The Internal Model of Complex Dynamic Systems

Christopher D. Wickens
University of Illinois

for
Contracting Officer's Representative
Michael Drillings

Office of Basic Research
Michael Kaplan, Director

June 1991

United States Army
Research Institute for the Behavioral and Social Sciences

distribution is unlimited.
U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

University of Illinois

Technical review by

Nehama Babin

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.
This report examines processing of multiple visual channels of information that pertain to complex dynamic systems. Three classes of results are described: (1) those that focus on the properties of the operator's mental representation or mental model of the system, (2) those that address the optimal means of displaying information about the system, and (3) those that focus on the cognitive biases observed when the mental model is updated by displayed information. Greatest emphasis is placed on the second of these classes, in which the principle of compatibility of proximity for object display integrality is developed.
18. SUBJECT TERMS (Continued)

Object integrity,
Information integration,
Individual differences,
Event-related brain potentials,
Heuristics,
Inferential processing.
The Internal Model of Complex Dynamic Systems
Final Report: Army Research Institute
Contract MDA903-83-K-0255
Christopher D. Wickens

ABSTRACT

This report summarizes the results of ARI Contract MDA903-83-K-0255 examining the processing of multiple visual channels of information pertaining to complex dynamic systems. Three classes of results are described. (1) Those that focus on the properties of the operator's mental representation or mental model of the system. (2) Those that address the optimal means of displaying information about the system. (3) Those that focus on the cognitive biases observed when the mental model is updated by displayed information. Greatest emphasis is placed on the second of these classes in which the principle of compatibility of proximity for object display integrality is developed.

1. OVERVIEW

We summarize below, the results of a 3-year research effort to examine the characteristics of human performance when integrating multiple sources of visual information about complex dynamic systems. These are systems, typical for example of the nuclear reactor, that have several interacting variables, fluctuating over time. The title of the research contract--The Internal Model of Complex Dynamic Systems--suggests that the focus of that research was on the cognitive representation of those systems. As our research progressed, however, it became apparent that there were essentially
two approaches to studying such representations. One approach is direct, through analysis and interpretation of performance and verbal protocols generated by operators working with such systems. Sometimes this approach may be accompanied by the imposition of different training techniques and subsequent examination of the effects of these techniques on comprehension. The other approach is indirect, and proceeds by examining manifestations of performance, inferred to depend upon such a model, as characteristics are manipulated of the display surface or task environment with which the operator is interacting.

Because approaches using verbal protocols encounter certain difficulties (particularly in domains that are heavily spatial and analog—see Wickens & Weingartner, 1986), we chose to employ the second, indirect, approach. However, in making such a choice, our research approach has lead to an inevitable ambiguity between the mental model (the knowledge which the operator has about the operation of a dynamic system), and the mental picture (the particular instantiation of that knowledge, at one point in time as inferred from the momentary state of display parameters). Thus, for example, an operator's mental model of a nuclear reactor will be the same, independent of whether water level is excessively high or low. But the mental picture will be (or should be) dramatically different in two circumstances.

The ambiguity between the mental model and the mental picture is manifest as follows: To the extent that our experimental manipulations vary some characteristic of the display or the task environment in which an operator is performing, and a difference is observed, we cannot be certain that these manipulations have influenced the operators ability to update the mental model (the fidelity of the mental picture), or have influenced the
model itself. Given, however, our assumption that the model within an individual is relatively stable over time, and given furthermore, the suggestion in one of our experiments that the model may be fairly constant between ability groups (Wickens & Weingartner, 1986), we assume that most of our manipulations have in fact focused on characteristics involved in updating the mental picture.

