Human Factors Psychology in the Support of Forecasting: 
The Design of Advanced Meteorological Workstations

ROBERT R. HOFFMAN

Adelphi University, Department of Psychology, Garden City, New York

(Manuscript received 12 September 1990, in final form 12 September 1990)

ABSTRACT

Advanced Meteorological Processing (AMP) systems will provide a workstation environment to support the activities of operational forecasters and research meteorologists. AMP system research and development projects are underway at laboratories of the National Weather Service, the U.S. Air Force, and a number of universities and private corporations. AMP systems will integrate artificial intelligence techniques with forecasting procedures, and will support the interpretation and integration of data from new remote sensing satellites and new ground-based radars. This article illustrates human factors research aimed at generating specifications for prototype AMP systems. A task analysis of forecasting deliberations and structured interviews with research meteorologists lead to a number of recommendations about the design of AMP systems, and to some ideas about needed research on graphics displays and the reasoning of expert meteorologists.

1. Introduction

Advances in remote sensing technology have yielded such a volume of data that meteorologists are sometimes inundated (cf., Chaters and Suomi 1975). Buoys, weather stations, and ships alone generate about 17000 sets of reports every day. The NEXRAD (NEXT generation weather RADar) demonstration project (Forsyth et al. 1985), yielded a number of different types of display products (e.g., time-height cross-section of wind velocities between 0200 EST and 0201 EST), totalling 450 h of display products over 59 days, or about 8.3 h of products per day. Operational NEXRAD installations will have the potential of providing any of 140 different display products every 5 min. Meteorological forecasting (and research) activities do not involve inspecting all of the available data, but only data that are pertinent to the forecasting problem of the day. The forecaster needs to have a clear idea of what to look for, where to look, and how to look (i.e., to know what questions to ask).

Along with the introduction of new data types and products has come an increased reliance on information processing systems. Advanced Meteorological Processing (AMP) systems are intended to provide a workstation environment that can support a number of related activities, such as data integration, single-station forecasting, and meteorological research. In addition, AMP systems will involve the analysis of weather data by Artificial Intelligence (AI) subsystems.

A number of Federal government laboratories, private companies, and university laboratories are currently developing AMP system prototypes (Dyer 1987; Seguin 1989).

The introduction of new data products along with new and complex information processing systems has meant that meteorologists have had to spend more time learning about and communicating with information processing systems, as opposed to doing actual forecasting. "The computer is driving the man rather than vice versa" (Boehm 1979). On the basis of such comments, one might wonder if the situation in meteorology is approximating the "worst case" scenario in terms of human factors. The factors that are known to increase workload in workstation environments are: 1) The display of too much information, 2) The display of irrelevant information, 3) The display of important information in the wrong places, 4) The display of information in an inflexible way, and 5) The display of information that lacks clear meaningfulness (Munk 1985). Such a situation can arise if it is assumed that human factors considerations are "common sense" and are taken care of in the ordinary course of design and prototyping (Kemeny 1979; Landauer 1987).

An example comes from the report on the National Weather Service's (NWS) Automation of Field Operations and Services (AFOS) project (Giraytys 1975), a prototype meteorological information processing workstation. After building the prototype, it was given a trial run, and then the participating forecasters were given a questionnaire. Some aspects of system design had been examined from a human factors perspective, but the forecaster's work procedures had not been

Corresponding author address: Robert R. Hoffman, Department of Psychology, Adelphi University, Garden City, NY 11530.

© 1991 American Meteorological Society
studied extensively. What the designers thought the meteorologists needed was not always what the meteorologists actually needed or wanted. AFOS cathode ray tube (CRT) displays were small, the use of color was restricted, and there was no animation (or "looping") capability. The interface used confusing command acronyms, and required users to type in many lines of commands in order to operate the system. Hence, several design changes were recommended (Seguin 1989).

Any AMP system development project will necessitate a great deal of research, based on an overall guiding philosophy and development plan, according to which the structurally separate components are designed in a functionally integrated way. A well-designed AMP system should support the interaction of machine and human as a complementary pair, a single system that capitalizes on the capabilities of each component and optimizes each component's ability to accomplish tasks (Hollnagel and Woods 1983).

The human factors approach to design research and development includes steps to be taken prior to prototyping (Sanders and McCormick 1987). One such step is to consult the relevant literature on human factors. That literature contains many design recommendations, some of which are very specific (e.g., the refresh rate of the cathode ray tubes). However, many take the form of general design principles. Any given design recommendation will not necessarily apply to a particular domain, and if it does apply, it may require modification and refinement. It is axiomatic in human factors psychology that recommendations from past research need to be adapted and modified in light of research on the specific domain at hand. Thus, another step in the human factors approach involves "bridging the gulf" from available relevant design principles (Gomez and Dumais 1986).

The research described here illustrates the generation and refinement of human factors guidance, with a focus on the design of AMP systems. The first question addressed by this research is, What do meteorologists do? That is, what are their goals, what tasks do they perform, and what are their reasoning processes? To answer this question, a "task analysis" of forecasting was conducted. The results suggest a description of the basic cognitive processes and reasoning strategies of forecasters—processes and strategies that will have to be supported by an AMP system. The second question addressed by the present research is, What do meteorologists need? That is, what kinds of displays and information processing capabilities would best support their activities? An answer is sought through "structured interviews" with experienced meteorologists.

