INTERACTION OF HUMAN COGNITIVE MODELS AND COMPUTER-BASED MODELS IN SUPERVISORY CONTROL

Thomas B. Sheridan

March 1984

CONTRACT N00014-83-K-0193
WORK UNIT NR 196-179
ENGINEERING PSYCHOLOGY PROGRAMS
OFFICE OF NAVAL RESEARCH
ARLINGTON, VIRGINIA 22217

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A142547
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Interaction of Human Cognitive Models and Computer-based Models in Supervisory Control	5. TYPE OF REPORT & PERIOD COVERED Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas B. Sheridan	8. CONTRACT OR GRANT NUMBER(s) N00014-83-K-0193
9. PERFORMING ORGANIZATION NAME AND ADDRESS Massachusetts Institute of Technology Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-196-179
11. CONTROLLING OFFICE NAME AND ADDRESS Engineering Psychology Group Office of Naval Research, Code 442EP 800 N. Quincy St., Arlington, VA 22217	12. REPORT DATE March 1984
	13. NUMBER OF PAGES 50
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)

Approved for public release; distribution unlimited

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

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

supervisory control	mental models	cognitive science
expert systems	knowledge representation	human factors
decision aids	decision theory	human performance
man-machine systems	computer aids	human-computer interactions

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

This report summarizes the first year's effort of a three-year research project on how knowledge is represented in decision aids and control systems and how the operators of such systems apparantly represent and utilize such knowledge.

(continued on next page)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The first section of the report discusses the relationship of computer-based supervisory control to computer-based decision-aiding (expert systems) by identifying component variables and functions and building up block diagrams.

The second section deals quantitatively with internal models, knowledge, and calibration, both with respect to expectations of the existence of identifiable states of the world and with respect to the overlap of meanings of terms (mental or linguistic encodings, "fuzzy" variables).

The third section discusses mental models and their importance in three kinds of activities supervisors must do in complex systems: (1) discovering how things work; (2) determining what is wanted out of the set of alternatives states of the attributes; (3) encoding and manipulating "fuzzy" concepts; (4) combining evidence and confidence; (5) deciding what to do.

The fourth section of the report deals with the human use of computer-based models in automatic control and in decision-aiding. It reports on three sets of experiments underway or completed: (1) A "satisficing" technique is being developed in part as a paradigm for studying how people decide what they want and balance multiple objectives (in this case in controlling a vehicle). (2) Preliminary experiments are reported on how subjects observe failures in a "black box" system, infer rules and assign "fuzzy" meanings to the terms used in their rules. (It is shown that fuzzy rule generation is both natural for humans to do and effective at predicting system response). (3) Current research is also reported on computer graphic aids for mixed human-computer planning of dynamic trajectories.

1. INTRODUCTION

1.1 Purposes

The purposes of this paper are to:

(1) define and characterize the relations between supervisory control and computer decision aids (expert systems), particularly with respect to the human operator's mental model and computer-based models of the controlled process or environment which might be incorporated in the decision aid or in an automatic controller,

(2) pose some salient paradigms for analysis and experimentation in regard to the above system relations, and

(3) report on some preliminary experimental results.

1.2 Fundamental relations

The exposition to satisfy purpose (1) above may best begin by a sequence of diagrams (Figs. 1-6) and definitions of terms (see glossary) at end of report.

Figure 1 shows the time-honored man-machine system in its simplest form, consisting of: (1) a human operator; (2) a controled process which, together with given objectives (goals, instructions, utilities) define the task of the operator. From displays he receives sensory feedback and he takes motor actions through controls with his hands, feet or voice. The symbol MM is the operator's internalized mental model of the task which lies outside him.

In Figure 2 a key new element is added: (3), a decision aid which is a question-answerer, advice-giver and which helps the operator decide what to do. The decision aid can be computer-based, taking the form of an expert system, or it can be a second human operator (or team of same). The symbol CM is a computerized model of the controlled process internalized within the decision aid.

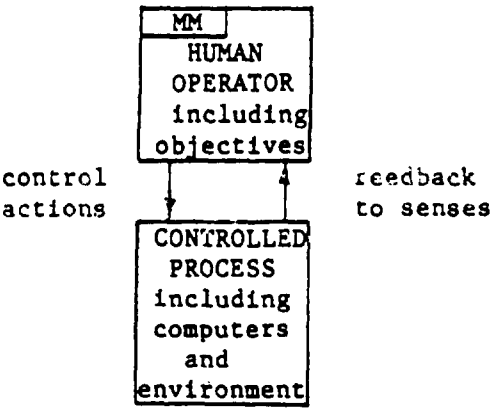


Figure 1. Simple man-machine system

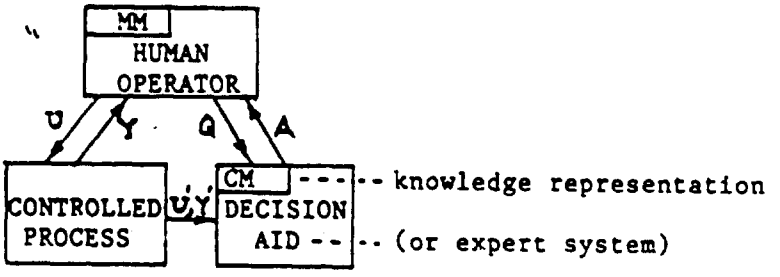


Figure 2. Decision-aided man-machine system

The controlled process may be simple manual skill (e.g., an olympic athletic event, where a human decision-aid takes the form of a coach standing nearby shouting advice or giving hand signals) or may itself embody a computer and/or automation. If the latter (e.g. an aircraft) it will likely involve a computerized internal model CM of the logic and/or differential equations (transfer function) of the controlled process and environmental forces. The latter is sometimes called a Kalman filter, estimator or observer and is an inherent part of a so-called optimal controller or modern control system.

A computer-based decision aid necessarily also embodies some form of internal model of the salient characteristics of the controlled process. Such systems have initially been implemented for important but non-time-stressed decision-making such as medical diagnosis/therapy, geological prospecting, or configuring of computer systems. These are not supervisory control systems as we will define them. Now there are plans for such decision aids to be implemented in supervisory control systems such as nuclear power plants and computer-aided gas well blowout systems, where time-stress, human safety and risk of capital are concurrent.

The artificial intelligence community often refers to a computer-based decision aid as an "expert system" and its internal model a "knowledge representation". The latter may encode knowledge about the world in various forms such as input-output equations or "if - then" relationships called "production rules" or just "productions". An even more general way of storing such knowledge is a "frame" wherein declarative facts, parameters, descriptors and semantic-networks and pointers to relevant information are stored in addition to strictly if-then or procedural representations.

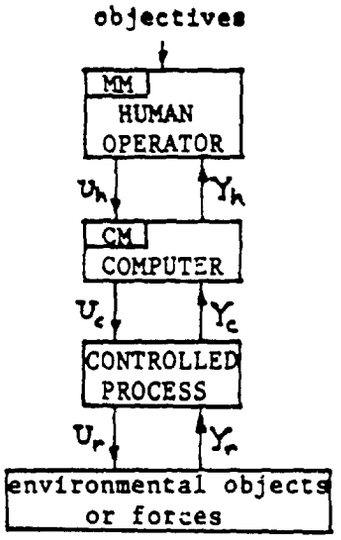
Using the usual symbolism of the control engineer we designate control, command or efferent variables as U, and measured state, feedback or afferent variables as Y. X are the true (but knowable only experimentally) state variables of the controlled process. U' is the subset of U available to the decision aid, while Y' is the set of state variable measurements available to the decision aid. (Both Y and Y' can have elements not common to the

other.) Q are questions asked by the human of the decision aid, and A are answers or advice given.

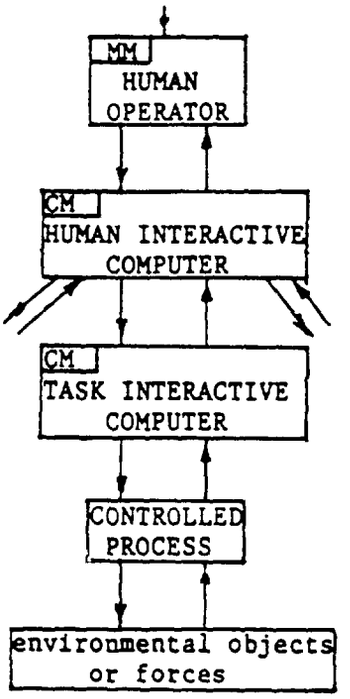
Note that the operator's displays and hand-foot-voice controls are no longer explicitly specified in the diagram. Both Q and U require some form of controls, and both A and Y some form of displays. Indeed the displays and controls for these two types of interaction between the human and the machine can be combined. The human engineering of the displays and controls per se is best decided in context and will not be dealt with in this paper.

Figure 3 illustrates what is meant by supervisory control, exclusive of the decision aid. In 3a a computer is imposed between human operator and controlled process, with a supervision of the high level control loop closed by U_h and Y_h through the human operator and the computer. A lower level control loop is closed by U_r and Y_r through the controlled process. Note that the latter is now defined to exclude the computer. Note also that exogeneous input variables called "objectives" (goals) and "environmental objects or forces" are added to complete the diagram.

Figure 3b shows the computer to be segmented into an executive or supervisory or human-interactive computer and one or many low-level or task-interactive computers. Often the latter is the case - such as where a single station coordinates many automatic subsystems. The former is multiplexed in time and space to the latter, as indicated schematically. Between the two types of computers is typically a spatial and temporal gap in the form of communications bandwidth constraints or time delay. This may be due to physical constraints on telecommunications or to multiplexing. Now we have a clean physical separation between the high-level human-interactive subsystem (HIS) and low-level task-interactive subsystem (TIS). The human-computer interaction within the HIS is in human-oriented language and is for the purpose of helping the human supervisor (1)plan, (2)teach (give commands) to the TIS, (3) monitor its automatic-execution of what is asked, (4)discover when it fails or completes its task, (5)intervene in emergencies or to assume manual control or reprogram the TIS, and (6)learn from accumulated experience. The one (or more likely many) TIS may be considered to be closely-watched and frequently reprogrammed robots. Examples of



(a) one control computer



(b) two control computers

Figure 3. Supervisory control

supervisory control are discussed more fully elsewhere (Sheridan,1982; Sheridan,1983).

Figure 4 shows that the functions of the high-level, supervisory, human-interactive computer can be divided into two parts, one to deal with command-giving by the human and "coordination of the troops" (the TIS), the second concerned with advice-giving and decision-aiding as was illustrated in Figure 2. Again note the presence of "internal models" in both parts, as discussed above.

By analogy to human organizations the decision-aid (expert system) functions as a policy or research staff, and the control-coordination functions as a production or command staff. The decision-aid also corresponds to those parts of the mammalian nervous system which support situation-assessing and cognitive contemplation, while the controller corresponds to those parts which support sensory-motor skill. It becomes clear that pure sensory function supports both of these, and indeed there are "chief-of-staff" or "executive coordination" functions of mediation of time and resources between them. The separation of physical elements between these functions is not clean in the animal or in the human organization or in the computer.

If the question-answer-advice-giver is another human we might just as well put him/her in the same position on the diagram as in the computer version, Figure 4. However if this second human is a full-fledged team member, also able to access the computer-based decision aid and also entitled to give commands, we have the much more complex situation diagrammed in Figure 5. This might be called a simplest model of the elements of a multi-person command and control system.