The paradigm that remains invariant across all of our experiments is that of a subject, attempting to update the mental picture of a complex, multivariate system whose state evolves over time. The evolution is governed by certain learnable (and partially predictable) rules which dictate the behavior of individual variables over time (their autocorrelation), and the relation between variables over time (cross correlation). Our interest is in the kinds of biases involved in this updating process, and how these biases may be influenced by environmental, ability and display factors. The following narrative of our research results, which refers to the more in-depth treatments given in the technical reports, is organized by research content rather than by chronological order. Specifically we focus on three themes. (1) Those experiments that have actually addressed internal model differences, (2) the role of display integration in updating the mental picture, and (3) the biases involved in this updating process. Some of the experiments that we describe cut across these three categories. Appendix A of this report presents the abstracts of technical reports for each experiment separately.
2. EXPERIMENTAL RESULTS

2.1 Internal Model Properties

Our initial effort to examine the internal model adopted an individual difference's approach. In Wickens and Weingartner (1985, 1986), we examined subjects' abilities to monitor a dynamic system for potential failures, as a function of whether the subjects were of high or low spatial and verbal ability, as these abilities were assessed by a battery of paper and pencil tests. Comparison was between one group of high spatial-low verbal ability, and a second of low spatial-high verbal ability. We purposefully confounded the manipulation of the two ability differences in order to identify two extreme positions on a verbal-spatial continuum. Although the strengths of both groups were independently validated by performance tests, there were no robust differences between them in system monitoring performance—only nonsignificant trends in three performance measures toward superior performances for the spatial group.

Such results might have been interpreted to suggest that each group used a mental model (or updating process) that capitalized on their ability strengths. However, this interpretation was contradicted by the results when failure detection was performed under dual task conditions. The fact that there was equally high interference for both groups, when the process monitoring task was time-shared with a spatial side task, and equally low interference when time-shared with a verbal side task, suggests that both groups employed the same spatial internal model. Our conclusion from these results was that the model appeared to be task determined more than ability determined.
Two other aspects of our research program have addressed individual differences in spatial ability, both producing generally negative results. Casey and Wickens (1986) failed to find that these differences discriminated good from poor performance on a process monitoring task, described in section 2.2 of this report. A second small study, not written into a report, by Jones and Wickens failed to find that high and low spatial ability groups discriminated performers on a test of intuitive integration of diagnostic numerical information regarding the status of a hypothetical energy system.

In Wickens and Weingartner's (1985, 1986) study, both groups received the same mixture of verbal and spatial (graphical) training, and so the knowledge imparted about the system was not expected to form a source of variance. In contrast, in a study just completed by Zhang and Wickens, we have explicitly manipulated training factors in order to establish the importance of semantic context in understanding the relation between dynamic system variables. Employing an eight variable dynamic gas-mixture system that was in many respects similar to that used by Wickens & Weingartner, we contrasted system monitoring performance between a group of subjects who received pretraining on the semantic context of the task, (i.e., explanation of the meaning of the variable interrelationships in terms of their gas properties of flow, pressure, temperature and so forth) with a group that only learned of the variables as abstractions. Pretraining on the semantic context provided no benefits to system monitoring performance. In contrast however, when the semantic context was provided concurrently with monitoring performance, and was represented by a static flow diagram of the system next to the variable indicators, this led to better failure detection performance than did a context-free display.
One final mini-experiment was carried out within the mental model context, and was designed to examine if subjects' representation of the continuously changing quantities in the monitored process used by Zhang and Wickens, was continuous and analog, or discrete. To do so we implemented two different versions of the display. One version represented each quantity as a continuously changing bargraph, and the second, as one of six discrete pieces of information each indicated by arrows located at one of three heights (indicating the quantity's value), and pointing either upward or downward (indicating the direction of change). Therefore the discrete display representation was highly compatible with the continuous information, indicating location and direction of change. Furthermore, analysis of the information content of this discrete display showed that it was just as diagnostic of failures as was the continuous display. However, the results indicated an overwhelming superiority of the continuous display in supporting failure diagnosis. In fact the experiment was discontinued after only a few subjects were run, because it was felt that additional data collected would be unnecessary and redundant in supporting this superiority. Therefore we concluded from this study that subjects do rely heavily upon the continuous changes in variable magnitude, even when it is not essential that they do so (since sufficient information to detect and diagnose failures was conveyed by the discrete display).