Following a discussion of the methods and results of the two studies, examples are provided of design recommendations, focusing on the AMP information processing capabilities, the graphics displays, and the workstation layout.

2. Study 1: Task analysis of forecasting

A principle that is axiomatic in human factors psychology is that system design should be based on a functional analysis of tasks in terms of the specific activities, goals, and needs of the system user (Buck 1983). Some previous research indirectly examined forecasters' tasks and goals by means of survey questionnaires, with the goal of revealing issues concerning precipitation probability forecasts (Murphy and Winkler 1974a). The present research utilized a more direct method for examining forecasters' tasks.

a. Method

Observations were made of 14 summertime bi-weekly mesoscale forecasting deliberations held at the Atmospheric Prediction Branch of the Air Force Geophysics Laboratory (AFGL).

1) PARTICIPANTS

Each deliberation was attended by four to ten research meteorologists, and had a designated leader, determined by a rotating schedule. Three of the meteorologists had experience in designing prototype AMP systems. Four had experience in operational forecasting for either civilian or military weather services.

2) THE CHARTROOM

The deliberations were held in a relatively small meteorological chartroom. (Operational meteorological chartrooms also typically contain telephones, teletypes, and other communications equipment not needed at the AFGL.) Typical of such chartrooms, it was dominated by meteorological charts showing surface observations, observations at various geopotential heights in the atmosphere, and the forecasts made by various models. In all, there were 28 different types, posted on two rows of clipboards spanning two entire walls. The charts' nature and arrangement followed meteorological traditions. (e.g., for the forecast products, time was represented across clipboards, with each having a different valid time). Also posted were photographic reproductions of half-hourly GOES visible and infra-red images. Against a third wall was a graphics CRT display of on-line ground-based regional radar data and lightning network data, and a CRT display of on-line regional observations in tabular form.

3) THE DELIBERATIONS

The observed deliberations were formal in that there was a year-round systematic comparative assessment of the predictions of various weather events (e.g., high and low temperatures, precipitation, etc.). However, the deliberations were informal in two respects. First, leaders were free to express their own forecasting styles. Second, the attending meteorologists were free to make
comments or ask questions of the leader. Occasionally, there were debates about the interpretation of data or about the peculiarities of various mathematical forecasting models.

The deliberations were held in the early afternoon, but the leaders prepared for their deliberation, in their usual manner, by analyzing the available data beginning in the early morning. Most leaders used colored felt-tip pens to indicate various fronts, high and low pressure areas, and other weather features directly on particular paper charts.

4) Procedure

The procedure was quite simple. During each deliberation, the observer kept note of the time, recorded each explanatory comment, and recorded the order in which the leader referred to any of the available data products (i.e., charts and satellite images).

b. Results and discussion: The cognition of forecasters

The deliberations were all fairly brief, having a mean duration of 16 min (range of 10 to 21 min). The average number of displays or charts referred to (including repeated references to some charts and displays) was 41 (range of 31 to 60). This works out to about 2.6 charts/displays per min. In the “gestural acts”, in which a forecaster referred to charts or displays and simultaneously made interpretive or explanatory comments, the forecasters would generally point to between one and three (sometimes four) charts or displays. This result has implications for the layout of AMP workstations, which are considered in more detail in the section entitled “Design Recommendations.”

Invariably, the leaders began their deliberations by citing the charts and satellite photos that depicted the current weather situation. This initial assessment would suggest some preliminary ideas about the nature of the mesoscale and synoptic-scale processes that were at work. Once an initial “mental model” was formed, the next step involved referring to specialized charts and displays, seeking evidence to either support or refute hypotheses that stemmed from their initial conceptual model. Which particular data were inspected was therefore a function of the weather scenario and the hypotheses that were being considered, or the forecast problems of the day. If observations suggested that microbursts might occur, a forecaster would inspect real-time images of high-resolution radar data. On the other hand, if a cold front was moving through, then hourly images of clouds at visible wavelengths would be inspected. One way or another, the forecaster’s initial mental model was tested and refined.1

---

1 Naturally, the observed deliberations invoked phenomena appropriate to the region and the time of year, such as severe storms. Indeed, the deliberations often focused on the weekend weather, with summer activities in mind.
mulate a new one or face the uncomfortable prospect of having no idea of what to expect next” (p. 255).

In considering the meteorologist's formation and testing of mental models, there is a potential for confusion given the various uses of the word “model” and the various meanings of the word “dynamic.” Mental models are themselves dynamic in the sense that all thoughts and mental images are dynamic (as opposed to being static like snapshots). Certainly, the dynamics of mental models are not the same on a material level as the atmospheric dynamics which the models represent. (This is part of what it means for something to be “representational.”) Forecasters’ mental models depict the kinematics of atmospheric events in a way that is consistent with the causal dynamics of the atmosphere (Godske et al. 1957, Ch. 17). Their mental models have conceptual or explanatory value in that they rely on the understanding and application of causal principles of physics and meteorology.

