 95, 32-38.

Kroll, N. E. A., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. Short-term memory while shadowing: Recall of visually and of aurally presented letters. Journal of Experimental Psychology, 1970, 85, 220-224.

Laming, D. Mathematical psychology. London: Academic Press, 1973.

Lamm, H. Will an observer advise higher risk taking after hearing a discussion of the decision problem? Journal of Personality and Social Psychology, 1967, 6, 467-471.

Lamm, H. & Kogan, N. Risk-taking in the context of intergroup negotiation. Journal of Experimental Social Psychology, 1970, 6, 351-363.

Lamm, H., Trommsdorff, G., & Kogan, N. Pessimism-optimism and risk-taking in individual and group contexts. Journal of Personality & Social Psychology, 1970, 15, 306-374.

Landauer, T. Rate of implicit speech. Perceptual & Motor Skills, 1962, 15, 646.

Langford, R. P. The effect of compressed speech on listening comprehension. Dissertation Abstracts, 1968, 29, 411A.

Laughery, K. R. & Harris, G. J. Visual and auditory intrusion errors in short-term memory. Journal of Experimental Psychology, 1970, 83, 101-106.

Laughery, K. R. & Pinkus, A. L. Effects of acoustic similarity, presentation rate, and presentation mode. Psychonomic Science, 1966, 6, 285-286.

Ledgard, H. F. Ten mini-languages: A study of topical issues in programming language. Computing Surveys, 1971, 3, 115-146.

Levine, J. M., Romashko, T., & Fleishman, E. A. Evaluation of an abilities classification system for integrating and generalizing human performance research findings: An application to vigilance tasks. Journal of Applied Psychology, 1973, 58, 149-157.

Levison, W. H., Elkind, J. I., & Ward, J. L. Studies of multivariable manual control systems; A model for task interference. NASA CR-1746, May, 1971.

Libby, W.L., Lacey, B. C., & Lacey, J. I. Pupillary and cardiac activity during visual attention. Psychophysiology, 1973, 10, 270-294.

Licklider, J. C. R. Influence of interaural phase relations upon the mashing of speech by white noise. Journal of the Acoustical Society of America, 1948, 20, 150.

Licklider, J. C. R. Man-computer communication. In C. A. Cuadra (Ed.), Annual review of information science and technology, 3. Chicago, Illinois: Encyclopedia Britannica, Inc., 1968.

Liebowitz, H. W. & Appelle, S. The effect of a central task on luminance thresholds for peripherally presented stimuli. Human Factors, 1969, 11, 387-392.

Lindner, W. A. Recognition performance as a function of detection criterion in a simultaneous detection-recognition task. Journal of the Acoustical Society of America, 1968, 44, 204-211.

Litecky, C. R. & Davis, G. B. A study of errors, error-proneness, and error diagnosis in COBOL. CACM, 1976, 19(1).

Lockhart, J. M. Extreme body cooling and psychomotor performance. Ergonomics, 1968, 11(3), 249-260.

Loftus, E. F. Leading questions and the eyewitness report. Cognitive Psychology, 1975, 7, 560-572.

Loftus, E. F. & Palmer, J. C. Reconstruction of automobile destruction: An example of the interaction between language and memory. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 585-589.

Loftus, G. R. Eye fixations and recognition memory for pictures. Cognitive Psychology, 1972, 3, 525-551.

Loftus, G. R. A framework for a theory of picture recognition. In R. A. Monty and J. W. Senders (Eds.), Eye movements and psychological processes. Hillsdale, N.J.: Lawrence Erlbaum, 1976.

Loftus, G. R. & Bell, S. M. Two types of information in picture memory. Journal of Experimental Psychology: Human Learning and Memory, 1975, 104, 103-113.

Lorens, S. A., Jr. & Darrow, D. W. Eye movements, EEG, GSR and EKG during mental multiplications. EEG and Clinical Neurophysiology, 1962, 14, 739-746.

Luchins, A. S. Mechanization in problem solving. Psychological Monographs, 1942, 54(6).

Luria, A. R. The mind of a mnemonist. New York: Basic Books, 1968.

Mackworth, J. F. The relation between the visual image and post-perceptual immediate memory. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 75-85.

Mackworth, J. F. Vigilance and attention: A signal detection approach. Middlesex, England and Baltimore, Maryland: Penguin, 1970.

Mackworth, N. H. The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology, 1948, 1, 6-21.

Mackworth, N. H. & Bruner, J. S. How adults and children search and recognize pictures. Human Development, 1970, 13, 149-177.

Mackworth, N. H. & Morandi, A. J. The gaze selects informative details within pictures. Perception and Psychophysics, 1967, 2, 547-552.

Maclay, H. & Osgood, C. E. Hesitation phenomena in spontaneous English speech. Word. 1959, 15, 19.

Maclay, H. & Osgood, C. E. Hesitation phenomena in spontaneous English speech. In L. A. Jakobovits & M. S. Miron (Eds.), Readings in the psychology of language. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1967, 305-324.

MacPherson, R. K. Thermal stress and thermal comfort. In A.T. Welford (Ed.), Man under stress. New York: John Wiley & Sons, 1974.

Magnuson, R. A. Computer assisted writing. Datamation, 1966, 12, 49-52, 57, 59.