Figure 6 adds one new element - a human experimenter who observes and in some sense controls the whole (rest of the) system. In this case the experimenter is not both observing the state-variables of the controlled process and, through his internal model of the controlled process, estimating other state variables, as was the case with the internal model in the computer. In this case the experimenter-observer is a researcher/engineer

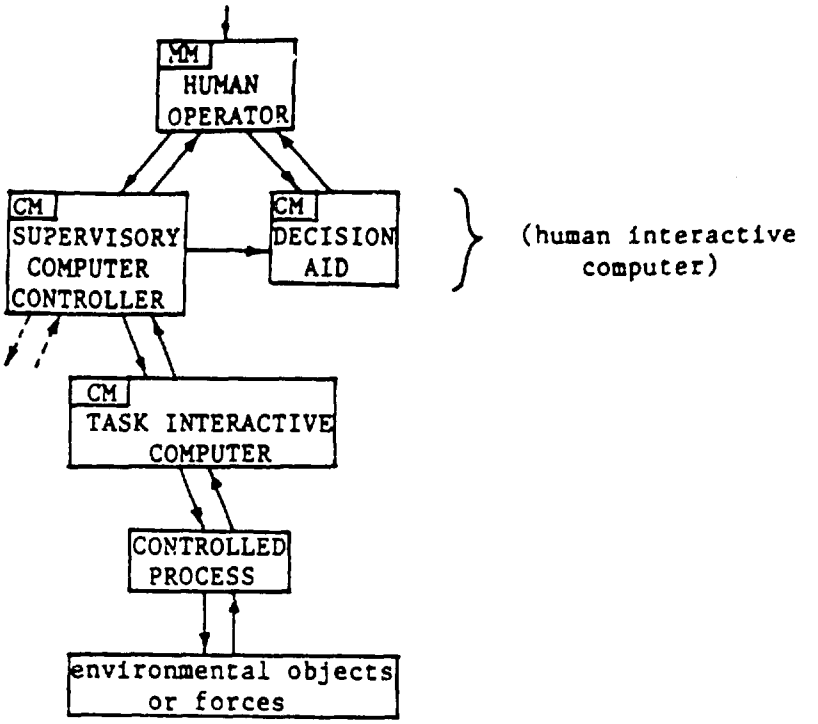


Figure 4. Supervisory remote control with two control computers and a decision aid

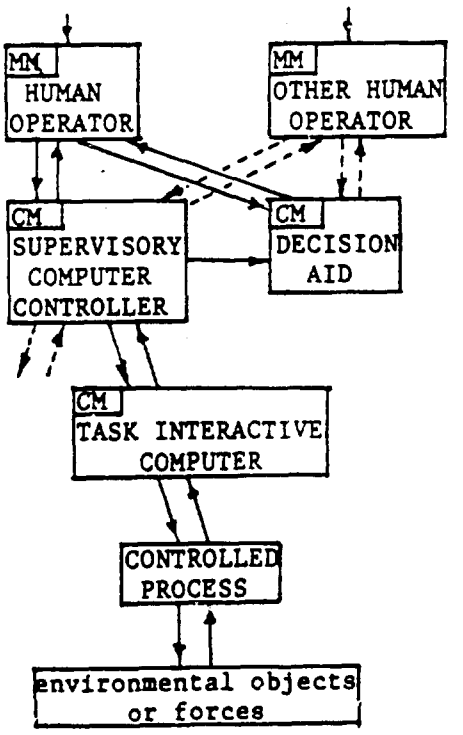


Figure 5. Multi-person supervisory control system

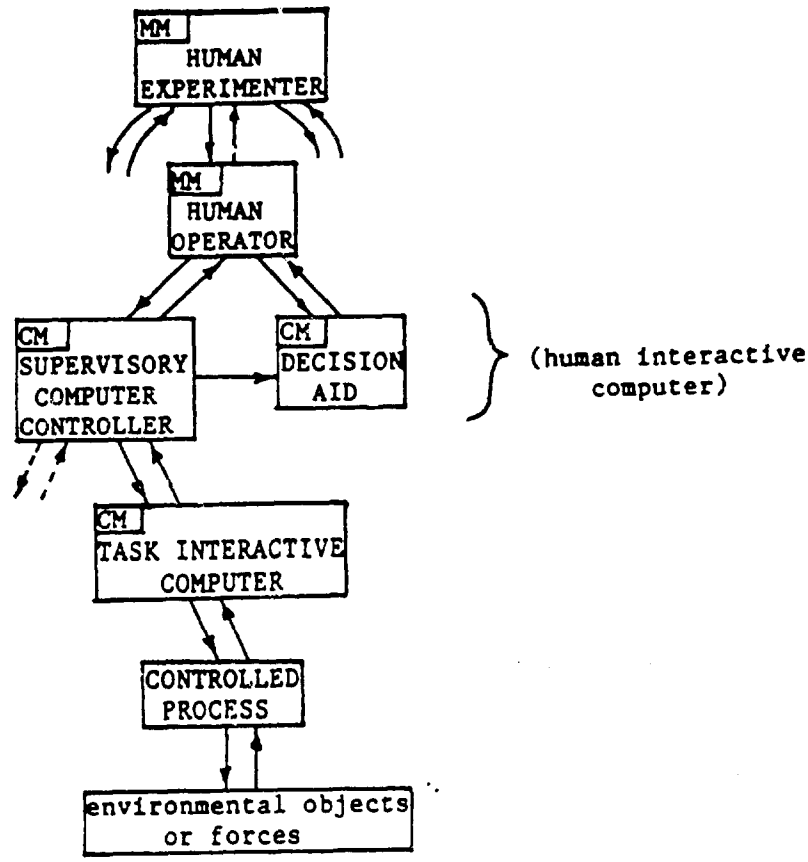


Figure 6. Experiment in supervisory control

who may first observe real supervisory control systems with built-in computerized decision aiding, who then sets up new experimental situations or simulations of known actual situations to do controlled experiments and build conceptual models of salient relationships. The mental model in the head of this observer is a model of the behavior of a system consisting of the human operator, the controlled process and the decision aid.

I and my colleagues in the present research are this observer. What we hypothesize and conclude and write in papers like this one emerge from that observer role and the mental models which develop in our heads. As indicated by (unlabeled) arrows, in such an experiment we can set the controlled task, giving it any degree of sophistication (including a computerized model (observer, estimator) which allows it to control automatically to a corresponding performance which we can determine). We can also set the decision aid, again including a computerized model (knowledge representation) of any quality or correspondence with reality. Finally we can set the human operator, by selection, training and instructions. Obviously there are many degrees of freedom for us to adjust - to determine a range of qualities of performance.

In the next section to follow we discuss internal models, what it means for an internal model to have knowledge, and how to characterize the correspondence of knowledge to truth or to the knowledge of another internal model, be it human or computer. In the third section we consider some of the most important research problems with regard to mental models of the human operator in relation to his supervisory control functions. In the fourth section we consider computer models and decision aiding (expert systems) in this supervisory control context. Here some relationships between mental and computer-based models are examined.

2. INTERNAL MODELS, KNOWLEDGE AND CALIBRATION

2.1 What is an internal model?

An internal model is a representation within a computer or (hypothetically) within a human operator of assertions about the world.

These assertions can be about tangible objects or about events, past, present or predicted future. They can be representations of systemic relations between elements coded as pictures or diagrams or tables, or values of isolated variables, or relationships between variables, such as "if - then" production rules (or simply "productions"), or can be input-output differential or difference equations. In the present context we assume the mental models are restricted to (1) the controlled process, (2) environmental objects or forces which must be rearranged or otherwise controlled (if not considered part of the controlled process) and (3) given objectives (goods and bads, benefits of various kinds of performance, costs or limits on uses of various resources).

2.2 Knowledge

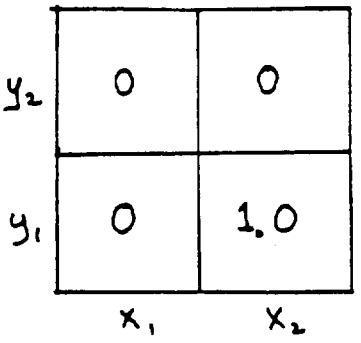
Knowledge is the representation of belief or expectation that certain states of the world are true and certain other states of the world are not true. More knowledge is a more refined or more specific representation of what is believed true and what is believed not true, i.e.,there is greater knowledge when there are more categories of what is believed not true.

In general any object or event or state of the world can be characterized in many dimensions of space,time, force, color or other attributes. For simplicity, assume there are only two dimensions x and y, each with only two levels. Then the state space has four categories, as shown in Figure 7a.

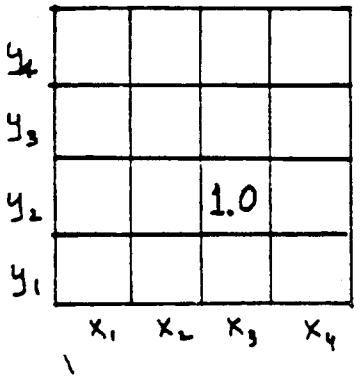
If an internal model represents x_2y_1 as the believed truth (and therefore the other three states as what is believed not the truth), then the amount of knowledge can be characterized by any function of the degree of specificity or concentration (Cooke, Mendel and Thijs, 1984). The most common measure is the reduction in entropy (confusion) in going from no expectation (all states are equally probable) to the given expectation. In this case where the given expectation has no uncertainty at all we have:

(entropy of no expectation) minus (entropy of given expectation) =

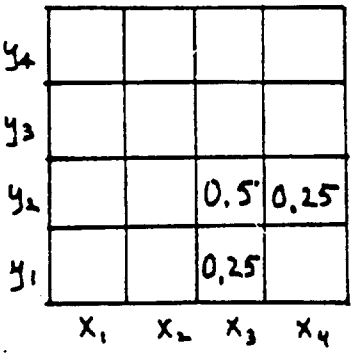
$$\log_2 4 - \log_2 1 = 2-0 = 2 \text{ bits}$$



(a) perfect knowledge with four states



(b) perfect knowledge with sixteen states



(c) imperfect knowledge with sixteen states

Figure 7. A way of representing knowledge

If each of x and y is specified to four levels, so that there are 16 possible states (Fig. 7b), representing the truth as x_3y_2 would constitute $\log_2 16 = 4$ bits of knowledge. If there is some expectation (probability) of several different states (as in Fig 7c) then, more generally, the knowledge K is

$$K = \sum_{ii} p_i p_j \log_2\left[\frac{1}{p_i p_j}\right] - \sum_{ij} p_i p_j \log_2\left[\frac{1}{p_i p_j}\right] =$$

(entropy of no expectation) minus (entropy of given expectation) =

$$\log_2 16 - \left(\frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4\right) = 2.5 \text{ bits}$$

Note that the information measure is only a function of expectation over different possible states, and is not related to the meaning of the alternative states (combinations of attributes). Below more will be said about meaning.

2.3 Correspondence of knowledge representations to each other: calibration

Note also that the above definition of knowledge says nothing about whether an internal model's representation of the believed truth is in fact true, or whether one internal model's representation corresponds to another such representation. To "know" in the sense we use it is "to act with certainty" about the world or to "have specific ideas". We all know people and machines that "act with certainty" or "have specific ideas" which are wrong, either because they don't agree with the world or because they don't agree with our own ideas!

Therefore, in addition to the notion of "knowledge" for a given internal model it is necessary to have the notion of "calibration", the correspondence

between the representation of knowledge in one model of the world internal to a person or computer and another representation of knowledge, (where the second is either the truth or it is the knowledge represented within a second person or computer).