The basic finding of this study dovetails with a general principle of human factors psychology: A well-designed AMP system needs to support the user’s needs and capabilities. Specifically, it needs to support their formation and testing of mental models. This basic principle was used as guidance in selecting and refining AMP system design recommendations.

3. Study 2: Structured interviews for AMP system design

As mentioned previously, many of the available human factors principles fall at a general or domain-independent level. Bridging the gulf to specific applications can be one of the more challenging aspects of system development. A few studies using questionnaire techniques have empirically examined meteorologists' preferences in terms of display products and display design (e.g., Giraytys 1975; Wilkins and Johnson 1975). The present study utilized structured interviews. A structured interview is preplanned, like a questionnaire, but does not force the respondent to use predetermined response categories. (For a detailed discussion of interview methodology, see Brule and Blount 1989.)

The structured interviews had four purposes: 1) To determine whether each of a number of candidate specifications would be judged reasonable by experts, 2) to generate refinements of the specifications, 3) to generate ideas about additional specifications, and 4) to generate ideas about how to make the prototype system more flexible and more acceptable to the users.

a. Method

1) PREPARATION OF THE MATERIALS

The relevant human factors literature presents research on workstation design, the design of interfaces, and other topics. Sources include texts on human factors and workstation design (cited below), but the most helpful sources were Kelly and Sauer's (1986) report on the design of the Air Force Advanced Integrated Workstation; the Technology Systems, Inc. (1984) report on the specifications for the University of Wisconsin-AFGL Meteorological Data Acquisition, Management, and Processing System (McIDAS); and the Department of Defense (1985) guidelines for information processing systems. Also especially helpful were two reports on prototype AMP systems (Chisholm et al. 1983; Forsyth et al. 1985), in which forecasters provided feedback about display and analysis products presented in a workstation-like environment.

Extracted from the literature were design recommendations which seemed pertinent to AMP systems, their displays, their operational modes, their interfaces, and the workstation layout. The literature search yielded 148 propositions, that is, single-sentence statements of general design principles and (more-or-less specific) design recommendations.

The propositions were entered into a computer file according to meaningful categories, such as display design, interface, workstation layout, etc. A printout of the file was used to provide structure to the interviews.

2) PARTICIPANTS

Ten AFGL meteorologists were interviewed. All were engaged in research at the time, although three had previous experience in operational forecasting. One was a specialist in NEXRAD. Three had extensive experience with meteorological information processing systems (i.e., hardware integration and software development).

3) PROCEDURE

In the structured interview, the interviewee went through the set of propositions one at a time, commenting on each one. The interviewer recorded every comment in the form of marginal notes on his copy of the printout. A comment could involve agreeing or disagreeing with a proposition; it could involve elab-
rating upon a proposition; it was also possible for a given proposition to be passed over—"No comment."

Each interview lasted between two and three hours, including breaks.

b. Results

Table 1 presents a summary of the results according to response type. The total number of agreement responses, disagreement responses, elaborative comments, and "no comment" responses was 1,864. Of that number, about 92 percent were agreement responses. The interviews also produced a number of refinements of the candidate specifications. The 10 interviewees yielded, on the average, 135 agreement responses, 39 elaborative comments, and only 6 disagreement responses. For some of the propositions, some interviewees had no comment. For example, operational forecasters might have some comments about the code-level specifications for the information processing system, but they might have a great deal to say about interface issues.

In general, agreement with the candidate specifications was very strong. Some of the specifications were of the "obvious" kind, but many of these were nonetheless regarded as important. Indeed, the most frequent responses were: "Yes, I agree," "Absolutely essential," and "Yes, definitely." Some of the specifications were not so obvious, and another frequent type of response was "Yes, that is often overlooked," and "Yes, good idea." The overall high rate of agreement suggests that the initial set of propositions formed a reasonable starting point.

4. Design recommendations

The upshot of these two studies is a set of candidate design recommendations for prototype AMP systems, some of which will be presented here for illustrative purposes. Of the examples presented here, some came initially from the references cited. If no reference is given for a particular recommendation, that indicates that the recommendation originated in the structured interviews. Although there should be some caution about the generality of the results due to the relatively small sample and the research orientation of many of the interviewees, all of the recommendations have been through some refinement, to the point where most of them seem to represent a consensus of forecasters and researchers. Many of these recommendations are manifested, for instance, in the AFOS system (NWS 1985) and in PROFS (Program for Regional Observing and Forecasting Service), a prototype forecasting system of NOAA (Schlatter 1985).

---

2 The full set of recommendations is available upon request, and includes more details on workstation layout, recommendations about the composition of the project team, and recommendations about the computer hardware that will be needed to do all the things that an AMP system needs to do.

---

a. Information processing subsystems

AMP systems will require advanced information processing (IP) subsystems for communication, processing, and storage of weather and climatological data, and also for the application of mathematical forecasting analyses. Examples of the design recommendations for AMP information processing subsystems are presented in Table 2.