Mandler, G. Organization and memory. In K. W. Spence and J. T. Spence (Eds.), Psychology of learning and motivation. Vol. 1. New York: Academic Press, 1967.

Marquis, D. G. Individual responsibility and group decisions involving risk. Industrial Management Review, 1962, 3, 8-23.

Martin, J. Design of man-computer dialogues. Englewood Cliffs, N.J.: Prentice Hall, 1973.

McCann, P. H. Variability of signal detection measures with noise type. Psychonomic Science, 1969, 15(6), 310-311.

McGrath, J. J. The effect of irrelevant environmental stimulation on vigilance performance. Technical Report 6, Human Factors Research, Inc., 1960.

McGrath, J. J., Harabedian, A., & Buckner, D. N. Review and critique of the literature on vigilance performance. Technical Report 206-1, December 1959, in Studies of Human Vigilance - an Omnibus of Technical Reports, Human Factors Research, Inc., January 1968.

Meadow, C. T. Man-machine communication. New York: John Wiley & Sons, 1970.

Merikle, P. M. Cognitive backward masking: A type of masking by nonoverlapping material. Paper presented at Eastern Psychological Association, Boston, 1974.

Mewhort, D. J. K. Scanning, chunking, and the familiarity effect in task recognition. Journal of Experimental Psychology, 1972, 93, 69-71.

Mewhort, D. J. K., Merikle, P. M., & Bryden, M. P. On the transfer from iconic to short-term memory. Journal of Experimental Psychology, 1969, 81, 89-94.

Michaelis, F. H. (Admiral, USN) The C3 generation. Signal, 1975, 28-31.

Miller, G. A. Intelligibility of speech: Effects of distortion. In C. E. Waring (Ed.), Combat instrumentation II. Washington, D.C.: Government Printing Office, 1946, 88-109.

Miller, G. A. The perception of speech. Harvard University Psycho-Acoustic Laboratory under Contracts AF 33-(038)-14343, AFCRC-TN-56-50, and N5ori-76 (Project NR142-201, Report PNR-180), 1950, 353-360.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, *63*, 81-97.

Miller, G. A. & Licklider, J. C. R. The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 1950, *22*, 167-173.

Miller, G. A. & Nicely, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, *27*, 338-352.

Miller, J. & Tanis, D. C. Recognition memory for common sounds. *Psychonomic Science*, 1971, *23*, 307-308.

Miller, L. Programming by non-programmers. IBM Research Report RC 4280, 1973.

Milner, B. & Teuber, H. L. Alteration of perception and memory in man: Reflections on methods. In L. Weiskrantz (Ed.), *Analysis of behavioral change*. New York: Harper & Row, 1968.

Milosevic, S. Effect of time and space uncertainty on a vigilance task. *Perception & Psychophysics*, 1974, *15*(2), 331-334.

Moray, N. *Attention*. New York: Academic Press, 1970.

Moray, N. Attention, control, and sampling behavior. Presented at NATO Conference on Monitoring Behavior and Supervisory Control, Berchtesgaden, (F.R.) Germany, March 8-12, 1976, Preprints pp. 160-182.

Morefield, M. A., Wiesen, R. A., Grossberg, M., & Yntema, D. B. Initial experiments on the effects of system delay on on-line problem-solving. Lincoln Lab., Massachusetts Institute of Technology, Lexington, Mass. Tech. Note ESD-TR-69-158, June 1969.

Morgan, C. T., Chapanis, A., Cook, J. S., III, & Lund, M. W. (Eds.), *Human engineering guide to equipment design*. New York: McGraw-Hill, 1963. (Sponsored by Joint Army-Navy-Air Force Steering Committee.)

Morris, W. (Ed.). *The American heritage dictionary of the English language*. American Heritage Publishing Co. & Houghton Mifflin Co., 1970.

Morton, A. S. The role of the commander in a command and control system. Mitre Corporation Technical Memorandum TM-3292, Mat 1962.

Morton, M. S. S. Computer-driven visual display devices - their impact on the management decision making process. Doctoral Dissertation, Harvard University, Graduate School of Business Administration , 1967.

Murdock, B. B. Recent developments in short-term memory. British Journal of Psychology, 1967, 58, 421-433.

Murdock, B. B. Decoding as a function of the number of bits per chunk. Journal of Experimental Psychology, 1968, 78, 1-7.

Murdock, B. B., Jr. Recent developments in short-term memory. British Journal of Psychology, 1967, 58, 421-433.

Murdock, B. B., Jr. & Walker, K. D. Modality effects in free recall. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 665-676.

Myers, D. G. & Lamm, H. The group polarization phenomenon. Psychology Bulletin, 1975.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Nelson, T. O., Metzler, J., & Reed, D. A. Role of details in the long-term recognition of pictures and verbal descriptions. Journal of Experimental Psychology, 1974, 102, 184-186.

Newell, A., Barnett, J., Forgie, J. W., Green, C., Klatt, D., Licklider, J. C. R., Munson, J., Reddy, D. R., & Wood, W. A. Speech understanding systems: Final report of a study group. Pittsburgh, PA: Carnegie Mellon University, 1971.