Calibration of knowledge may be defined by the conventional chi-square statistic, where $p(x_i y_j) = p_{ij}$ is the probability density or expectation of the mental or computer model being calibrated and $q(x_i y_j) = q_{ij}$ is the reference expectation:

$$C = \sum_{ij} \sum [(p_{ij} - q_{ij})^2 / p_{ij}]$$

This is really a "miscalibration" measure.
It may also be defined by the relative entropy (Cooke, Mendel and Thijs, 1984):

$$C' = \sum_{ij} \sum (q_{ij} \log q_{ij}/p_{ij})$$

Note that in either case when $p_{ij} = 0$ for any q_{ij} finite the measure goes to infinity - which makes sense when it is realized that if a prior gives zero credibility to any hypothesis then no amount of evidence can force Bayesian updating to modify that zero.

Note also that in contrast to the above definition "calibration of knowledge" or "correspondence" could also mean accuracy (simple attribute distance between corresponding states), precision (standard deviation of the distribution of difference measures around their mean), Pearson-product moment correlation, and so on.

2.4 Correspondence of meanings within a single internal model: fuzzy sets

Knowledge was defined above in relation to a person's or computer's probability density of expectation over states of the world defined by one or more attributes (Figure 8a). In that case the meanings of those different states were assumed to be clearly distinguishable from one other. That is, there was no uncertainty in the mind of the human operator or the memory address of the computer other than which of them was believed to be true. The degree of certainty is in the strength of belief or expectation of existence where the truth itself is not observable. In fact only one state is or will ever be true, and that fact is clear to all concerned.

There is another type of uncertainty not in expectation of existence, but in meaning or identification. This is where information is coded or represented in the mind or computer memory in such a way that a term or symbol or concept can have shades or degrees of meaning among different states (objects or events within a set).

First let us consider unambiguous or distinct designation of sets. Suppose for a one-dimensional set X the term "small" is designated to represent the set including states x_1 and x_2, and exclude x_3, x_4 and x_5. In this case we can represent membership graphically by Figure 8b, where membership in the set "small" is 1 for x = 1, 2 and is 0 elsewhere. "Small" means both x=1 and x=2 for sure and not the others for sure. Then the set "large" might mean x_3, x_4, x_5 (and not x_1 and x_2).

Now let the membership function be continuous between 0 and 1, and let "small" clearly and strongly and appropriately mean x_1, less clearly and strongly and appropriately x_2 and so on until x_5, for which "small" is clearly inappropriate. This is called a "fuzzy membership function" and "small" is called a "fuzzy set". Expectation or confidence about the existence of any x is not considered here. Once the relative strengths of meaning are specified those meanings exist for certain, whatever the value of membership, over the whole fuzzy set.

It is clear that two or more fuzzy sets can be defined over any set of attribute states (Figure 8d).

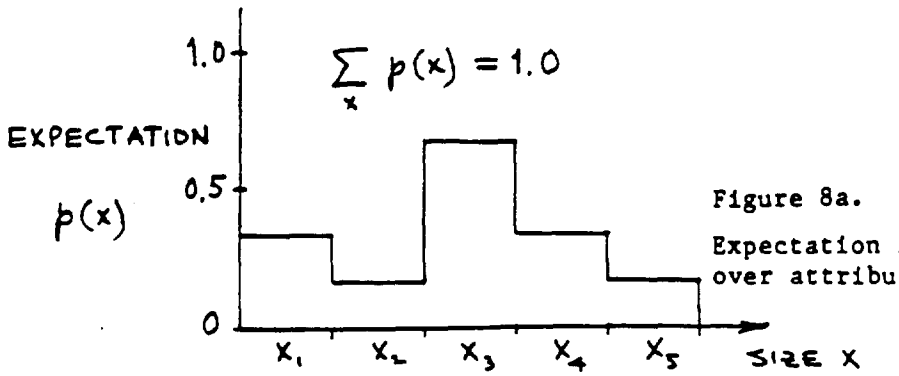


Figure 8a.

Expectation function
over attribute states

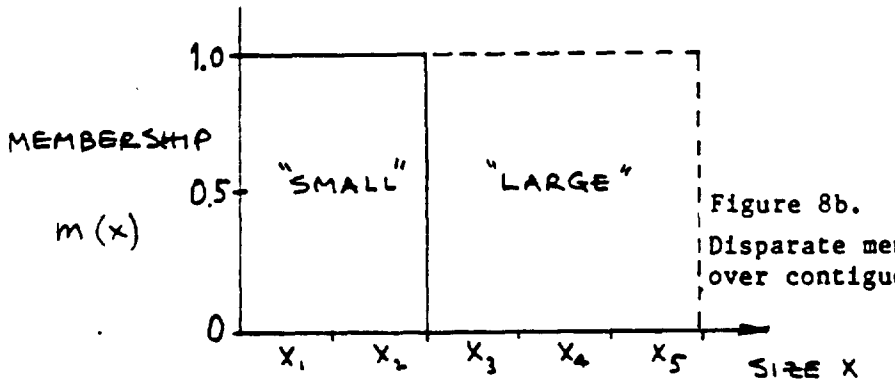


Figure 8b.

Disparate membership functions
over contiguous sets of attribute state

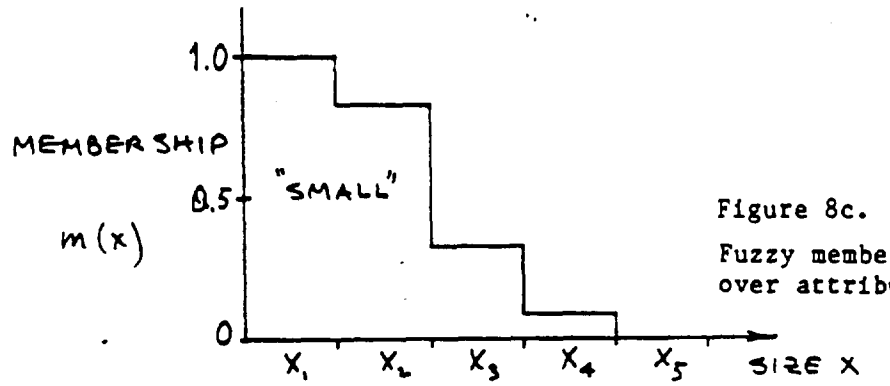


Figure 8c.

Fuzzy membership function
over attribute states

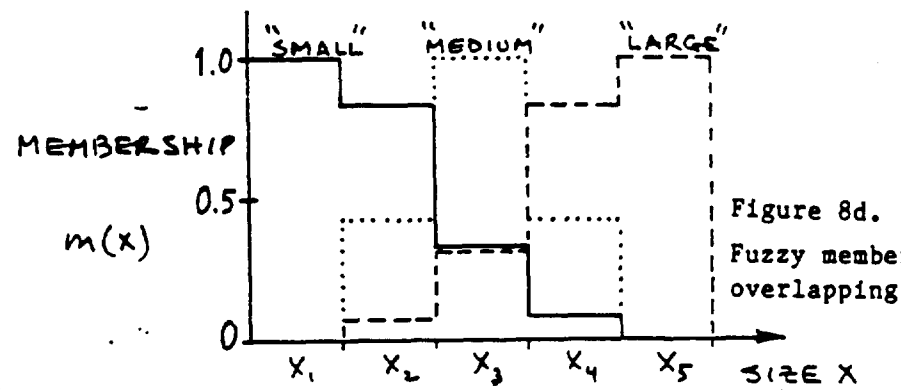


Figure 8d.

Fuzzy membership functions
overlapping attribute states

The theory of fuzzy sets has been developed by Zadeh (1965) and others because of the need to make these theoretical distinctions between expectation of existence and meaning and to be able to combine and manipulate assertions couched in fuzzy language, which is evidently what people do. There are various calculi (procedures for combining statements) which have been developed for this purpose. The most popular simply assumes that the membership m of any particular value or state of world attribute x in a representation combining two fuzzy sets inclusively, e.g. "small or medium", is the greater or stronger or most true of the two component memberships at that x:

$$m_{s1 \cup s2}(x) = \max [m_{s1}(x) \cdot m_{s2}(x)]$$

Similarly the membership in an exclusive combination (e.g. "small and medium") is the lesser or weaker or least true of the two component memberships at that x:

$$m_{s1 \cap s2}(x) = \min[m_{s1}(x), m_{s2}(x)]$$

From many such statements, presumably couched in fuzzy terms because that is the way people think about the world, e.g., "When size is large or medium, and speed is fast, do -", a "state action matrix" can be built up. This in effect specifies a precise response (or conclusion) u for every possible input (state), $x_i y_j$, as in Figure 9.

Note that the state action matrix is a summary of the implications of the set of fuzzy statements (based on the particular combinatorial calculus used). If a human operator communicates a number of fuzzy assertions to a computer and the computer derives the state action matrix the latter becomes the computer's internal model of knowledge as provided from that operator. And a display of the state action matrix to that human operator becomes feedback as to what the computer understood (can conclude) from what the human operator asserted. It also indicates the strength or relative truth corresponding to each u (in a sense the computer's degree of confidence that that u is appropriate for the given state). Seeing this feedback display the human operator can offer more assertions bearing on regions of the state

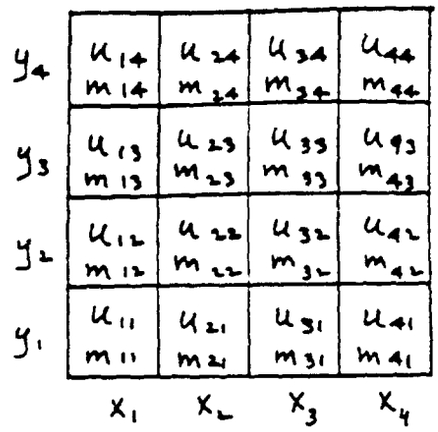


Figure 9. State action matrix

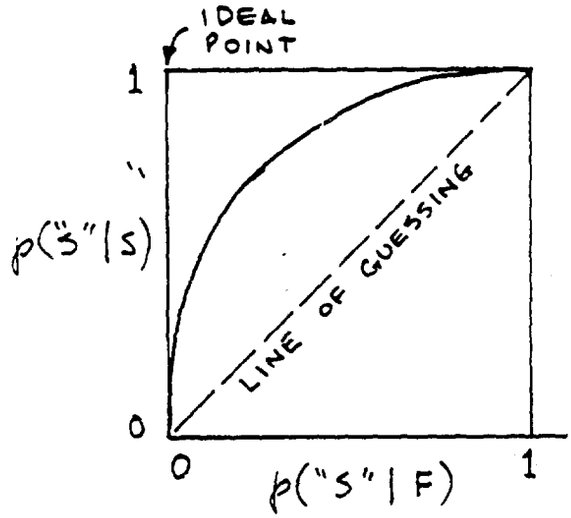


Figure 10. ROC curve for confidence (predicting one's own success)

space where m is weak, i.e., where the computer is not confident what a particular circumstance $x_i y_j$ implies about what to do u_{ij}.

3. MENTAL MODELS

3.1 Is "mental" admissible?

Behavioral scientists have always approached the subject of "how humans think" with trepidation. Early in this century, indeed, logical positivism and the methodology of "operationism" acclaimed by Bridgeman (1928) and other physicists so indimidated psychologists and reinforced the "behaviorists" that for a psychologist to purport to be studying "thinking" was to ask for ostracism from the scientific elite. Today computer technology has encouraged study of both artificial intelligence and cognitive science. It is now acceptable to speculate about thinking. However it is still as difficult as ever to establish generalizable models based on controlled experiments. There remain physiologists and "hard" behavioral scientists who eschew all that is "mental".