Some of the recommendations in Table 2 refer to the "user friendliness" of the command-level interface (i.e., prompts, menus, commands, etc.). Rather than relying on complex, cryptic command codes and acronyms, the interface should facilitate the user's comprehension of system operations. Thus, human-machine interactions should approximate a naturalistic dialogue (Hollan 1984). There is some controversy over the needed degree of approximation, but in any event, the user must be able to "navigate" easily. For example, if hierarchies of operations or menus are complex (i.e., many layers deep), then users may end up having to rely heavily on commands such as "Undo." If so, the hierarchy is probably in need of greater breadth. To permit easy navigation, the IP subsystems will have to be sensitive to the differences between novices and those who are intimately familiar with system operations. For instance, graphic icons may help novices, but experienced users may want to skip over certain operations or menu levels that novices need to work through (Chapinis et al. 1983; Graesser et al. 1987).

b. Display design

Psychologists have conducted a great deal of research on the question of how people perceive graphic symbols (e.g., Farrell 1984; Thorndyke and Goldin 1983). In this research, participants are presented with arrays or map-like displays of geometric or cartographic symbols, in which symbol type, size, color, and other variables are systematically manipulated. The typical task for the participants is to search the display for particular targets, while the experimenter measures such variables as search reaction time, accuracy, and the effects of learning over trials. This research has implications for the design of graphics displays of all types. For example, it is well-established that the perception of symbols is strongly affected by visual discriminability (Remington and Williams 1986). Poor discriminability is manifested in meteorology, for example, when it is easier to perceive individual numerical values on a paper chart than off of a CRT display.

At the most general design level, a good display should be as simple as possible given what it has to do. It should present data clearly and unambiguously, it should take a minimum amount of time to interpret, and it should require minimal display-reading skills (granted that "minimal" may require considerable training). Some example recommendations for AMP
Another important problem involving the use of color is that of accommodating meteorologists who have a color vision weakness. Most color vision problems involve a weakness in the perception of reds or greens (which appear yellowish-grey to the color weak person). Color weak individuals can discriminate colors, but often do so on the basis of brightness or degree of apparent grey tonality. Psychological research on color discrimination suggests that color displays for such individuals should involve either blue or purple, but not both. They should involve either red or yellow, but not both. They should use a highly saturated yellow—if not highly saturated, yellows can appear grey to color weak individuals (Miller-Jacobs 1984).

When colors are presented as separate patches, individuals with normal vision can readily discriminate about two dozen colors (with hue, brightness, and saturation all varying). However, problems arise when colors are applied to actual data sets. Cartographic and meteorological displays contain various convoluted shapes, and not isolated patches of color. Indeed, a significant problem in all domains of cartography is that there are more things to be represented than there are discriminable colors (Hoffman and Conway 1989; Tufte 1983).

For example, one of the NEXRAD wind products (Forsyth et al. 1985; Figs. 7 and 13), shows arrays of wind bars in which various colors indicate various wind speeds. For example, green bars indicate 20-kt winds, purple indicates 30-kt winds, and a pale red (pink) indicates 40-kt winds. In theory, the green would be clearly discernable from the purple, which in turn would be clearly discernable from the pink. However, a particular set of data happened to contain dense clusters of bars which were mostly purple or pink. As one could imagine, discriminability was low even though the displays were high resolution. (Also see Fig. 17 in Schlatter 1985).

It is important that there be a noticeable difference (technically, a “supra-threshold chromatic contrast”) between each of the colors, and between colors and their backgrounds (De Corte 1986). Common experience among scientists who use colored displays is that discriminability can begin to break down if more than about seven colors are used (McCormick 1976). Typical false-color remote sensing displays use anywhere from 5 up to 35 different colors (Hoffman and Conway 1989). The typical basic color palette consists of very bright and highly saturated tones—such as a set consisting of blue, green, yellow, red, and purple (or violet) (along with black, white, and grey tones). Color sets that go beyond the basic set are likely to include orange, or perhaps a second shade of blue (or a blue-grey), a
second shade of green (such as a yellow-green), or violet (in addition to purple). Note the potential discriminability problems with an expanded set: Red and violet are sometimes hard to discriminate, as can be yellow and orange, as can be yellow and yellow-green (Miller–Jacobs 1984).

For some applications, the number of colors must be reduced rather than expanded (e.g., Krebs et al. 1978). For example, a display that involves overlays (i.e., three spatial dimensions) and animation (i.e., the time dimension) will probably have to be restricted to a few colors because of the great processing load placed on the graphics computer.

d. Workstation design

The activities of forecasters and researchers differ considerably, even though they rely on the same workspace components—desk, shelf space, charts, CRTs, and keyboards. It was generally acknowledged in the structured interviews that the workspace for an individual researcher can be a cubicle-sized station. With regard to the main workstation, one of the first issues to be addressed is the question of how many CRT displays the main workstation should have.