Newell, A. & Simon, H. A. Human problem solving. Englewood Cliffs, New Jersey: Prentice Hall, 1972.

Nickerson, R. S. Auditory codability and the short-term retention of visual information. Journal of Experimental Psychology, 1972, 95, 429-436.

Nickerson, R. S. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. Canadian Journal of Psychology, 1965, 19, 155-160.

Nickerson, R. S. A note on long-term recognition memory for pictorial material. Psychonomic Science, 1968, 11, 58.

Nickerson, R. S. Man-computer interaction: A challenge for human factors research, invited paper, prepared for IEEE-GMMS ERS International Symposium on Man-Machine Systems, Cambridge, England, August 1969, Ergonomics 1969, 12, 501-517. Also in IEEE Transactions on Man-Machine Systems. 1969, 10, 164-180.

Nickerson, R. S. Can characters be classified directly as digits vs. letters or must they be identified first? Memory & Cognition, 1973, 1, 477-484.

Nickerson, R. S. Effects of correlated and uncorrelated noise on visual pattern matching. In P. M. A. Rabbitt & S. Dornic (Eds.), Attention and performance V. New York and London: Academic Press, 1975.

Nickerson, R. S. Some characteristics of conversation. Proceedings of NATO Advanced Study Institute Conference on Man-Computer Interaction, Mati, Greece, 5-18 September 1976.

Nickerson, R. S. Short-term retention of visually presented stimuli: Some Evidence of visual encoding. Acta Psychologica, 1976, 40, 153-162.

Nickerson, R. S. & Feehrer, C. E. Decision making and training: A review of theoretical and empirical studies of decision making and their implications for the training of decision makers. NAVTRAEQUIPCEN 73-C-0128-1, 1975.

Nickerson, R. S. & Freeman, B. Discrimination of the order of the components of repeating tone sequences: Effects of frequency separation and extensive practice. Perception and Psychophysics, 1974, 16, 471-477.

Nickerson, R. S. & Pew, R. W. Oblique steps toward the human factors engineering of interactive systems. BBN Report NO. 2172, 1971.

Nickerson, R. S. & Pew, R. W. Visual pattern matching: An investigation of some effects of decision task, auditory codability and spatial correspondence. Journal of Experimental Psychology, 1973, 98(1), 36-43.

Nickerson, R. S. & Stevens, K. N. Teaching speech to the deaf: Can a computer help? IEEE Transactions on Audio & Electroacoustics, 1973, 21, 445-455.

Nolte, L. W. Theory of signal detectability: Adaptive optimum receiver design. Journal of the Acoustical Society of America, 1967, 42(4), 773-777.

Nolte, L. W. & Jaarsma, D. More on the detection of one of M orthogonal signals. Journal of the Acoustical Society of America, 1967, 41(2), 497-505.

Norman, D. A. Models of human memory. New York: Academic Press, 1970.

Norman, D. A. & Bobrow, D. G. On data-limited and resource-limited processes. Cognitive Psychology, 1975, 7, 44-64.

Norman, D. A. & Bobrow, D. G. On the analysis of performance operating characteristics. Psychological Review, 1976, 83(6), 508-510.

Norman, D. A. & Rumelhart, D. E. Memory and knowledge. In D. A. Norman and D. E. Rumelhart (Eds.), Explorations in cognition. San Francisco: W. H. Freeman, 1975.

Noton, D. A theory of visual pattern perception. IEEE Transactions on Systems Science and Cybernetics, 1970, SC-6, 349-357.

Noton, D. & Stark, L. Scanpaths in eye movements during pattern perception. Science, 1971, 171, 308-311.

Noton, D. & Stark, L. Eye movements and visual perception. Scientific American, 1973, 225, 35-43.

O'Hanlon, J., Jr., Schmidt, E. A., & Baker, C. H. A study to determine the effects of placebos upon performance in a vigilance task. Technical Report 750-3, January 1964, in Studies of Human Vigilance - an Omnibus of Technical Reports, Human Factors Research, Inc., January 1968.

Ochsman, R. B. & Chapanis, A. The effects of 10 communication modes on the behavior of teams during co-operative problem solving. International Journal of Man-Machine Studies, 1974, 6, 579-619.

Paivio, A. Imagery and verbal processes. New York: Holt, Rinehart, & Winston, 1971.

Palmer, S. E. The effects of contextual scenes on the identification of objects. Memory & Cognition, 1975, 3, 519-526.

Parasuraman, R. Vigilance: A topical bibliography 1968-1976, unpublished, 1976a.

Parasuraman, R. Task classification and decision processes in monitoring behavior. Unpublished doctoral dissertation, University of Aston in Birmingham, United Kingdom, 1976b.

Parasuraman, R. & Davies, D. R. A taxonomic analysis of vigilance performance. Presented at NATO Conference on Vigilance II: Relationships among Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy, August 3-6, 1976. To be published in a volume with approximately that title, R. R. Mackie (Ed.), New York: Plenum (in press).

Parkinson, S. R., Parks, T. E., & Kroll, N. E. Visual and auditory short-term memory: Effects of phonemically similar auditory shadow material during the retention interval. Journal of Experimental Psychology, 1971, 87, 274-280.