2.2 Display Proximity and the Mental Picture

This series of studies focused on the following research question: Given that information along a set of channels must be integrated (a characteristic of monitoring most dynamic systems), should the display of those channels also be integrated, or physically "proximate"? An
affirmative answer to this question defines the proximity compatibility principle (Wickens, 1986). A corollary of this principle is that, given that information along separate channels should be treated independently, the displayed channels should not be proximate. Independent processing results from the task requirement to focus attention on a smaller subset of variables while filtering others; or to process these variables in parallel as part of a multi-task requirement. In most of these studies, display proximity has been defined by the representation of the dimensions of a single object (Casey, Kramer, & Wickens, 1984; Casey & Wickens, 1986; Casey, 1986; Jones & Wickens, 1986; Barnett & Wickens, 1986; Barnett, Goettl, Kramer, & Wickens, 1986; Goettl, Kramer, & Wickens, 1986). This object representation was typically compared with a representation of the variables as separate bargraphs. In Wickens (1986) I present the collective results of these (and other relevant) studies in a single integrative framework, which provides the theoretical background for the proposed compatibility of proximity principle. Presented below is a brief outline of the relevant studies, with more detail provided in Wickens (1986), and in the individual reports.

Our initial study (Casey & Wickens, 1986; Casey, 1986); compared three display formats for representing a simulated temperature process monitoring task. The subjects monitored changes in the temperature of five centrally heated chambers to detect abnormalities in the heat flow to a given chamber. Information was conveyed by separate bargraphs, an integrated pentagon display, or a schematic face display. (The perceptual salience of changes in each of these displays had previously been equated in a psychophysical calibration study reported in Casey, Kramer & Wickens, 1984). The results obtained by Casey & Wickens failed to provide evidence of benefits for the
face display in the process monitoring task. They suggested that the separate bargraph display was best, but that its advantage over the pentagon was less in detection (indicating that any variable had failed) than in diagnosis (identifying which variable had failed).

This difference between detection and diagnosis we attributed to differences in the processing characteristics of the two tasks—a greater degree of integration required in the former; a greater degree of attention focusing in the latter. To address this hypothesis further, our next study (Jones & Wickens, 1986), compared the bargraph and pentagon display in a task that required the maximum degree of integration—averaging of the five dynamically changing values. In contrast to the results of Casey and Wickens' study, there was here a clear unambiguous advantage to the object display when full integration was required.

Both experiments also incorporated manipulations of cross correlation between variables—a manipulation of the "configural" properties of the set of variables. In both, the effects of high correlation were either to leave the performance of the object display relative to that of the separate display unchanged (Casey & Wickens), or reduced (Jones & Wickens). Such a trend—indicating that the object display appears to help most (or hurt least) those conditions with a lower correlation between variables—appears to be consistent with other findings in the literature (Wickens, 1986).

A third investigation (Casey & Wickens, in preparation) has tried to resolve and clarify more explicitly the differences between Casey and Wickens (1986) results, and others which have shown more consistent object display advantages (Carswell & Wickens, in press; Jones & Wickens, 1986). A fundamental difference between these sets of experiments, which is consistent with the proximity compatibility principle has been the mapping
of display variables to failure states. In Casey and Wickens (1986) this mapping was one-to-one. In the other studies, it was many-to-one. A further difference is that the failures to be in detected Casey and Wickens' (1986) study involved changes in the correlational structure between variables. In Carswell and Wickens, these failures had involved changes in the causal structure. Therefore both the structure (correlational versus causal) and the failure mapping (many-1 versus 1-1) were varied within this third experiment. Unfortunately however, the results remain ambivalent. Like the earlier study of Casey and Wickens, an advantage to the bargraph display was found, and did not appear to be modulated by changes in mapping or structure. The issue here remains unresolved.