“Each day, the 100 or so facsimile charts and the output of teletypewriter circuits are hung on walls or posted on clipboards. A trained meteorologist can quickly scan this information and form impressions of large-scale weather events. He has great difficulty, however, in sorting out details or rapidly finding that one item needed . . . In terms of AFOS, does this mean that we need to provide as many graphic displays as chart panels that the forecasters can now see in one glance? This might require upwards of twenty displays” (Giraytys 1975, p. 112).

The question of “how many displays” is linked to the more fundamental question of how many different kinds of data the meteorologist needs to see. Operational AMP workstations will involve a compromise between two extremes: Having a single CRT display which relies heavily on an overlay capability, vs. having a separate display for each of the things to be displayed. In an ideal world, cost factors would not matter. However, high-quality color graphics CRTs (and their associated hardware and software) are quite costly. No solution to this problem jumps out of a ready-made formula in the human factors literature. However, a tentative empirical solution is suggested in the results from the task analysis described earlier. It was found that the forecasters generally referred to between one and three charts/displays in each of their gestural or explanatory acts. It thus seems likely that AMP workstations will have at least two CRT for data display (in addition to a display for system operations). An early generation of the PROFS workstation had two CRTs (one for system operations, one for data displays), but forecasters who worked with it suggested that future generations preserve the overlay capability and yet include at least one more CRT for data display (Schlatter et al. 1985). The AFOS meteorological workstation (Giraytys 1975; NWS 1985) had three data displays.

The availability of more than two graphics CRTs would give the user control over which information to display, when to display it, where to display it, and how to display it. Meteorologists could customize the location of information and its flow to fit their needs and operating style. In other words, the CRTs could be used in a distinct “operational mode” for each of the functionally distinct activities that is revealed by task analysis.

In the observed forecasting deliberations, the forecasters invariably began by referring to the annotated chart of surface observations and the latest GOES satellite imagery. An “electronic chartroom” mode is conceivable in which the workstation CRTs would display surface observations, GOES imagery, and atmospheric data from various geopotential heights (e.g., the jet stream). During regular forecasting operations, one CRT could continuously loop a series of satellite images, another could loop radar reflectivity, and the other two CRTs could be available for inspection of other products that may be of use (e.g., Doppler velocity, upper air plots, etc.).

A “scenario building” mode would operate as an electronic felt-tip pen (Chisholm et al. 1983; Doswell and Maddox 1986; Forsyth et al. 1985). It would allow the user to call up various data sets or displays (radar, GOES, symbolic data) and overlay them. Alternatively, the user could tag the salient weather features (i.e., fronts, regions of precipitation, regions of high cloud

---

**Table 3. Example of general design principles and design specifications for AMP system graphics.**

1. The graphics component of an AMP system will have to display a number of traditional data types: point observations (rawinsonde, radiosonde, surface reports, ship reports), imagery (GOES satellite images, radar, lightning networks), products (e.g., charts from the National Weather Service and the Air Weather Service), and the outputs of various mathematical models.

2. Each graphics CRT should be able to display each of the available images and data sets. Each CRT should be able to display more than one product or data set, using either a sectoring or an overlay capability.

3. AMP displays should have as an option the presentation of data in the form of function graphs rather than data tables. This is recommended because of the ease with which the visual system perceives relationships such as position, length, angle, and area (Cleveland and McGill 1985; Poulton 1985). If function graphs are to depict rates of change or differences between functions, there should be an option of depicting the slopes or differences themselves.

4. Displayed elements (i.e., lines, symbols) should be large enough to facilitate the basic perceptual operations of detection, discrimination, and identification (given the size and spacing that is appropriate to the scale of the displayed information) (McCormick 1976). Displays should minimize the visual flicker and ragged edges that are caused by digitization.
tops, etc.), and then "drag" them onto a central CRT—using the overlay capability to build up a three-dimensional representation of the scenario being evaluated (such as the structure of storms). In effect, the user could construct a scientifically valid graphical depiction (involving three spatial dimensions and one temporal dimension) of their mental model.

In the refinement of their mental models, meteorologists usually rely on the outputs of mathematical models. In a "mathematical model" mode, a dedicated function key would animate the displays (i.e., time series looping). Model-based predictions or series of satellite or radar images would be displayed successively across various time intervals.

Given the above considerations, one can begin to conceive of possible layouts for the main workstation itself. Human-factored workstations that have more than one CRT (Kelly and Sauer 1986; Monk 1985) place the CRTs in a wrap-around parabola in which the CRTs fall slightly below the horizontal line of sight of the user. At the same level are such things as clocks, bulletin board, and communications equipment. On a broad shelf at waist level are keyboards, graphics pad, and desk space. The specific dimensions for the parabola, desk space, etc. can be derived from known rules of anthropometry (Davis and Swezey 1983; Kantowitz and Sorkin 1983; Morgan et al. 1963, NASA 1978; Van Cott and Kinkade 1972).

The AMP workstation, especially its prototype, will have to be built upon adjustable height-angle CRT mounts, so that the workstation can be made to fit particular facilities or particular operational needs. A single seated user or small groups of users would need a narrow workstation radius. However, a workstation for large forecasting operations or for briefings might need a wide radius. In addition, a traditional chart wall might adjoin the workstation. Alternatively, a chartroom might be separate from the workstation room, to support deliberations without interfering with workstation operations. Workstation design involves many details that flesh out the workstation layout (cf., Eastman Kodak Company 1983; Human Factors Society 1988; Kelly and Sauer 1986; Sanders and McCormick 1987). (See Footnote 2.)