As an engineer-psychologist who respects this conservative, hard-line experimental operationism I nevertheless feel (after Descartes I suppose) that my consciousness is real and verbal report is legitimate evidence of same. So is choice behavior when the question is couched in semantically acceptable terms. Based on such behavioral responses a psychophysics of mental models seems tractable.

Assuming that "mental is admissible", five problems emerge for the operator which amount to five key mental activities of supervisory control: (1) discovering how things work, (2) determining what is wanted, (3) mental manipulation of fuzzy ideas or "chunks" (4) combining evidence with confidence, and (5) deciding what to do. These are discussed first as problems for the supervisory operator which in turn become problems of the cognitive researcher. In subsequent sections they are discussed as problems for the designer of computerized decision aids.

3.2 Discovering how things work

DeKleer and Seely-Brown (1982) in their studies of how humans understand simple electro-mechanical devices, distinguishes three kinds of mental activities: "device topology" (definition of elements, and structuring of connections between elements in terms of variables), "envisioning" (development of a qualitative simulation of what causes what and is constrained by what) and "running" (testing of the simulation). They make the point that a successful simulation must be "consistent" (a variable cannot be caused to have two different values), be "corresponding" (agree with empirical facts) and be "robust" (apply to a wide range of situations). Presumably at the start structure and function are disparate, and then are correlated in the testing and iterative development of the simulation.

In considering how things work (or how they don't work) Rouse (1980) has pointed out the tendency of human subjects in fault searching tasks to seek confirming evidence for fault hypotheses but pay little attention to disconfirming evidence which may be just as useful in discriminating. This would suggest that people also may neglect or not know how to use evidence of what doesn't function or what the structure is not.

Rusmussen (1981), again concerned primarily with fault detection, has made a useful distinction between "topographic" and "symptomatic" search. Topographic search is systematically observing over a whole set of given items to find one or more items which stand out as being different by some given criterion. Symptomatic search means searching through a whole set of symptoms or patterns or criteria to find one which matches the observation on a given item. Both strategies may apply to search for cause of a given consequence, search for structural connexity, search for a plausible explanation or search for a fault.

In the present context of supervisory control the search for cause-consequence, connexity and explanation is applied to the controlled process. Without understanding how that process works presumably one can control only by making actual trials and errors. With understanding one can run mental experiments, "simulations" in DeKleer's terminology, make predictions and determine what control is likely to work before commiting to a single actual trial. In this manner one can readily discover polarity and

sensitivity of effect, order of integration, or essential nonlinearities.

Along with understanding of the endogenous input-output characteristics of controlled processes is understanding of the exogenous properties of the external forcing functions and constraints - disturbances, obstacles, time and resource limits, etc.

Later some experimental paradigms are presented for studying how subjects use computer aiding to discover how things work and how best to control them.

3.3 Determining what is wanted

In order to do mathematical optimization the analyst customarily presupposes an objective function which specifies which performances or states of the system under consideration are good and which are bad and what is the precise tradeoff between achieving different levels of different objectives or "goods". The constraints on resources, combined with the physical laws governing the behavior of the controlled process generally prevent the process from achieving the highest level of every objective; some compromise must be made. The question is what compromise is best, and the analyst can (usually) determine this, or at least come close, by simultaneous solution of the given objective equations and the given process equations.

With real people and real controlled processes to talk of optimal control is mostly nonsense. A precise objective function is seldom if ever available. Real human operators have a sense of what is better and what is worse relative to situations they have thought about. To be posed in rapid-fire succession with a large number of hypothetical situations (combinations of attributes) which are far from what they think they want is very difficult for people to respond to. This is especially true if these situations are cast in the form of lotteries as required by multi-attribute utility theory, e.g., "would you rather have a lottery consisting of 0.7 probability of x_1 and 0.3 probability of x_2 or would you prefer an even chance of x_3 vs x_4?" For these reasons, many multi-attribute utility measuring techniques, though theoretically rigorous, are empirical disasters.

A somewhat different approach is proposed by March and Simon (1958) who call it "satisficing". The idea is to discover the neighborhood of achievable state space with which a person (judge) is satisfied that he can't do very much better. And the idea is to avoid having to compare states which are unrealistic (undesired or unachievable).

It is clear in any case that the supervisory controller does have a significant problem in determining what he most wants from the bewildering array of what he can have, and in pondering this choice typically doesn't even know what is achievable and what is not. (This will be dealt with further in the next section).

A different aspect of the human operator determining what is wanted is the mediation between determining what he himself wants and what some outsider (the boss, the co-worker, the regulation), wants. Insofar as there is good calibration on what is wanted by several different entities (a notion yet to be developed) there should be a "confidence" (see Section 3.5) to go ahead.

3.4 Mental manipulations of fuzzy chunks

Miller (1956), in his researches on immediate memory, has made convincing the idea that humans recode detailed information in coarser pieces called "chunks", and that this chunking can be hierarchically recursive (fleas have little fleas, etc). Ordinary experience strongly suggests that such chunks are fuzzy, i.e., that certain concepts or mnemonics have a stronger meaning or elicit stronger association or recall for some attribute states then for others.

With respect to knowledge, one can always test how precise is a subject's expectation over specified unambiguous states of the world. However if the only way a subject can encode that expectation is to use fuzzy sets which spread over a large number of states, then that subject cannot discriminate very well. The greatest achievable concentration of expectation (over the given states, not over the fuzzy sets) is where there is no chunking and no fuzziness. (Wisdom may be expressed in simple terms - a few

chunks arranged in a pithy but fuzzy statement. But if those chunks or fuzzy terms do not permit discrimination with respect to what they don't mean, then surely there is no wisdom).

A person's knowledge may be stated in expectation of terms that we would normally call fuzzy (e.g., 0.7 expectation of small, 0.3 expectation of large) and another person's knowledge may be similarly stated (e.g. 0.6 expectation of small, 0.4 expectation of large). Calibration of one person's knowledge against that of the other is straightforward if the (fuzzy) terms are the only ones that are used - with no need to worry about how fuzzy memberships spread over some set of attribute states. However when one is interested in calibration with respect to attribute states and the "large" and "small" fuzzy membership functions of the two persons differ as well as their expectations of large and small, then the calibration problem is more interesting. This real life problem is only suggested here; no solution has been developed, but we are giving thought to it.

3.5 Combining evidence and confidence

It is known that people are conservative decision-makers in using evidence they have available to update their subjective expectations (Phillips, Hayes and Edwards, 1966). In many ways they hedge their judgements toward the mean, toward "no expectation" or the "maximum entropy position". This would be irrational were their sensing, memory and data processing noise-free. It makes sense, however, if viewed as rationality embedded in self-noisiness, or as lack of calibration of one's knowledge or good-bad criteria with those of another signifiant entity (human or computer or true state of the world).

After many empirical trials one usually can be subjectively confident that a coin is fair, i.e., 50-50. After no trials at all one may expect heads and tails on a 50-50 basis but with no confidence. After 9 heads and one tail the empirical knowledge suggests 9 to 1, but the confidence limits on the probability militate in the direction of the 5 to 5, and awareness of self-noisiness may militate still further in that direction.

The best decision-maker will know when to have confidence (when to predict his own success or predict that his empirical knowledge will yield the best result) and when to hedge the expectation toward the mean because of assumed internal noise or entropy or poor sample. Figure 10 suggests a way of analyzing a person's skill at doing this, using a conventional relative operating characteristic (ROC) curve of signal detection theory. In this case, one asks the decision-maker to predict his own success ("S"). One then cross-plots the probability of predicting success given actual success (S) vs. probability of prediciting success given actual failure (F). The upper left hand corner is perfection, the diagonal is the line of guessing, and the curve shown is typical of where subjects might lie on average. Most important, the lower-left to upper-right direction corresponds with the decision-maker's general confidence in himself (i.e., to predict his own success). Note that degree of true success is an orthogonal direction on the graph to degree of confidence. This use of signal detection theory is another idea we intend to pursue further.

3.6 Deciding what to do

If one knows how things (the controlled process, the forcing functions, the obstacles, the resources) work, if one is satisfied as to what is wanted (and satisfied with the consistency of what different authorities want), if one has coded (chunked) at a proper level and not in a different way than another significant person or computer, if one is properly atuned as to when to accept the evidence and when to hedge toward the no-confidence (entropic) position, then one is ready to decide what to do.

In complex supervisory control systems these factors may require more mental capacity than a human supervisor has, and a decision-aid is likely to be of help.

4. COMPUTER MODELS AND DECISION AIDING

4.1 Two types of computer models

In section 1 the idea was introduced that models of external reality can

be represented inside the computer. Two types of such models were identified as appropriate to supervisory control: (1) model of a dynamic process which serves to "observe" or estimate state variables which are inconvenient to measure, from which automatic control is implemented; (2) a model of a process which, together with a user interface, provides a human operator advice and responds usefully to queries he makes.

Figure 11 diagrams the first type as is used in linear control systems. Because this use of internal models is well known, and its operation at the automatic (task interactive computer) level of supervisory control is straightforward and well documented, we do not concern ourselves further with it in this report. We are more interested in exploring the uses and abuses of the second type of computer model and how it interacts with the human operator's mental model. These days this latter type of computer use is called an "expert system". More generally we have called it a "decision aid".

4.2 Expert systems: general needs

The knowledge base component of an expert system (Figure 12, Gevarter, 1982) is usually thought of as a set of "if-then" production rules. Possibly separate from this general store of rules may be a model of the system status, candidate hypotheses, or conclusions reached thus far about the present reality. A human expert may input rules; also, data from the ongoing process may be input to the system status model. In an executive position over these elements may be a rule interpreter and whatever else is needed to provide an interface with a human user. (Whether the system status model is separate from the general store of rules depends on the type of model, in conventional equation-based simulations they are integrated).

Expert systems can operate top-down (deductively), i.e., deduce detailed statements from abstract knowledge of how thngs work. They can also operate bottom-up (inferentially). They can be driven by events (what is the meaning of what just occurred?) or by cumulative data (what does it all mean?) or by expected results or goals (how can we get there and how near are we getting?)