Two further experiments have examined object display integration in tasks that are more typical of a decision domain than a process monitoring domain. In Barnett and Wickens (1986) (reported in Barnett, Goettl, Kramer, & Wickens 1986), the "system" that was monitored was the subject's internal representation of the advisability of carrying out a course of action (to either continue or abort a simulated flight mission). This representation was updated on the basis of a series of discrete diagnostic cues. The cues varied in their diagnosticity (D) and reliability (R)—dimensions which formally combine multiplicatively to reveal the cue's total information worth. The data indicated that the ability to integrate D and R information from a given cue was facilitated by a rectangle (object) display, over a bargraph display, and the ability to integrate across cues was also facilitated by proximity in both space and time. When task demands called for focusing attention on one of the attributes (D or R), the advantage of object integrality evaporated. Thus both the compatibility of proximity principle and its corollary were confirmed.
Goettl, Kramer, and Wickens (1986) (in Barnett, Goettl, Kramer, & Wickens 1986) also employed a judgment task, typical of that employed in multicue probability learning. Subjects viewed displays of three cue values, represented either as bargraphs or as apexes of a triangle, and attempted to infer their predictive relation to a criterion value. Feedback was provided after each response. A between condition manipulation of attention focusing was incorporated to test both the principle and its corollary. During one block of trials, all three cues were relevant, requiring full integration. During a second block, one cue was made irrelevant, requiring subjects to focus attention on a smaller subset. The results indicated that during the first block, there was no difference between the object and bargraph displays. During the second block, requiring attention focusing, an advantage for the separate display emerged. The results then support the proximity compatibility corollary.

Collectively then, and in conjunction with other studies, discussed in Wickens (1986), these experiments provided general support for the proximity compatibility hypothesis and its corollary. As more integration is required, the relative advantage of proximate displays increases; or their relative cost decreases.

2.3 Biases and Heuristics in Information Integration

The thrust of these efforts was to identify manifestations of different decision making heuristics, well documented in structured static problems employed by investigators such as Tversky, Einhorn, and their colleagues as these biases might be manifest in the more dynamic process monitoring domain. In Jones and Wickens' (1986) study for example, we examined for the presence of the anchoring and adjustment heuristic, asking if subjects'
estimate of the state of the process was revised on the basis of new evidence, as sufficiently as was optimal. The results indicated that there was a general trend, across all conditions, away from a conservative anchoring strategy; that is, subjects tended to adjust more rather than less than the optimal amount. This heuristic, along with two others was examined in more detail by Barnett and Wickens (1986b). In this experiment the scenario was similar to that employed by Barnett and Wickens (1986a) and described in Section 2.2: a sequential decision making task in a military flight scenario, in which the advisability of carrying out or aborting an airborne mission was repeatedly updated on the basis of a series of discrete cues. The results of this experiment, in contrast to that of Jones and Wickens did reveal a conservative anchoring tendency in hypothesis revision. The source of this discrepancy between the two studies is not immediately apparent, although it may be inherent in the greater degree of autocorrelation within the cues used in the Barnett and Wickens (1986b) study. Furthermore Barnett & Wickens required the integration of more disparate sources of information. In neither study however were the biases particularly strong.

The primary focus of Barnett and Wickens' study was to examine the interactions between the display locations of the cues, influencing their perceptual salience, the simulated source of the cues, indexing their psychological salience, and their information value, defining their optimum weighting. It was hypothesized that imposing various forms of cognitive stress (time stress and dual task loading), might increase the manifestation of two heuristics—a bias toward more salient display locations, and a bias to subjectively equate the information value of the different cues. While subjects in general performed quite optimally, the two forms of stress were
found to induce differential tendencies toward the two heuristics; time stress slightly increased a left position bias (perceptual salience); task loading produced an overall reduction in decision quality and a slight (but not significant) cognitive leveling over different cue information values.