The task analysis and structured interview results presented here represent a preliminary attempt to describe the cognition of forecasters and research meteorologists, and a preliminary attempt to develop recommendations for the design of AMP systems. The resultant recommendations were referred to as "candidates" since many of them do not yet fully specify the design features of AMP systems. These point to areas where further research is needed.

5. Research needs

A host of questions about system operation will have to be addressed. It will be important to assess the quality of the human-computer interface early in the design process (Gordon et al. 1987). Will users find the system easy to use? Will they be able to understand how it works? How long will it take people to learn to use the system? What kinds of errors will they make? How long will it take them to recover from errors? How long will it take them to complete various tasks?

A number of the outstanding issues center on the design of graphics displays and the utilization of AI technology.

a. Color coding issues

The purpose of using color graphics involves more than just facilitating perception by presenting symbols or areas in color. It involves using colors to code information. Specific colors from a color palette are put into correspondence with particular values of a variable. For example, GOES infra-red data are coded in terms of grey tones, with the brightness of the tones corresponding systematically to the amount of measured infra-red energy. A color code can involve variation in hue, brightness, or saturation dimensions, or it can involve variations in more than one of these variables. In one of the NEXRAD displays, the brightness levels of two colors are used to indicate the different speeds at which winds are approaching (blue shades) or receding (red shades). The combined brightness and color differences make the relative wind speeds and directions readily apparent in the display (cf., Forsyth et al. 1985, Fig. 1).

Psychological research has shown that performance at searching a display is better if one knows beforehand something about the color and symbol codes that are to be looked at (Smith 1962). Hence, workstation systems usually rely on conventional symbols and colors, often based on common associations (Woodson 1981). In the case of meteorological cartography, many symbols are conventional, such as the symbols for fronts, lightning, cloud types, pressure changes, and wind speeds and directions. Some of those symbols rely on meaningful associations, such as the use of angled lines to denote the rising and falling of air pressure, the use of a jagged streak symbol to denote lightning, or the use of an asterisk-like symbol to denote snow. Some of the color codings in meteorology are conventionalized as well. For example, shades of red and blue are sometimes used to depict relative temperatures.

The extent to which AMP system displays should rely on conventional color or symbolic associations is an open, empirical question. Many types of meteorological color coding schemes are conceivable, and even conventional associations can be used in novel ways. For example, one researcher at the AFGL capitalizes on the color association of Doppler shifting, and codes temporal/movement information by enhancing clouds (i.e., in GOES images) such that leading edges are red and trailing edges are blue. Furthermore, just because
color is used to code information systematically, it does not always follow that the coding effectively portrays information.

For instance, a color coding scheme that is common in most areas of remote sensing maps values of the measured variable onto the primary colors of the visible spectrum. The "naive rainbow" is a nice concept, but in actual practice it leads to interpretation anomalies (Artis and Carnahan 1982). For example, in winter aerial thermography, poorly insulated houses can appear (to a novice) like trees (i.e., green blobs). On the other hand, trees are about the hottest things around, and they appear anomalously hot (i.e., white/yellow blobs). The NEXRAD dBZ (return signal strength) display uses a naive rainbow code which has a number of dark colors interspersed with the brighter ones. This code is redundant in that it is made up of combinations of hue and brightness which depict variations in signal strength. The variations of hue with brightness, however, are not systematic. In general, the use of brightness level in color codes (covarying hue, or covarying hue and saturation) in order to generate large color sets can lead to interpretation problems if it turns out that a critical difference in information value happens to fall at a poorly discriminable hue, brightness, or saturation boundary. Such possibilities cannot always be anticipated, but become apparent when specific color codes are applied to specific data sets.

AMP system users should be able to manipulate color and color coding schemes so as to enhance or suppress contrasts, but they need to appreciate the potential effects, vis-à-vis interpretation, of ill-conceived experimentation with color schemes. Meteorological forecasters and researchers who have developed prototype AMP systems have many opinions about color and color coding, but appear to have reached some consensus (cf., Schlatter 1985). Even points of consensus, however, can be researched and refined. Research on the perception and interpretation of color coded remotely-sensed imagery has only just begun (Hoffman and Conway 1989), but we know that empirical answers can be obtained to the lingering questions that AMP system technology presents. Should all the display colors be of equal subjective magnitude? Should some displays (e.g., GOES infra-red) be left as a grey scale, or should all univariate data be presented in colors? Would colors work better if we break from tradition and use a palette of pastels rather than bright, highly saturated colors?

b. The graphic depiction of mental models

In his treatise on cartographic communication, Tufte (1983) distinguished efficient interpretation from being forced to "puzzle out" the meaning of displays. Research has shown that the interpretation of cartographic information can be facilitated by using three-dimen-

sional or perspective representations, relative to flat, map-like displays (Bemis et al. 1988; Ware and Beatty 1988). Furthermore, research has shown that the direct representation of objects facilitates problem solving by permitting the manipulation of relevant variables and conceptual relations (Norman and Draper 1986). In other words, a main purpose of an interface is to support the development of mental models (Cechile et al. 1989; Doswell and Maddox 1986).