Both Feigenbaum (1980) and Buchanan (1981) have commented on the

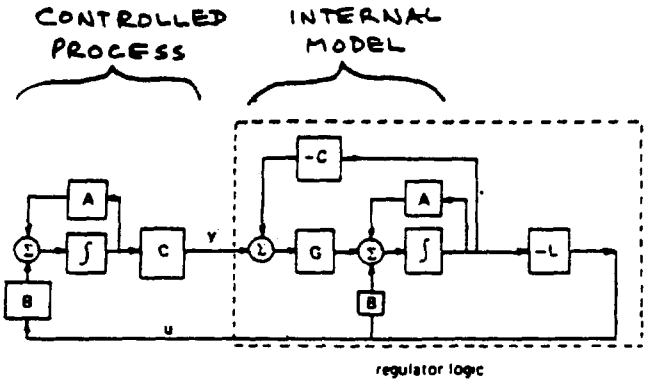


Figure 11. Computer model internal to modern control system (for purpose of estimating)

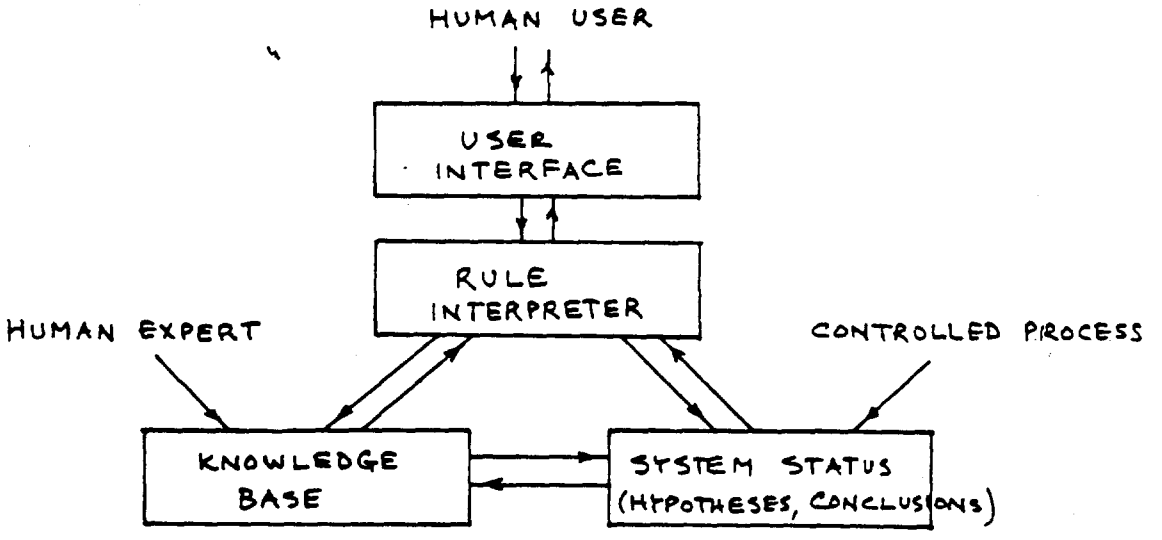


Figure 12. Components of an expert system

"bottleneck" of acquiring knowledge from a human expert, on the need for steering the expert to provide the right kind of information, on working with multiple experts, and on making expert systems user-friendly. Bonissone (1979) has discussed sources of uncertainty in expert systems as: (1) unreliability of information in the knowledge base; (2) imprecision of language in the rules; and (3) incomplete information. He suggests representing uncertainty as a fuzzy interval, with the degree of necessity as a lower bound and degree of possibility (usually associated with the fuzzy membership function) as an upper bound.

4.3 Expert systems for supervisory control

It is clear that expert systems can be useful to help the supervisory controller with the key mental activities described in Section 3.1: (1) discovering how things work; (2) determining what is wanted,(3) manipulation of fuzzy ideas or "chunks"; (4) combining evidence with confidence; and (5) deciding what to do.

The first of these is a fundamental problem of both artificial intelligence and cognitive psychology, and we hope to pursue it later in our research. The second problem is seen as particularly relevant for supervisory control, is one we are now experimenting with, and will be discussed in Section 4.4. We have explored the third problem experimentally in the context of failure detection, and a summary of that work is provided in Section 4.5. The fourth problem is one we are considering pursuing with respect to expert systems, but it will not be discussed beyond what was said already in Section 3.4. Some experiments are in progress in the fifth area, particularly with regard to graphic machine aids for dynamic problems, and these are described in Section 4.6.

4.4 Expert systems to help the operator decide what he wants

Wierzbicki and colleagues (1982) have devised a simple computer aid to implement "satisficing", which may be described as follows (Figure 13). Consider a multiattribute space. For purposes of simplicity Figure 9 shows a space of only two attributes x and y. The computer is given the complete set

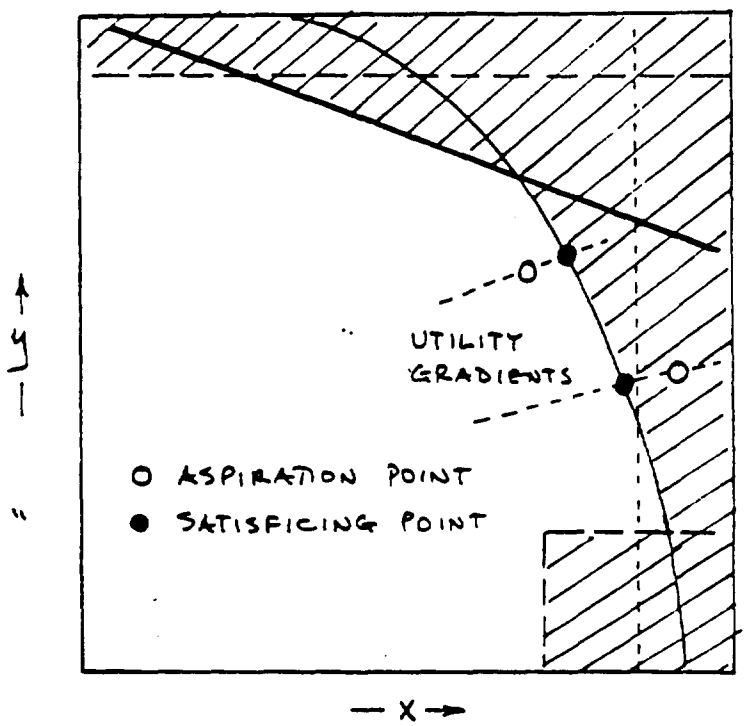


Figure 13. Simple example of satisficing

of constraining equations, which might be absolute limits which can be tolerated (e.g., on time, money, force, energy) or may be tradeoffs between these variables imposed by the laws of nature for the particular controlled process and/or environment. This means that a particular region of state space is unachievable (shown shaded). Initially, however, the computer has no knowledge of what the human user wants or how he would order the achievable states on a good-bad scale. (Nor, indeed, does the user usually know explicitly what is wanted; if he did he could make an immediate choice of what he likes best from what is available to him).

The satisficing procedure starts with the user selecting some reasonable "aspiration" state that he might like to achieve, and indicating which directions of the attributes he would consider an improvement if he could have them and roughly how important those attributes are relative to each other. From this information the computer can determine which is the "best" point on the boundary of what is achievable based on the aspiration point provided and the good-bad gradient relative to this aspiration. This "satisfacing point" may be better or worse than the aspiration state.

The user, having discovered one satisficing state that can be achieved (that presumably is in the neighborhood of what he wants) may then wish to pose a few other aspiration states which differ by some attributes. He may think, for example, "well, if I can't have as much of x as I'd like maybe at a different y I can achieve more of x" or "if I can do better than expected relative to that compromise let me try another combination and maybe I can do even better still". In this way few or many alternatives can be explored until he is fully satisfied (satisficied!). Never in the process is he posed with comparing states which seem unrealistic (undesired or unachievable) and never is he confronted with weighing hypothetial probabilities of consequences.

James Roseborough (1984) is developing a flexible computer-graphic system to provide this capability in a dynamic context (Wierzbicki's demonstration was with a static problem). This will serve as an experimental apparatus for studying how supervisory controllers might interact with such a decision aid to explore options and to clarify what they want. The

particular example problem Roseborough has chosen is that of maneuvering a large vehicle into a terminal location, where the following attributes of performance must be traded off: (1) time to complete the task; (2) energy expenditure; (3) deviation from an ideal path (to avoid obstacles); (4) deviation from an ideal final location. To determine these dependent variables for different ship trajectories some complex nonlinear equations must be solved. The emphasis, however, is on discovering what facilitates human decision of what is wanted.

4.5 Experiments toward the end of an expert system for failure detection using fuzzy sets

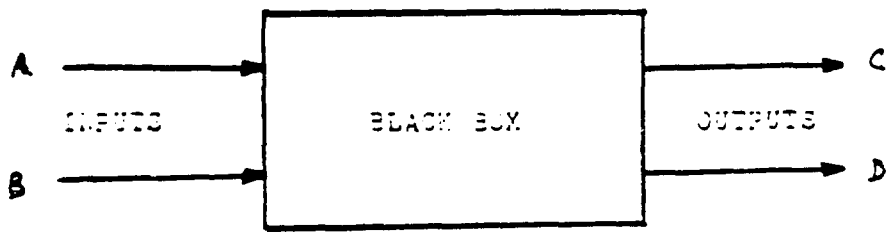
Now we come to the use of fuzzy sets, an area in which a master's thesis of Frank Laritz has been completed (1983). That thesis is summarized here; the complete document is available.

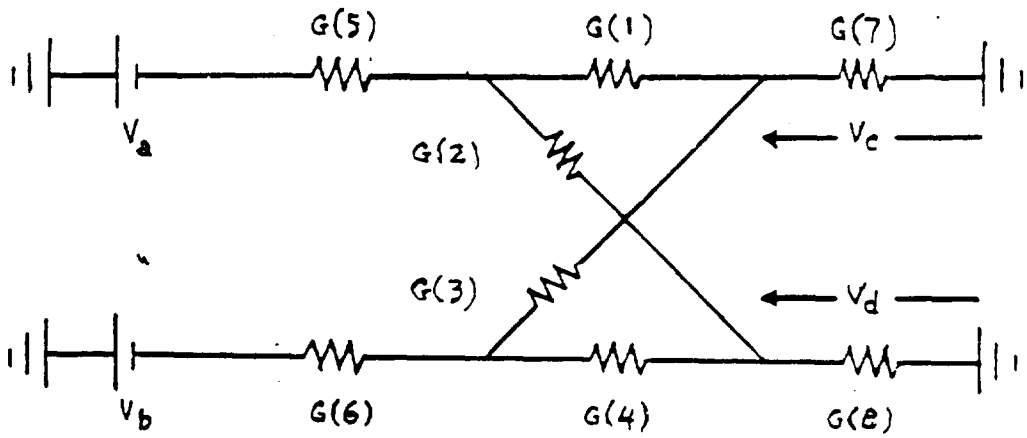
Laritz had five subjects repeatedly adjust two "inputs" A and B to a "black box" to any value between 10 and 100, set a "failure mode" to any one of four available settings including "no failure", and observe two "outputs" C and D. The contents of the black box were not revealed. The subjects' task was to correlate inputs and outputs with failure modes and from this infer rules by which to assert whether and in what mode the black box had "failed" as a function of the two inputs and two outputs.

Actually the black box was a simple resistor network as shown in Figure 14 in which one of the resistors 1,2,3,4 was selectively opened (or none was).

After each subject had completed a number of trials (they were all really learning trials) he was asked to formulate rules in terms of easy-to-remember descriptors for the four variables like "low", "medium" and "high". Using these descriptors he was to generate rules such as:

"when A is low and B is medium or high and C is high and D is medium or high, the failure is mode 2".





Resistors Allowed to Fail: G(1) = 61 mhos
G(2) = 87
G(3) = 59
G(4) = 76

Other Resistors: G(5) = 95
G(6) = 55
G(7) = 74
G(8) = 85

Figure 14. Simple resistor network comprising Laritz' "black box"

There could be any number of such (fuzzy) descriptors and any number of such rules, and the subjects were free to format them in tables or however they wished. They could also combine variables in forms such as C/D and C-D.

The subjects were also asked to produce functions of each descriptor (fuzzy set) defining what they "meant". Each function specified "membership" m or "truth" as a function of the values of the corresponding variable (in the range 10-100). Two of the five subjects observed the black box behavior first, then devised the rules, and lastly devised membership functions. The others chose to invent terms and define the membership functions first.

As an example Figure 15 lists the rules given by one subject (JR) and Figure 16 presents his membership functions. Note that certain regions of A,B and D-C were (apparantly intentionally) not covered by his membership functions (and rules). For contrast the membership functions of a second subject DM are also shown (Figure 17).

For each subject independently the experimenter derived the state-action matrix (failure mode as a function of input and output numerical values) using the conventioanl "max m" for "OR" and "min m" for "AND". He then proceeded to evaluate each resulting expert system not only against single complete failures (the basis on which the subjects made up their rules) but also on multiple complete failures and single partial failures (5% changes rather than 100% changes in resistance). For a given set of inputs and outputs each subject's expert system yielded a "truth value" for each failure mode for each combination of A,B,C,D. A simple procedure is to assert failure for that mode having the greatest truth value greater than some threshold and no failure for truth less than that threshold. Laritz used this as one decision criterion (which he called the "most true" criterion) but also counted the number of times m for each mode exceeded 0.5 (the "times true" criterion), and the sum of truth values for each mode ("truth summation" criterion). Figure 18 summarizes the rather impressive success of subject JR's expert system, and for comparison Figure 19 summarizes that of subject DM. The performances of the other fuzzy expert systems lay somewhere in between.