This issue of cognitive leveling of information value was one explicit focus of a set of experiments reported in Wickens, Bosco, Kramer, Mane, Coles, and Donchin (1986). In these experiments, which adopted a process monitoring paradigm, subjects were asked to update their hypotheses of the momentary state of a time-varying process, on the basis of discrete cues of differing information value. The three experiments varied in their level of complexity, and in the degree of fidelity to a real process monitoring environment in which indicators of different information worth are physically located at different display locations. Across the three experiments, we found that as realism and system complexity increased, a progressively greater (non-optimal) tendency to treat all cues as if they were of the same information worth was observed.

A second thrust of this experiment was to assess whether the P300 component of the event related brain potential was sensitive to the differential extraction of information from these cues. That is, whether larger P300's were produced by processing cues of greater information value. Of course, such an assessment was dependent upon whether our performance measures in fact revealed that information was differentially extracted. In the simplier conditions, in which this was the case, we did indeed find that extraction of more information was accompanied by a larger P300. This finding at once helps to validate a theory of P300 amplitude, and to suggest its utility as a tool in display design.
3. APPLIED IMPLICATIONS TO ARMY NEEDS

We believe that the research reported above suggests a number of important general contributions to cognitive psychology and information processing theory. From this research however we are also able to identify five specific implications with more direct applications to the Army's interests. These are outlined briefly as follows:

3.1 Individual Differences and Multiple Task Performance

Wickens and Weingartner (1985, 1986) found that paper and pencil ability measures provided valid predictors of single task performance, but failed to predict the loss of that performance when resources were diverted to a concurrent task, a result recently replicated by Derrick, McCloy et al (1986) and consistent with interpretation of these results in the literature, recently reviewed at a conference on standardization of performance testing, held at Vaalsbroek, Netherlands. Such results point to the continuing need to develop adequate predictors of time-sharing performance.

3.2 Multiple Resource Theory and Task Prediction

The results of Wickens and Weingartner's study also provide an important data point for our efforts to develop the Multiple Resource Model of Human Performance into a predictive model, of use to systems designers, that will predict when and how task components will interact (Wickens, 1984). Code interference (i.e., interference between two spatial tasks) has been frequently demonstrated in laboratory situations using simple tasks; but in more complex operational settings, these demonstrations have often confounded the amount of visual (scanning) interference with the difference
between verbal and spatial tasks. The current data clearly substantiate the finding of code interference, in a paradigm in which scanning is absent, and in a task with some degree of operational complexity.

3.3 The Object Display

As noted in Section 2.2, the research program focused heavily on the concept of the object display. The potential efficiency of the object as a means of reducing information processing load in information integration tasks was demonstrated. Also demonstrated were some of the costs associated with its use, when focused attention or independent processing is required. Together these demonstrations provide the basis for a theory-based set of design guidelines or principles, which should aid the formatting of complex multi-element displays.

3.4 Stress Effects and Heuristics

These effects and biases demonstrated in our experiments reported in Section 2.3 were relatively small. This finding itself may be important, as indicating the constraints on non-optimal behavior that are shown as characteristics of real world situations are imposed (specifically cross and auto correlation). Equally important has been the demonstration of the feasibility of making measurements of these heuristics in the paradigms described. Given this feasibility, it is hoped that they may be examined further in more operational settings.

3.5 The ERP as a Tool for Display Evaluation

Finally, we have demonstrated in Wickens, Bosco, Mane, Kramer, Coles, and Donchin (1986) that the P300 component of the event related brain
potential can provide an index of the information that is extracted from visual displays. The data however, suggest that our ability to discriminate different amounts of information extraction, is not great. However our ability to discriminate attended for unattended stimuli in one display is strong, and reinforces the potential value of the EPP as a tool in display evaluation.
4. REFERENCES

(This list includes only those references not supported by the ARI contract. The latter are reported in Section 5 below.)


5. ARI PUBLICATIONS & TECHNICAL REPORTS

Publications


TECHNICAL REPORTS