Parks, T. E., Kroll, N. E. A., Salzberg, P., & Parkinson, S. R. Persistence of visual memory as indicated by decision time in a matching task. Journal of Experimental Psychology, 1972, 93, 437-438.

Patterson, J. H. & Green, D. M. Discrimination of transient signals having identical energy spectra. Journal of the Acoustical Society of America, 1970, 48, 894-905.

Paul, D. D. & Sutton, S. Evoked potential correlates of response criterion in auditory signal detection. Science, 1972, 177, 362-364.

Pellegrino, J. W., Siegel, A. W., & Dhawan, M. Short-term retention of pictures and words: Evidence for dual coding systems. Journal of Experimental Psychology: Human Learning and Memory, 1975, 104, 95-102.

Peterson, C. R. & Beach, L. R. Man as an intuitive statistician. Psychology Bulletin, 1967, 68(1), 29-46.

Peterson, L. R. & Peterson, M. J. Short-term retention of individual verbal items. Journal of Experimental Psychology, 1959, 58, 193-198.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. The theory of signal detectability. Transactions IRE Professional Group on Information Theory, 1954, PGIT-4, 171-212.

Pew, R. W. Notes from a Workshop on Perception and Information Processing, 9 July 1976.

Pew, R. W. & Rollins, A. M. Dialog specification procedures. BBN Report No. 3092, U.S. Dept. of Agriculture, August 1975.

Phillips, L. D. & Edwards, W. Conservation in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-357.

Phillips, W. A. & Baddeley, A. D. Reaction-time and short-term visual memory. Psychonomic Science, 1971, 22, 73-74.

Pick, A. P. Improvement of visual and tactual form discrimination. Journal of Experimental Psychology, 1965, 69, 331-339.

Pitz, G. F., Reinhold, H., & Geller, E. S. Strategies of information seeking in deferred decision making. Organizational Behavior and Human Performance, 1969.

Plomp, R. The ear as a frequency analyzer. Journal of the Acoustical Society of America, 1964, 36, 1628.

Pollack, I. & Ficks, L. Information of elementary multidimensional auditory displays. Journal of the Acoustical Society of America, 1954, 26, 155-158.

Pollack, I. & Spence, D. Subjective pictorial information and visual search. Perception and Psychophysics, 1968, 3, 41-44.

Pomerantz, J. R. & Garner, W. R. Stimulus configuration in selective attention tasks. Perception & Psychophysics, 1973, 14, 565-569.

Posner, M. I. Components of skilled performance. Science, 1966, 152, 1712-1718.

Posner, M. I. Short-term memory systems in human information processing. Acta Psychologica, 1967, 27, 267-284.

Posner, M. I. Abstraction and the process of recognition. In G. H. Bower (Ed.), Advances in learning and motivation. New York: Academic Press, 1969.

Posner, M. I. On the relationship between letter names and superordinate categories. Quarterly Journal of Experimental Psychology, 1970, 22, 279-287.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. Journal of Experimental Psychology Monograph, 1969, 79(1, Pt. 2).

Posner, M. I. & Keele, S. W. Decay of visual information from a single letter. Science, 1967, 158, 137-139.

Posner, M. I. & Keele, S. W. On the genesis of abstract ideas. Journal of Experimental Psychology, 1968, 77, 353-363.

Posner, M. I. & Keele, S. W. Retention of abstract ideas. Unpublished manuscript, University of Oregon, 1969. (Cited in Posner, 1969).

Posner, M. I. & Mitchell, R. F. Chronometric analysis of classification. Psychological Review, 1967, 74, 392-409.

Potter, M. C. Meaning in visual search. Science, 1975, 187, 965-966.

Potter, M. C. Short-term conceptual memory for pictures. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 509-522.

Potter, M. C. & Levy, E. I. Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology, 1969, 81, 10-15.

Poulton, E. C. Effect of fatigue upon sonar detection. TRC Report No. T75-4042, September 1972a.

Poulton, E. C. Environment and human efficiency. Charles C. Thomas, 1972b.

Poulton, E. C., Hitchings, N. B., & Brooke, R. B. Effect of cold and rain upon the vigilance of lookouts. Ergonomics, 1965, 8, 163-168.

Pylyshyn, Z. W. What the mind's eye tells the mind's brain: A critique of mental imagery. Psychological Bulletin, 1973, 80, 1-24.

Rabow, J., Fowler, F. J., Bradford, D. C., Hofeller, M. A., & Shibuya, Y. The role of social norms and leadership in risk taking. *Sociometry*, 1966, *29*, 16-27.

Rahe, R. The pathway between subjects' recent life changes and their near-future illness reports: Representative results and methodological issues. In B. S. & B. P. Dohrenwend (Eds.), *Stressful life events: Their nature and effects*. New York: John Wiley & Sons, 1974.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, *81*, 274-280.

Reid, R. H. Grammatical complexity and comprehension of compressed speech. *The Journal of Communication*, 1968, *18*, 236-242.

Reitman, J. S. Mechanisms of forgetting in short-term memory. *Cognitive Psychology*, 1971, *2*, 131-157.

Rim, Y. Risk-taking and need for achievement. *Acta Psychologica*, 1963, *21*, 108-115.

Rim, Y. Leadership attitudes and decisions involving risk. *Personnel Psychology*, 1965, *18*, 423-430.