(1,1) If A is high and B is low and D is significantly greater than C, then the system is in failure mode 1.

(2,1) If A is high and B is low and C is significantly greater than D, then the system is in failure mode 2.

(3,1) If A is low and B is high and D is significantly greater than C, then the system is in failure mode 3.

(4,1) If A is low and B is high and C is significantly greater than D, then the system is in failure mode 4.

(5,1) If A is high and B is high and D is slightly greater than C, then the system is in failure mode 0.

Figure 15. Fuzzy decision rules inferred by subject JR

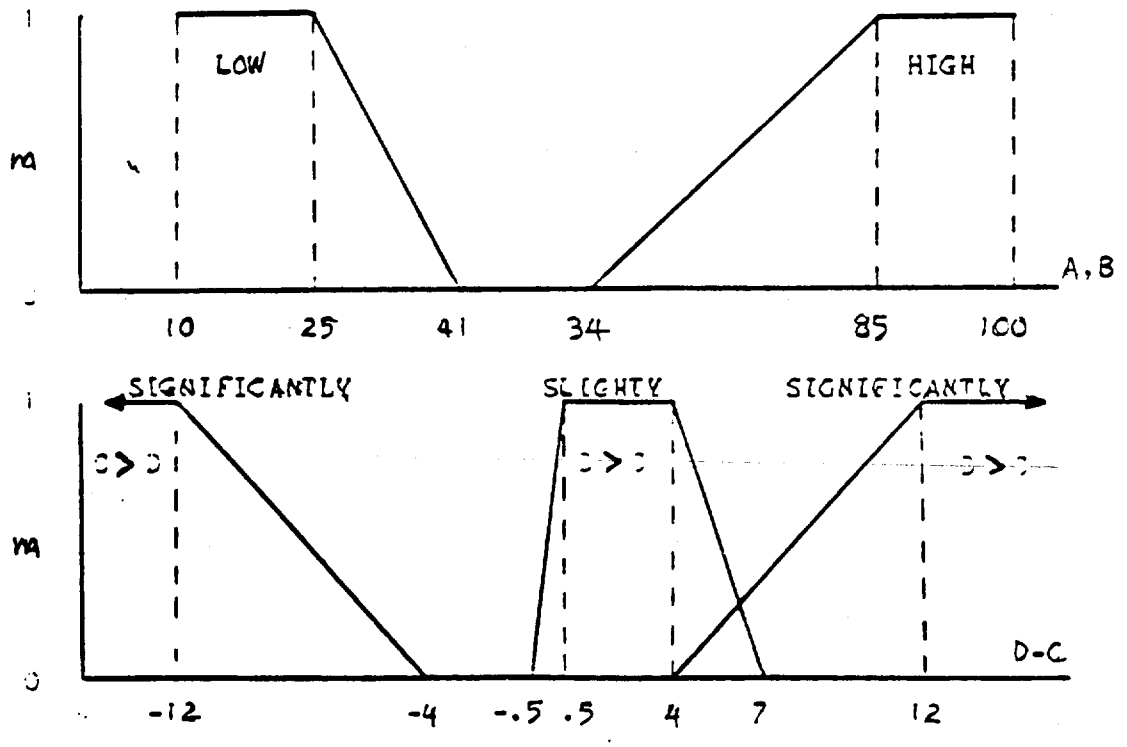
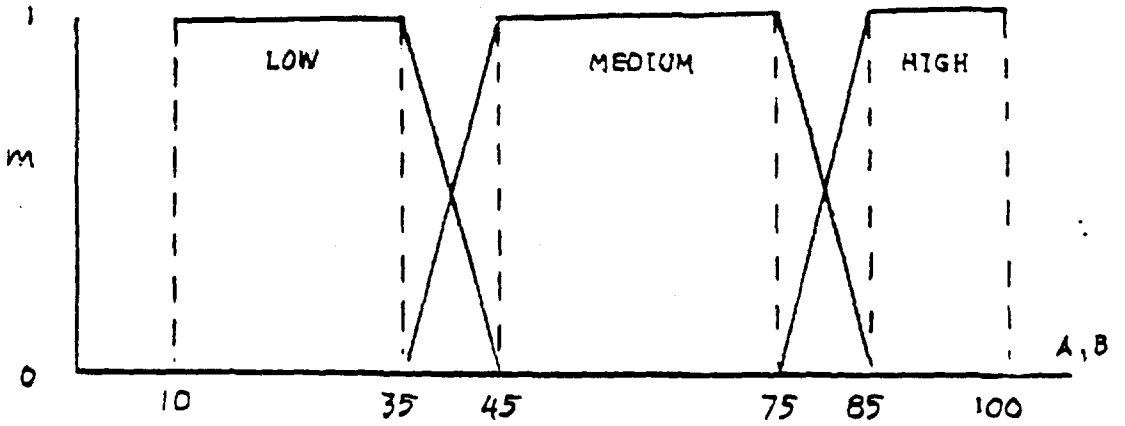


Figure 16. Membership functions devised by subject JR



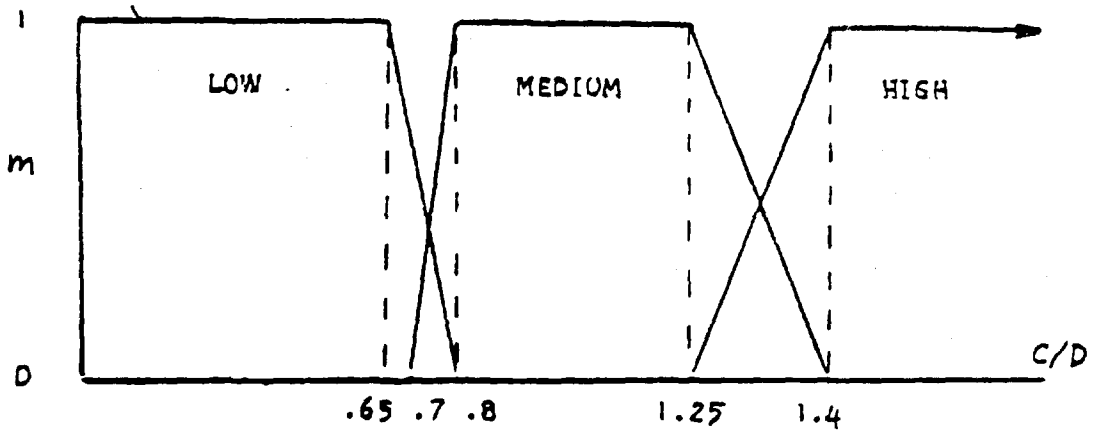


Figure 17. Membership functions devised by subject DM

TEST 1: SINGLE COMPLETE FAILURE

ACTUAL FAILURE	IDENTIFIED FAILURE		
	MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
0	0	0	0
SCORE:	5/5	5/5	5/5

TEST 2: MULTIPLE COMPLETE FAILURES

ACTUAL FAILURES	IDENTIFIED FAILURE		
	MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1,3	3	3	1
1,4	1,4	1	1
2,3	2,3	3	3
2,4	2,4	2	2
SCORE:	4/4	4/4	4/4

TEST 3: SINGLE PARTIAL FAILURE

ACTUAL FAILURE	IDENTIFIED FAILURE		
	MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1	1 (55%)	1 (50%)	1 (55%)
2	2 (75%)	2 (70%)	2 (55%)
3	3 (70%)	3 (70%)	3 (65%)
4	4 (80%)	4 (75%)	4 (65%)
SCORE:	4/4	4/4	4/4
TOTAL SCORE:	13/13	13/13	13/13

* = INCORRECT DECISION

Figure 18. Results of applying JR's expert system

TEST 1: SINGLE COMPLETE FAILURE

ACTUAL FAILURE	IDENTIFIED FAILURE		
	MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1	1	1	1
2	* 1,2	2	2
3	* 1,3	* 1	* 1
4	* 1,4	4	4
0	0	0	0
SCORE:	2/5	4/5	4/5

TEST 2: MULTIPLE COMPLETE FAILURES

ACTUAL FAILURES	IDENTIFIED FAILURE		
	MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1,3	1,3	1	1
1,4	* 1,4,0	1	1
2,3	* 2,3,4,0	2	2
2,4	2	2	2
SCORE:	2/4	4/4	4/4

TEST 3: SINGLE PARTIAL FAILURE

ACTUAL FAILURE	IDENTIFIED FAILURE		
	MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1	1 (80%)	1 (55%)	1 (50%)
2	* 4	* 4	* 4
3	3 (70%)	3 (65%)	3 (60%)
4	* 1,4	4 (30%)	* 1
SCORE:	2/4	3/4	2/4
TOTAL SCORE:	6/13	11/13	10/13

* = INCORRECT DECISION

Figure 19. Results of applying DM's expert system

From this part of the experiment Larritz concluded:

1. The method of observing trends, then formulating rules, and then defining fuzzy values captures more of the human's ingenuity and pattern recognition ability and provides a better expert failure detection system than the method or creating fuzzy values, then gathering data, and then deducing rules.

2. If the second method is used, it is best to put the membership functions for the fuzzy values on paper at the outset so that there will be no loss of information later.

3. Expert systems using non-fuzzy values require perfect failure rules. When the rules are not perfect, the expert system does not perform well.

4. Although not explicitly defined for this purpose in the investigation, the fuzzy expert systems did remarkably well in detecting and locating multiple and partial failures. This means that fuzzy methods have some robustness.

5. The decision method can be chosen to suit the strength and tightness of the rules. Stronger rules require less margin for error.