An AMP system's graphics component will need a robust "display engine" (Hollan 1984) which capitalizes on the ability of dynamic color graphics to condense a large amount of numerical data into an immediately perceptible form, while at the same time preserving scientifically valid conceptual relations (Jagacinski & Miller 1978; Jasperson and Venne 1987). A number of meteorologists have cited the utility of time series looping or animation of images (e.g., Schlatter 1985). This comes as little surprise since the mental models of meteorologists seem to be four-di-

mensional. That is, they involve dynamics in the sense of structural change over time and in the sense of the interaction of physical forces. One of the interviewee meteorologists, who had experience in operational forecasting, commented that tradition in meteorological cartography is so ingrained that some meteorologists' mental models depict isolines and fronts. This is not to deny the importance of dynamics, however. As an exercise, one can show a synoptic chart to a meteorologist and ask him or her to say what is happening in the weather at a given location at the depicted time (a "nowcast"). Sometimes this is a difficult thing to do, precisely because a single chart freezes out most of the important dynamics (Dyer 1987).

All of the experts in the structured interviews agreed that meteorology needs displays which depict the structure and dynamics of the atmosphere and the meteorologist's hypotheses (see also Chisholm et al. 1983; Forsyth et al. 1985; Seguin 1989). Considerable programming and research will be needed to create the needed displays, to create the operational modes, and to assess their utility.

The "human factor" that must not be left out is meaningful input from forecasters themselves, in a wide range of field locations, before any prototype specifications are set in concrete. Furthermore, it should not be assumed that tomorrow's forecasting tasks will be the same as today's. The moral is that AMP system prototypes will have to be flexible.

c. The promise of artificial intelligence

In order to be an effective manager of information, the AMP system information processing subsystem will almost certainly incorporate some techniques from AI (Dyer 1989). In theory, AI techniques could help meteorologists deal with the "data overload" problem, by
facilitating the integration of new data types into forecasting and research operations (Chisholm et al. 1983). In theory, an AI capability could facilitate the interpretation process by supporting meteorologists’ reasoning and their formulation of scientifically valid mental models. The use of such AI techniques as “expert systems” does not entail the goal of replacing the human operator. Rather, the goal is to create an “intelligent assistant” (Gordon 1988).

Such is the promise of AI applications, now a major topic of research (Fischhoff 1983). There has been considerable debate over whether computer systems can mimic experts successfully or can algorithmically reproduce human behavior. Most investigators see tremendous potential for expert systems (e.g., Reddy 1988), although others question whether the potential can ever be realized (e.g., Dreyfus and Dreyfus 1986; Graubard 1988). What is clear is that in order for expert systems technology to be successfully applied, the systems will need a number of capabilities that are not available in a majority of the current systems (Berry and Broadbent 1986; Carroll and Olson 1987; Kidd and Cooper 1985). The application of AI to meteorology in general, and AMP systems in particular, lies at the cutting edge of research in computer science, cognitive psychology, and human factors psychology. Indeed, some AMP system development projects are backing off from the incorporation of AI techniques, based on the belief that AI technology is not yet “ready” (Seguin 1989).

How are the mental models of meteorologists to be expressed in the language of AI? There are many different ways to represent knowledge in AI systems, including rules, frames, object-oriented representations, and networks. It is not always obvious a priori which formats will work in any given domain or subdomain of knowledge. Without some prototyping, it is rarely obvious which formats will work best. In their expert system for single station forecasting, Jasperson and Venne (1987) successfully used frames to describe various meteorological entities. Each entity, such as a front, pressure system, or air mass, was described in terms of its type, location, direction, intensity, and speed.

The most commonly used format for representing expert knowledge relies on the use of IF–THEN rules. As a given case is analyzed, the reasoning rules form a path or “forward chain” from input data to hypotheses (also called “data driven” reasoning), or a “backward chain” going from goals to the data that could confirm or disconfirm hypotheses (also called “knowledge driven” reasoning). A simple example of such a chain would be the statement: “If there is to be radiative fog, confirm the required values for cloud cover, wind speed, and surface moisture.” However, when one gets down to the gory details, the actual reasoning rules in expert systems can be quite cryptic: “IF

wind is 360 to 400, THEN Condition 1 is met, ELSE Condition 1 is not met.” A less cryptic and more detailed example, created for illustrative purposes, is presented in Table 4. (Once the details in a rule are specified—What are “baggy” isobars?—the actual implementable rule will be much more detailed and cryptic than the example presented in Table 4.)