Ronken, D. A. Monaural detection of a phase difference between clicks. *Journal of the Acoustical Society of America*, 1970, *47*, 1091.

Rouse, W. B. Design of man-computer interfaces for on-line interactive systems. *IEEE Proceedings*, *63*(6), June 1975.

Rubin, A. D. & Risley, J. F. The PROPHET system--An experiment in providing a computer resource to scientists. BBN Paper 1976.

Rundus, [illegible] Output order and rehearsal of multi-trial free recall. *Journal of Verbal Learning & Verbal Behavior*, 1974, *13*, 656-663.

Russo, J. E. & Rosen, L. D. An eye fixation analysis of multialternative choice. *Memory and Cognition*, 1975, *3*, 267-276.

Sackman, H. Outlook for man-computer symbiosis: Towards a general theory of man-computer problem solving. Paper presented at the Advanced Study Institute on Man-Computer Interaction, Mati, Greece, September 1976.

Sackman, H. & Citrebaum, L. On-line planning towards creative problem solving. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Sage, G. H. & Bennett, B. The effects of induced arousal on learning and performance of a pursuit motor skill. Research Quarterly, 1973, 44(2), 140-149.

Salancik, J. R. & Kiesler, C. A. Behavioral commitment and retention of consistent and inconsistent attitude word-pairs. In C. A. Kiesler, The psychology of commitment: Experiments linking behavior to belief. New York: Academic Press, 1971.

Salthouse, T. A. Using selective interfaces to investigate spatial memory representations. Memory & Cognition, 1974, 2, 749-757.

Scarborough, D. L. Stimulus modality effects on forgetting in short-term memory. Journal of Experimental Psychology, 1972, 95, 285-289.

Schachter, S. Deviation, rejection and communication. Journal of Abnormal & Social Psychology, 1951, 46, 190-207.

Scherr, A. L. An analysis of time-shared computer systems. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Mass., MAC-TR-18, 1965.

Schrenk, L. P. Objective difficulty and input history in sequential decision making. Human Factors, 1964, 9, 49-55.

Schum, D. A. On the behavioral richness of cascaded inference models: Examples in jurisprudence. Rice University, Report No. 75-1, May 1975.

Schum, D. A. The weighing of testimony in judicial proceedings from sources having reduced credibility. Human Factors, 1975, 17(2), 172-182.

Schum, D. A. & DuCharme, W. M. Comments on the relationship between the impact and the reliability of evidence. Organizational Behavior and Human Performance, 1971, 6, 111-131.

Scott, R. J. & Gerber, S. E. Pitch-synchronous time-compression of speech. Proceedings of 1972 conference on Speech Communication and Processing. IEEE and AFCRL, Newton, Mass., April, 1972, 63-65.

Scott, R. J. Temporal effects in speech analysis and synthesis. Unpublished doctoral dissertation, University of Michigan, 1965.

Selye, H. *The stress of life*. New York: McGraw-Hill, 1956.

Senders, J. W. The human operator as a monitor and controller of multi-degree-of-freedom systems. *IEEE Human Factors in Electronics*, 1964, *HFE-7*, 103-106.

Senders, J. W. An investigation of the visual sampling behavior of human observers. BBN Report No. 1246, 1965.

Senders, J. W. On the distribution of attention in a dynamic environment. In A. F. Sanders (Ed.), *Attention and performance*. Amsterdam: North Holland, Publ. (*Acta Psychologica*, 1967, *27*.)

Senders, J. W., Elkind, J. E., Grignetti, M., & Smallwood, R. P. An investigation of the visual sampling behavior of human observers. NASA CR-434, 1965.

Senders, J. W. & Posner, M. J. M. A queueing model of monitoring and supervisory behavior. Presented at NATO Conference on Monitoring Behavior and Supervisory Control, Berchtesgaden, (F.R.) Germany, March 8-12, 1976, Preprints pp. 183-197.

Shackel, B. & Shipley, P. Man-computer interaction: A review of the ergonomics literature and related research. EMI Electronics Report DMP 3472, Feltham, Middlesex, 1970.

Shaffer, W. O. & Shiffrin, R. M. Rehearsal and storage of visual information. *Journal of Experimental Psychology*, 1972, *92*, 292-296.

Sheehan, J. J. & Drury, C. G. The analysis of industrial inspection. *Applied Ergonomics*, 1971, *2*, 74-78.

Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 156-163.

Sheridan, T. On how often the supervisor should sample. *IEEE Transactions on Systems, Science and Cybernetics*. 1970, *SCC-6*, 140-145.

Shiffrin, R. C. & Gardner, G. T. Visual processing capacity and attentional control. *Journal of Experimental Psychology*, 1972, *93*, 72-82.

Shiffrin, R. M. & Schneider, W. Towards a unitary model for selective attention, memory scanning, and visual search. Indiana Mathematical Psychology Program Report No. 75-10, 1975.

Shiller, P. H. & Stryker, M. Single-unit recording and stimulation in superior colliculus of the alert rhesus monkey. Journal of Neurophysiology, 1972, 35, 915-924.

Shneiderman, B., Mayer, R., McKay, D., & Heller, P. Experimental investigations of the utility of flowcharts in programming. Indiana University Computer Science Department, Technical Report No. 36, Bloomington, Indiana, 1975.