6. Expert systems which have approximately the same number of rules for each failure mode perform better than those with an uneven distribution

As a second experiment Laritz used himself as a subject on a black box resistor network that was much more complex (sufficiently so that he had no advantage over a subject who did not know what was inside). Again there were two adjustable inputs and two resulting outputs but this time eight failure modes. Laritz first experimented and observed, then derived his rules (Figure 20), then defined his membership functions, and finally derived his expert system on the same basis as before. His results showed that his expert system worked perfectly on complete failures but faltered on multiple

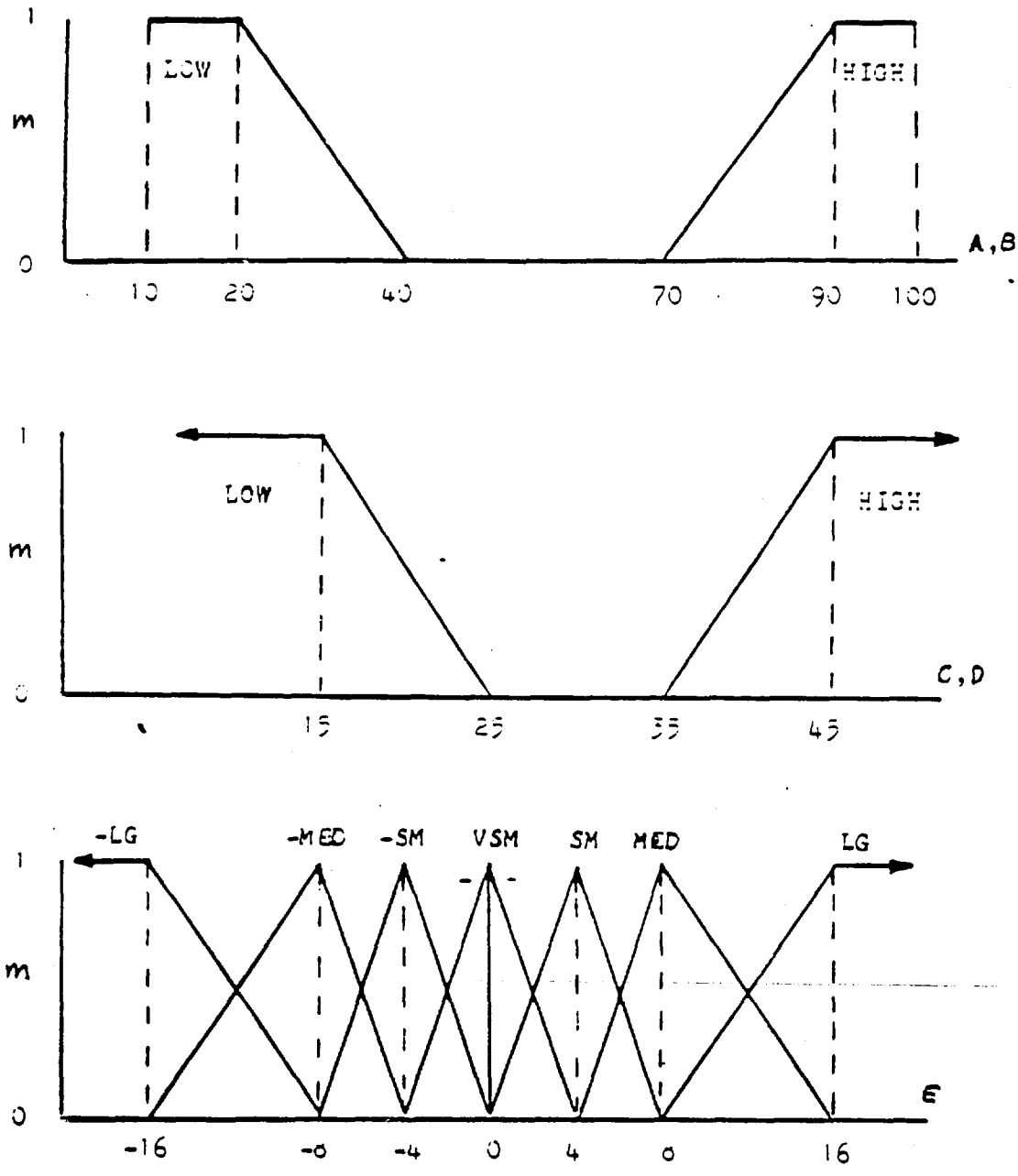


Figure 20. Membership functions devised by Laritz for variables of complex resistor network

complete failures and partial failures (Figure 21). Further attempts to refine his decision rules showed little gain in discriminabililty.

These preliminary experiments with a small number of subject strongly suggest that a fuzzy expert failure identification system, given relatively little knowledge from persons who are "expert" in the behavior of a sufficiently simple system under complete failures, can perform well in such identifications. But when the system is complex and failures are multiple or partial and the expert's knowledge is not derived on the basis of experiencing such failures, such an expert system cannot be expected to perform very well.

4.6 Experiments in graphic decision-aids for dynamic trajectory planning

The purpose of these experiments, being done by Leon Charny (1984), is to explore the relative allocations of decision-making to human and computer in the process of deciding what to do. The computer in this case is assumed to have at least some of the constraint information, some of the relative worth information, some search-and-compare capability, and be able to communicate with the human user in some common terms. At the same time it seems reasonable to expect the human to know some things and do some searching not shared by the computer.

We have chosen a "simplest possible" form of representing a dynamic decision problem - both conceptually, to help ourselves as researchers to abstract, understand and generalize - and in the form of a graphic display to make the task clear for our experimental subjects. The experimental task (Figure 22) is to choose, at each of a succession of time steps, a single response. The initial constraint is that responses adjacent in time must also be adjacent with respect to the response scale (could be magnitude, spatial location, etc.); in other words the response is continuous (apart from the discrete reticulation of the time and response scales). Other constraints may be added such as obstacles, local acceleration, limits, etc.

In first informal experiments, Charny assigned semi-random numbers to all response states (every combination of response level and time step),

TEST 1: SINGLE COMPLETE FAILURE

ACTUAL FAILURE	IDENTIFIED FAILURE: MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
0	0	0	0
SCORE:	9/9	9/9	9/9

TEST 2: MULTIPLE COMPLETE FAILURES

ACTUAL FAILURES	IDENTIFIED FAILURE: MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1,2	2	2	2
1,3	1	1	1
1,4	1	1	1
1,5	1	* 1,7	1
1,6	1	1	1
1,7	* 5	* 5	* 5
1,8	8	8	8
2,3	2	2	2
2,4	4	4	4
2,5	* 4	* 4	5
2,6	6	6	6
2,7	7	7	2
2,8	2	* 1,2	* 1
3,4	4	4	4
3,5	5	5	5
3,6	6	6	3
3,7	7	7	7
3,8	8	8	8
4,5	* 7	* 7	* 7
4,6	6	6	6
4,7	* 5	* 5	* 5
4,8	* 6	* 6	* 6
5,6	* 1	* 1	* 1
5,7	* 0	* 0	* 0
5,8	* 2	* 0	* 2
6,7	* 4	* 8	* 8
6,8	6	6	6
7,8	* 4	* 4	* 1
SCORE:	18/28	16/28	18/28

TEST 3: SINGLE PARTIAL FAILURE

ACTUAL FAILURE	IDENTIFIED FAILURE: MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1	1 (55%)	1 (55%)	* 2
2	2 (65%)	2 (65%)	2 (25%)
3	3 (60%)	3 (60%)	* 2
4	* 3	* 3	* 2
5	* 2	* 2	* 4
6	* 8	* 8	* 8
7	* 4	* 4	* 2
8	8 (55%)	8 (55%)	* 2
SCORE:	4/8	4/8	1/8
TOTAL SCORE:	31/45	29/45	28/45

* = INCORRECT DECISION

Figure 21. Results of applying Larritz' expert system to failures of complex resistor network

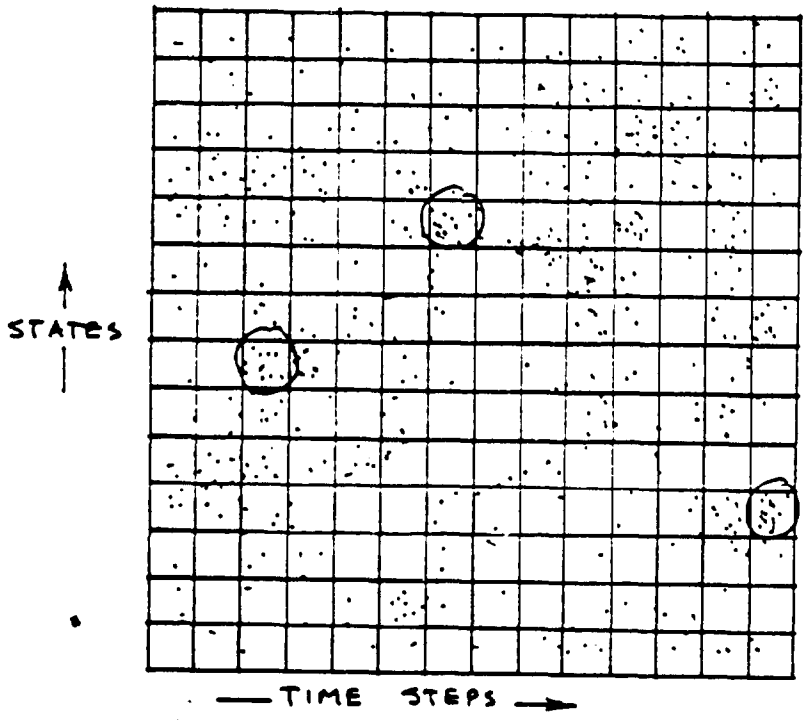


Figure 22. Charney's experimental task

implying the reward for transversing that state. These rewards were cumulative, and the goal was simply to amass the greatest possible cumulative reward. In these first experiments the computer display simply brighted at various state points as a function of incremental reward, and our (initially skeptical) hypothesis was that the operator, having such a nice qualitative analog display, could visually identify the best or near best paths with ease. This proved not well founded.

Next experiments had the operator spotting points or regions of state space intermittently, then having the computer use a "greedy algorithm", simply taking the most lucrative next step as it went along to connect the intermittent points chosen by the operator. This did not work so well either.

What worked much better was to have the human specify intermittent points ("I think the path should go through here") again while looking at the state-by-state brightness displays, then have the computer do a dynamic program solution for intervening points. This is much easier and faster for the computer than to do a dynamic program for the whole trajectory, since by human selections of intermittent points large regions of state space may be eliminated from computer consideration. This is because of the constraint that at adjacent time steps the states must be adjacent, which limits the computer search to a very small number (e.g., 3^2 in the simplest case of one attribute dimension as compared to N^2).

Because the above problem is extremely simple for the computer, even as the state space is made slightly larger or two dimensional, we may obtain a resource cost measure (e.g., computer time) for each trajectory. In order to explore what might be more typical in a more complex application we could require the operator in our experiments to make some tradeoff between the cost of his own time and that of the computer's time - in addition to trying to maximize his cumulative reward from the game.

Clearly if the computer were asked to do dynamic programs connecting every other human-selected point then it is possible that a different trajectory might be obtained, and the earlier combined man-machine result

would be inferior to a fully machine-determined result. So far this has not been the case; instead the experiments served to verify that the human hunches were correct. In any case when the machine doesn't know about certain constraints or because of inordinate amounts of dynamic programming or if time is simply too costly to use over the whole state space the computer cannot be trusted to make more global verifications of the human's intuition.

We are also interested in what happens when the human operator may then review what the computer decided was best. He may find that it has done something suprising that was foolish or did not respect a constraint that he, the operator, had been expecting it to respect. By iterating the procedure once or twice one might see significant refinements.

REFERENCES

Bridgeman, P.W., 1928, The Logic of Modern Physics, N.Y. Macmillan.

Bonissone, P., 1979, The Problem of Linguistic Approximation in Systems Analysis, Ph.D. Thesis, Dept. of EECS, Univ. of Michigan, Ann Arbor.

Buchanan, B.G., 1981, Research on Expert Systems, Stanford Univ. Computer Science Dept., Rep. No. STAN-CS-81-837.

Charney, L., 1984, Ph.D. thesis in progress.

Cooke, Mendel and Thijs, 1984, Calibration and Subjective Knowledge, Delft Univ. of Technology, Netherlands, Unpublished paper.

DeKleer, J. and Seely-Brown, J., 1982, Foundations of Envisioning, Proc. AAAI National Conf. on Artificial Intelligence.

Fergenbaum, E.A., 1980, Knowledge Engineering, The Applied Side of Artificial Intelligence, Computer Science Dept., Memo HPP-80-21, Stanford Univ., July.

Gevarter, W., 1982, An Overview of Expert Systems, Nat'l. Bu. Stds., Report NBSIR 82-2505, May.

Keeney, R.L. and Raiffa, H., 1976, Decisions with Multiple Objectives, N.Y., Wiley.

Laritz, F.J., 1983, The Use of Fuzzy Sets in Failure Detection, MIT SM Thesis, December.

March, J.G. and Simon, H.A., 1958, Organizations, Wiley, N.Y.