Shneiderman, B. & McKay, D. Experimental investigations of computer program debugging and modification. Proceedings of the Sixth Congress of the International Ergonomics Association, July 1976, 557-562.

Shulman, H. & Fisher, R. Expected value as a determinant of attention. Journal of Experimental Psychology, 1972, 93, 343-348.

Shvartz, E., Magazanik, A., & Glick, Z. Thermal responses during training in a temperate climate. Journal of Applied Physiology, 1974, 36(5), 572-576.

Sidorsky, R. C. Some determinants of computer-aided decision effectiveness. Proceedings of the 80th Annual Convention of American Psychological Association, 1972, 697-698.

Siegel, S. & Zajonc, R. B. Group risk-taking in professional decisions. Sociometry, 1967, 30, 339-350.

Sime, M. E. "So I said in the most natural way, if x=0 then begin...": The empirical study of computer language. MRC Social and Applied Psychology Unit Memo No. 132, 1976.

Simon, H. A. How big is a chunk? Paper presented at the meetings of EPA, Atlantic City, New Jersey, May 1970.

Simon, H. A. & Barenfield, M. Information-processing analysis of perceptual processes in problem solving. Psychological Review, 1969, 76, 473-483.

Smith, E. E. Effects of familiarity on stimulus recognition and categorization. Journal of Experimental Psychology, 1967, 74, 324-332.

Smith, J. J., et al. Cardiovascular response of young men to diverse stresses. Aerospace Medicine, 1974, 45(6), 583-590.

Smith, S. L. An on-line model of traffic control in a communication network. MITRE Corp. Technical Report MTR 2813, March 1974.

Snapper, K. J. & Fryback, D. G. Inference based on unreliable reports. Journal of Experimental Psychology, 1971, 87, 401-404.

Snapper, K. J. & Peterson, C. R. Information seeking and data diagnosticity. Journal of Experimental Psychology, 1971, 87, 429-433.

Soelberg, P. O. Unprogrammed decision making. Industrial Management Review, 1967, 8, 19-29.

Sperling, G. The information available in brief visual presentations. Psychological Monographs, 1960, 74(11, Whole No. 498).

Sperling, G. A model for visual memory tasks. Human Factors, 1963, 5, 19-31.

Sperling, G. Successive approximations to a model for short-term memory. Acta Psychologica, 1967, 27, 285-292.

Sperling, G. Attention operating characteristics for visual search. Presented at Ninth Annual Mathematical Psychology Meetings, New York University, September, 1976.

Sperling, G., Budiansky, J., Spivak, J. G., & Johnson, M. C. Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. Science, 1971, 174, 307-311.

Squires, K. C., Squires, N. K., & Hillyard, S. A. Vertex evoked potentials in a rating-scale detection task: Relation to signal probability. Behavioral Biology, 1975a, 13, 21-34.

Squires, K. C., Squires, N. K., & Hillyard, S. A. Decision related cortical evoked potentials during an auditory signal detection task with cued observation intervals. Journal of Experimental Psychology: Human Perception and Performance, 1975b, 1, 268-279.

Standing, L., Conezio, J., & Haber, R. N. Perception and memory for pictures: Single-trial learning of 2560 visual stimuli. Psychonomic Science, 1970, 19, 73-74.

Starr, S. J., Metz, C. E., Lusted, L. B., & Goodenough, D. J. Visual detection and localization of radiographic images. Radiology, 1975, 116, 533-538.

Steinberg, J. D. Acoustics, the sense of hearing. Electrical Engineers Handbook. New York: John Wiley & Sons, 1936, 9-32.

Sterling, T. D. Humanizing computerized information systems. Science, 1975, 190, 1168-1172.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 1969, 57, 421-457.

Stevens, A. L. The role of inference and internal structure in the representation of spatial information. Unpublished Ph.D. dissertation, University of California at San Diego, 1976.

Stevens, R. T. Testing the NORAD command and control system. IEEE Trnsactions on Systems Science and Cybernetics, 1968, SSC-4, 47-51.

Stewart, T. F. M. The software interface and displays. HUSAT Memo No. 95, Department of Human Sciences, Univ. of Technology, Loughborough, 1974.

Stewart, T. F. M. The specialist user. Proceedings of a NATO Advanced Study Institute, Mati, Greece, September 1976 (in preparation).

Stoner, J. A. F. A comparison of individual and group decisions involving risk. Unpublished Master's Thesis, Massachusetts Institute of Technology, School of Industrial Management, 1961.

Sweet, A. Temporal discrimination by the human eye. American Journal of Psychology, 1953, 66, 185-198.

Swets, J. A. Signal detection theory applied to vigilance. Presented at NATO Conference on Vigilance II: Relationships among Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy, August 3-6, 1976. To be published in a volume with approximately that title, R. R. Mackie (Ed.), New York: Plenum (in press).

Swets, J. A. & Birdsall, T. G. The human use of information, III. Decision-making in signal detection and recognition situations involving multiple alternatives. Transactions IRE Information Theory, 1956, IT-2,, 138-165.

Swets, J. A. & Birdsall, T. G. Deferred decision in human signal detection: A preliminary experiment. Perception & Psychophysics, 1967a, 2(1), 15-18.

Swets, J. A. & Birdsall, T. G. Research on signal detection by human observers. Quarterly Progress Report 10, Bolt Beranek and Newman Inc., August, 1967b. (Work sponsored under Contract No. NAS2-2676, with the National Aeronautics and Space Administration).

Swets, J. A. & Kristofferson, A. B. Attention. Annual Review of Psychology, 1970, 21, 339-366.

Swets, J. A. & Sewall, S. T. Invariance of signal detectability over stages of practice and levels of motivation. Journal of Experimental Psychology, 1963, 66, 120-126.

Swets, J. A., Shipley, E. F., McKey, M. J., & Green, D. M. Multiple observations of signals in noise. Journal of the Acoustical Society of America, 1959, 31, 514-521.

Tanner, W. P., Jr. Theory of recognition. Journal of Acoustical Society of America, 1956, 28, 882-888.

Taylor, M. M., Lindsay, P. H., & Forbes, S. M. Quantification of shared capacity processing in auditory and visual discrimination. In A. F. Sanders (Ed.), Attention and performance. Amsterdam: North Holland Publishing Co., 1967, 223-229.

Taylor, R. W. Man-computer input-output techniques. IEEE Transactions on Human Factors in Electronics, 1967, 8, 33-48.

Teger, A. I. & Pruitt, D. G. Components of group risk-taking. Journal of Experimental Social Psychology, 1967, 3, 189-205.

Ternes, W. & Yuille, J. C. Words and pictures in an STM task. Journal of Experimental Psychology, 1972, 96, 78-86.

Thomas, I. B., Hill, P. B., Carroll, F. S., & Garcia, B. Temporal orders in the perception of vowels. Journal of the Acoustical Society of America, 1970, 48, 1010-1013.

Treisman, A. M. Strategies and models of selective attention. Psychological Review, 1969, 76, 282-299.

Turvey, M. T. The effects of rehearsing analyzed information upon the retrieval of unanalyzed information. Psychonomic Science, 1966, 6, 365-366.

Turvey, M. T. Repetition and the preperceptual information store. Journal of Experimental Psychology, 1967, 74, 289-293.

Turvey, M. T. On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. Psychological Review, 1973, 80, 1-52.

Turvey, M. T. Constructive theory, perceptual systems, and tacit knowledge. In W. B. Weimer & D. S. Palermo (Eds.), Cognition and the symbolic processes. Hillsdale, N.J.: Erlbaum Assoc., 1974.

Tversky, A. & Kahneman, D. Belief in the law of small numbers. Psychology Bulletin, 1971, 76, 105-110.

Tversky, A. & Kahneman, J. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Tversky, B. Pictorial and verbal encoding in a short-term memory task. Perception and Psychophysics, 1969, 6, 225-233.

Tversky, B. Eye fixations in prediction of recognition and recall. Memory & Cognition, 1974, 2, 275-278.

Tversky, B. & Sherman, T. Picture memory improves with longer on time and off time. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 114-118.

Tyler, D. M., Waag, W. L., & Halcomb, C. G. Monitoring performance across sense modes: An individual differences approach. Human Factors, 1972, 14(6), 539-547.

Underwood, B. J. & Schulz, R. W. Meaningfulness and verbal learning. New York: Lippincott, 1960.

Van Cott, H. P. & Kinkade, R. G. (Eds.). Human engineering guide to equipment design (revised edition). Washington, D.C.: U.S. Government Printing Office, 1972. Library of Congress Catalog Card Number 72-600054.

Vickers, D., Leary, J., & Barnes, P. Adaptation to decreasing signal probability. Presented at NATO Conference on Vigilance II: Relationships among Theory, Physiological Correlates and Operational Performance, St. Vincent, Italy, August 3-6, 1976. To be published in a volume with approximately that title, R. R. Mackie (Ed.), New York: Plenum (in press).

Voekler, C. H. An experimental study of the comparative rate of utterance of deaf and normal hearing speakers. American Annals of the Deaf, 1938, 83, 274-284.

Volkman, F. C., Schick, A. M. L., & Riggs, L. A. Time course of visual inhibition during voluntary saccades. Journal of the Optical Society of America, 1968, 58, 562-569.

Von Wright, J. M. Selection in visual immediate memory. Quarterly Journal of Experimental Psychology, 1968, 20, 62-68.

Von Wright, J. M. On selection in visual immediate memory. Acta Psychologica, 1970, 33, 280-292.

Wallach, H. & Bacon, J. The constancy of the orientation of the visual field. Perception and Psychophysics, 1976, 19, 492-498.

Wallach, M. A. & Kogan, N. Sex differences and judgment processes. Journal Personality, 1959, 27, 555-564.

Wallach, M. A. & Kogan, N. The roles of information, discussion and consensus in group risk taking. Journal of Experimental Social Psychology, 1965, 1, 1-19.

Wallach, M. A., Kogan, N., & Bem, D. J. Group influence on individual risk taking. Journal of Abnormal and Social Psychology, 1962, 65, 75-86.

Wallach, M. A., Kogan, N., & Burt, R. B. Group risk taking and field dependence-independence of group member. Sociometry, 1967, 30, 323-338.