Miller, G.A., 1956, The Magical Number Seven Plus or Minus Two; Some Limits on Our Capacity for Processing Information, Psychol. Rev. 63, 81-97.

Phillips, L.D., Hayes W.L. and Edwards, W., 1966, Conservatism in Complex Probablistic Inference, IEEE Trans. Human Factors in Electronics, 7, 7-18.

Rasmussen, J., 1981, Models of Mental Strategies in Process-Plant Diagnosis, in Rasmussen J. and Rouse W.B. (eds.), Human Detection and Diagnosis of System Failure, Plenum Press, N.Y.

Roseborough, J., 1984, Ph.D. thesis in progress.

Rouse, W.G., 1980, Systems Engineering Models of Human-Machine Interaction, N.Y., Elsevier Press.

Sheridan, T.B., 1982, Supervisory Control: Problems, Theory and Experiment for Application to Human-Computer Interaction in Undersea Remote Systems, MIT Man-Machine Lab Rep., March.

Sheridan, T.B., 1983, Supervisory Control of Remote Manipulators, Vehicles and

Dynamic Processes: Experiments in Command and Display Aiding, MIT Man-Machine Systems Lab. Rep. March.

Wierzbicki, A. P., 1982, A Mathematical Basis for Satisficing Decision-Making, Mathematical Modeling, 3, 391-405.

Zadeh, L. A., 1965, Fuzzy Sets, Information and Control, 8, 338-353.

GLOSSARY

CALIBRATION - the degree to which the expectations of one person or machine correspond to those of another person or machine.

COMPUTER MODEL - a computerized representation of the state or configuration of the controlled process or environment, used either as part of an automatic controller (estimator, observer) or as part of an expert system (knowledge representation).

COMPUTER ESTIMATOR, OBSERVER - a computer model which, when calibrated to make some state variables correspond to the actual controlled process, provides an estimation of other state variables which are difficult to measure in the actual controlled process.

CONTROL DECISION - judgement by a human operator or a machine about what actions to take to make the controlled process improve with respect to the objective function, given the task constraints.

CONTROLLED PROCESS - a machine and/or natural system which a human operator and/or another machine forces to conform to a given criterion.

DECISION AID - a computer or other device which provides useful information to a human operator in making worth, constraint or control decisions

DYNAMIC - changing with time.

ENVIRONMENTAL OBJECTS/FORCES - objects and forces in the environment of the controlled process which affect its behavior.

EXPERT SYSTEM - a computer-based decision aid programmed with assertions obtained from a human expert in relevant subject matter to answer questions posed by a naive person.

FRAME - a general device for storing knowledge including production rules,

declarative facts, parameters, descriptors, semantic networks and pointers to other frames.

FUZZY SET - a set of objects or events (attribute states) which can be referred to by the same term which vary in their strength of meaning (membership, truth) with that term.

HUMAN INTERACTIVE COMPUTER - a computer or part of a computer designed to communicate with a human operator to interpret his commands or to answer his questions.

HUMAN INTERACTIVE SYSTEM - that part of a supervisory control system consisting of a human operator, a human interactive computer, displays and hand-controls.

HUMAN OPERATOR - a person who intentionally interacts with a machine in order to produce a result.

INTERNAL MODEL - a mental representation of relevant information in the mind of a human operator, or a computer representation of relevant information.

KNOWLEDGE - (a) coded information representing some characteristics of an object or event; (b) the degree to which an expectation for some classes of events is greater than that for others; (c) formally $\sum_i p_i \log p_i/c$ or "relative negentropy".

MAN-MACHINE SYSTEM - a system in which a person interacts with a machine to produce an observable result.

MEMBERSHIP FUNCTION - the quantitative specification of the relative strength of meaning (membership, truth) of various objects or events (attribute states) associated with a given term.

MENTAL MODEL - a hypothetical mental representation of the state or configuration of the controlled process and/or environment.

MULTIPLEX - to alternately connect a communication channel to different information sources and/or sinks.

OBJECTIVE FUNCTION - a scalar function of significant performance variables which defines overall goodness (worth, utility) and scales tradeoffs for relative goodness among all significant combinations of states of performance variables.

OPTIMIZE - to maximize an objective function

PRODUCTION RULE - a statement that if certain events are true then other events will occur.

SATISFICE - to achieve a marginally achievable level of performance judged to be satisfactory.

STATE - a particular configuration of an object or event, defined by the concurrent value of each of its variables or attributes.

STATE ACTION MATRIX - the specification of what action or response to make for each state of a system.

SUPERVISORY CONTROL - a situation wherein a person interprets what is happening with the aid of a computer and specifies subgoals to a computer which automatically strives to implement those subgoals.

SYMPTOMATIC SEARCH - systematic test of an object or event against a number of criteria to determine which criteria, if any, are met.

TASK - in a supervisory control system, all elements exterior to the referenced element or set of elements. The human operator's task is to satisfy the objectives by acting on the human interactive computer, controlled process and environment, insofar as these elements exist. The task of the human interactive system is to control the task interactive system, while the task of the latter is to accomodate the former.

TASK INTERACTIVE COMPUTER - that part of a supervisory control system consisting of a task interactive computer, a controlled process and associated actuators and sensors.

TASK CONSTRAINTS (PROCESS CONSTRAINTS, or just CONSTRAINTS) - rules that govern the way the controlled process must behave, including "laws of nature" and limits on resources such as time, energy, memory and money.

TASK INTERACTIVE SYSTEM - that part of a supervisory control system consisting of a controlled process, a task interactive computer, and associated artificial sensors and actuators.

TEAM DECISION MAKING - interaction of two or more persons with each other and/or with a machine to make a decision.

TOPOGRAPHIC SEARCH - systematic comparison of all objects or events in a set to look for those which stand out by some criterion.

UTILITY DECISION - judgement by a human operator of the relative goodness, worth or utility of some object or event, experienced in reality or specified as a hypothesis in terms of a combination of states of performance variables.

TEST 1: SINGLE COMPLETE FAILURE

ACTUAL FAILURE	IDENTIFIED FAILURE		
	MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
0	0	0	0
SCORE:	9/9	9/9	9/9

TEST 2: MULTIPLE COMPLETE FAILURES

ACTUAL FAILURES	IDENTIFIED FAILURE		
	MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1,2	2	2	2
1,3	1	1	1
1,4	1	1	1
1,5	1	* 1,7	1
1,6	1	1	1
1,7	* 5	* 5	* 5
1,8	8	8	8
2,3	2	2	2
2,4	4	4	4
2,5	* 4	* 4	5
2,6	6	6	6
2,7	7	7	7
2,8	2	* 1,2	* 1
3,4	4	4	4
3,5	5	5	5
3,6	6	6	3
3,7	7	7	7
3,8	8	8	8
4,5	* 7	* 7	* 7
4,6	6	6	6
4,7	* 5	* 5	* 5
4,8	* 6	* 6	* 6
5,6	* 1	* 1	* 1
5,7	* 0	* 0	* 0
5,8	* 2	* 0	* 2
6,7	* 4	* 8	* 8
6,8	6	6	6
7,8	* 4	* 4	* 1
SCORE:	18/28	16/28	18/28

TEST 3: SINGLE PARTIAL FAILURE

ACTUAL FAILURE	IDENTIFIED FAILURE		
	MOST-TRUE	TIMES-TRUE	TRUTH-SUMMATION
1	1 (55%)	1 (55%)	* 2
2	2 (65%)	2 (65%)	2 (25%)
3	3 (60%)	3 (60%)	* 2
4	* 3	* 3	* 2
5	* 2	* 2	* 1
6	* 8	* 8	* 8
7	* 4	* 4	* 2
8	8 (55%)	8 (55%)	* 2
SCORE:	4/8	4/8	1/8
TOTAL SCORE:	31/45	29/45	28/45

* = INCORRECT DECISION

Figure 21. Results of applying Larritz' expert system to failures of complex resistor network

DISTRIBUTION LIST

Engineering Psychology Group
Office of Naval Research
Code 442EP
800 N. Quincy St.
Arlington, VA 22217 (3 cys.)

CDR. Paul Girard
Code 250
Office of Naval Research
Code 442EP
800 N. Quincy St.
Arlington, VA 22217 (3 cys.)

Physiology Program
Office of Naval Research
Code 441NP
800 North Quincy Street
Arlington, VA 22217

Manpower, Personnel & Training Programs
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Sciences Division
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Special Assistant for Marine Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR James Offutt, Officer-in-Charge
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375

Naval Training Equipment Center
ATTN: Technical Library
Orlando, FL 32813

Commanding Officer
Naval Health Research Center
San Diego, CA 92152

Dr. Robert Blanchard
Navy Personnel Research and Development Center
Command and Support Systems
San Diego, CA 92152

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 4023
Pacific Missile Test Center
Point Mugu, CA 93042

Dr. W. Moroney
Human Factors Section
Systems Engineering Test Directorate
U.S. Naval Air Test Center
Patuxent River, MD 20670

Dr. Harry Crisp
Code N 51
Combat Systems Department
Naval Surface Weapons Center
Dahlgren, VA 22448

Dr. Edgar M. Johnson
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
U.S. Army Human Engineering Labs
Aberdeen Proving Ground, MD 21005

Director, Organizations and Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Robert G. Smith
Office of the Chief of Naval Operations, OP987H
Personnel Logistics Plans
Washington, D.C. 20350

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Defense Technical Information Center
Cameron Station, Bldg. 5q
Alexandria, VA 22314 (12 copies)

Dr. Clinton Kelly
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Mr. H. Talkington
Engineering & Computer Science
Code 09
Naval Ocean Systems Center
San Diego, CA 92152

Dr. L. Chmura
Naval Research Laboratory
Code 7592
Computer Sciences & Systems
Washington, D.C. 20375

CDR C. Hutchins
Code 55
Naval Postgraduate School
Monterey, CA 93940

CDR Tom Jones
Naval Air Systems Command
Human Factors Programs
NAVAIR 330J
Washington, D.C. 20361

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 61R
Washington, D.C. 20362

U.S. Air Force Office of Scientific Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D.C. 20332

Dr. Daniel Kahneman
University of British Columbia
Department of Psychology
Vancouver, BC V6T 1W5
Canada

Dr. M. D. Montemerlo
Human Factors & Simulation
Technology, RTE-6
NASA HQS
Washington, D.C. 20546

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. T. B. Sheridan
Dept. of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Paul E. Lehner
PAR Technology Corp.
P.O. Box 2005
Reston, VA 22090

Dr. Stanley Deutsch
NAS-National Research Council(COHF)
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Dr. Amos Freedy
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washingotn, D.C. 20064

Dr. Christopher Wickens
Department of Psychology
University of Illinois
Urbana, IL 61801

Larry Olmstead
Naval Surface Weapons Center
NSWC/DL
Code N-32
Dahlgren, VA 22448

Capt. Robert Biersner
Naval Medical R&D Command
Code 44
Naval Medical Center
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Dr. Marvin Cohen
Decision Science Consortium, Inc.
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. John Payne
Graduate School of Business
Administration
Duke University
Durham, NC 27706

Dr. William B. Rouse
School of Industrial and Systems
Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Dr. Richard Pew
BOlt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. Edward R. Jones
Chief, Human Factors Engineering
McDonnell-Douglas Astronautics Co.
St. Louis Division
Box 516
St. Louis, MO 63166

Dr. Lola Lopes
Information Sciences Division
Department of Psychology
University of Wisconsin
Madison, WI 53706

Mr.Joseph G. Wohl
Alphatech, Inc.
3 New England Executive Park
Burlington, MA 01803

Dr. Hillel Einhorn
Graduate School of Business
University of Chicago
1101 E. 58th St.
Chicago, IL 60637