Wallach, P. M. & Adams, S. K. The utility of signal-detection theory in the analysis of industrial inspector accuracy. AIIE Transactions, 1969, 1, 33-44.

Wallas, G. The art of thought. New York: Harcourt Brace, 1926.

Warm, J. S., Epps, B. D., & Ferguson, R. P. Effects of knowledge of results and signal regularity on vigilance performance.

Bulletin of the Psychonomic Society, October 1974, 4(4A), 272-274.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. Auditory sequence: Confusion of patterns other than speech or music. Science, 1969, 164, 586-587.

Watson, C. S. & Clopton, B. M. Motivated changes of auditory sensitivity in a simple detection task. Perception & Psychophysics, 1969, 5, 281-187.

Watson, C. S. & Nichols, T. L. Detectability of auditory signals presented without defined observation intervals. Journal of the Acoustical Society of America, 1976, 59(3), 655-667.

Weinberg, G. M. The psychology of computer programming. New York: Van Nostrand Reinhold, 1971.

Weisstein, N. & Harris, C. S. Visual detection of line segments: An object-superiority effect. Science, 1974, 186, 752-755.

Welford, A. T. (Ed.). Man under stress. New York: John Wiley & Sons, 1974.

Welford, A. T. Arousal, channel-capacity and decision. Nature, 1962, 194, 365-366.

Wheeler, D. D. Processes in word recognition. Cognitive Psychology, 1970, 1, 59-85.

Whitfield, D. Human factors aspects of interactive conflict resolution in air traffic control. Paper presented at Sixth Congress of the International Ergonomics Association, University of Maryland, 11-16, July 1976. Symposium in Computer Aiding in Cognitive Tasks.

Whitfield, D. & Stammers, R. Human Factors aspects of computer aiding for air traffic controllers. Paper presented at NATO Training Course on Human Engineering, Utrecht, May 1976.

Wickelgren, W. A. Short-term memory for repeated and non-repeated items. Quarterly Journal of Experimental Psychology, 1965a, 17, 14-25.

Wickelgren, W. A. Distinctive features and errors in short-term memory for English vowels. Journal of the Acoustical Society of America, 1965b, 38, 583-588.

Wickelgren, W. A. Distinctive features in short-term memory for English consonants. *Journal of the Acoustical Society of America*, 1966, *39*, 388-398.

Wickens, C. D. The effects of divided attention on information processing in manual tracking. *Journal of Experimental Psychology*, 1976, *2*(1), 1-13.

Wickens, C. D. & Gopher, D. Strategies and attention allocation in dual task performance. Presented at Psychonomics Society Meeting, Denver, Colorado, 1975.

Williams, L. G. The effect of target specification on objects fixated during visual search. *Perception and Psychophysics*, 1966, *1*, 315-318.

Williges, R. C. Manipulating the response criterion in visual monitoring. *Human Factors*, 1973, *15*(2), 179-185.

Wingfield, A. Acoustic redundancy and the perception of time-compressed speech. *Journal of Speech and Hearing Research*, 1975, *18*, 96-104.

Wohlstetter, R. *Pearl Harbor: Warning & decision*. Stanford: Stanford University Press, 1962.

Wolford, G. Perturbation model for letter identification. *Psychological Review*, 1975, *82*, 184-199.

Wolford, G. & Hollingsworth, S. Evidence that short-term memory is not the limiting factor in the tachistoscopic full-report procedure. *Memory & Cognition*, 1974, *2*, 796-800.

Wollen, K. A., Weber, A., & Lowry, D. H. Bizarreness versus interaction of mental images as determinants of learning. *Cognitive Psychology*, 1972, *3*, 518-524.

Wood, L. E. Visual and auditory coding in a memory matching task. *Journal of Experimental Psychology*, 1974, *102*, 106-113.

Woodson, W. E. & Conover, D. W. *Human engineering guide for equipment designers* (2nd ed.). Berkeley, Calif.: University of California Press, 1964. (First edition 1954).

Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

Wright, P. & Kahneman, D. Evidence for alternative strategies of sentence retention. Quarterly Journal of Experimental Psychology, 1971, 23, 197-213.

Wyndham, C. H. Research in the human sciences in the gold mining industry. American Industrial Hygiene Association Journal, 1974, 35(3), 113-136.

Yarbus, A. L. Eye movements and vision. New York: Plenum Press, 1967.

Yasukawa, K. The effect of comments on program understandability and error correction. Unpublished paper.

Young, E. A. Human factors in programming. International Journal of Man-Machine Studies, 1974, 6(3).

Young, L. On adaptive manual control. IEEE Transactions Man-Machine Systems, 1969, MMS-10, 292-332.

Youseff, Z. I. & Peterson, C. R. Intuitive cascaded inference. Organizational Behavior and Human Performance, 1973, 10, 349-358.

Zinchenko, V. P. & Vergiles, N. Y. Formation of visual images: Studies of stabilized retinal images. Translated by Consultants Bureau, New York: Plenum Press, 1972 Cited in Gould, 1976.

Zubeck, J. P. Behavioral and physiological effects of prolonged sensory and perceptual deprivation--a review. In Man in isolation and confinement. Chicago: Aldine Publishing Co., 1973, 8-83.