It will be necessary to adapt tools and techniques from AI to the domain of meteorology, to integrate rule-based systems with meteorological dynamics and the reasoning patterns of meteorologists. A number of researchers have demonstrated that rule-based decision aids can be created which focus on important and recurring weather events, such as isolated severe storms, microbursts, wind shear, precipitation, hail, rain/snow line, snowfall, temperature highs and lows, and lightning (Campbell and Olson 1987; Moninger 1985; Zubrick 1984). An example is the “Zeus” system for predicting visibility (fog formation and dissipation) at certain air bases (Stunder et al. 1987). First, key parameters (such as cloud cover, humidity of the surface air layer, and strength of surface winds) were defined, as were the dynamic processes that occur (such as the loss of heat due to radiation, the cooling of air, and dew deposition). The parameters and processes were represented as a decision tree for an implementable model.

The AI sub-component of an AMP system will probably operate in the form of expert system modules, like Zeus. Each module would specialize in building and displaying models for particular situations. However, all of the specialist expert systems will rely on a knowledge base containing general physical principles (e.g., fluid dynamics), general principles of meteorology (e.g., front formation, the behavior of high and low pressure areas), principles that relate atmospheric events to features of the earth (e.g., the effects of mountain ranges, bodies of water, etc.), and mesoscale principles (those dealing with weather patterns on a local rather than global scale) (Dyer 1989; Jasperson and Venne 1987).

Before one can begin to build such expert system modules, one must capture and express the knowledge and reasoning patterns of meteorologists. The topic of expert knowledge is currently a major focus of cognitive research. For example, research has revealed reasoning heuristics (or “rules of thumb”) and judgmental biases (Hoffman et al. 1990; Wickens 1984). A number of studies, in the tradition of mathematical psychology, have examined meteorologists’ forecasting “hit rate” for specific predictions such as the probability of precipitation (Murphy and Brown 1984; Murphy and Winkler 1974b). The research has yielded some evidence for overconfidence bias in the judgments of meteorologists (Allen 1982). However, there has been relatively little research on the knowledge and ongoing reasoning of forecasters. Exceptions to this are studies
| IF | PRESENCE OF A WARM OR COLD FRONT,                      |
|    | AND UPPER AIR VORTICITY ADVECTION OVERSPLAYING THE FRONT, |
|    | AND A 500-MILLIBAR TROUGH OVERTAKING THE FRONTAL ZONE,  |
|    | AND PRESENCE OF MOISTURE IN THE LOW LEVELS OF THE ATMOSPHERE, |
|    | AND AN UNSTABLE ADIABATIC LAPSE RATE,                   |
| THEN-DO | INSPECT GOES VISUAL AND IR IMAGES,                   |
| AND IF | A WIDENING OF THE FRONTAL CLOUD BAND,                |
|        | AND THE DEVELOPMENT OR EXPANSION OF AN AREA OF HIGH CLOUDS OR DECREASING TEMPERATURE TO THE NORTHEAST OF THE NEW CYCLONE'S CENTER, |
| THEN-DO | INSPECT SURFACE OBSERVATIONS AND AVAILABLE RADAR DATA, |
| AND IF | STEADY OR INTENSIFYING PRECIPITATION IN THE REGION OF THE CLOUD GROWTH, |
|        | AND REGIONAL SURFACE PRESSURE ANALYSIS SHOWS AN AREA OF "BAGGY" ISOBARS ALONG OR JUST AHEAD OF THE FRONT, |
|        | AND PRESSURE FALLS OF 2-3 MILLIBARS OVER A THREE-HOUR PERIOD ALONG THE FRONT AND IN THE VICINITY OF THE "BAGGY" ISOBARS, |
|        | AND AN INTENSIFYING AREA OF WIND CONVERGENCE (DIVERGENCE OR STREAMLINE ANALYSIS), |
| THEN | CYCLOGENESIS IS PROBABLY OCCURRING.                   |

by Schlatter (1985) and by Moninger and Stewart (1986) and the task analysis reported here.

Experts' knowledge is extensive and highly coherent. In other words, knowledge elicitation is not the mere extraction of isolated "bits" of knowledge (Hoffman 1987). Furthermore, the conceptual basis of experts' reasoning is often tacit—that is, the details of reasoning may not appear explicitly in their verbal deliberations or judgments. Tacit knowledge, including knowledge of skills and procedures, may have to be elicited using special interview techniques or other methods from experimental psychology (Hart 1986; Hoffman 1991; Perby 1989). The elicitation and characterization of the knowledge and reasoning of expert meteorologists is likely to occupy researchers for some time.

Acknowledgments. The research reported here was conducted at the Atmospheric Prediction Branch of the U.S. Air Force Geophysics Laboratories (AFGL), with the support of a grant from the Summer Faculty Research Program (SFRP) (Contract No. F49620-85-0013) of the U.S. Air Force Office of Scientific Research. The author would like to thank Rodney Darrah and the staff at Universal Energy Systems, Inc.; A. S. Jursa and Agnes Bain, who were the AFGL co-ordinators for the SFRP; Rosemary Dyer and Donald Chisholm of the Atmospheric Prediction Branch of the AFGL; and the AFGL meteorologists and researchers whose expertise and patience made this research possible: Al Boehm, Robert d'Entremont, Ralph Donaldson, Jerry Freeman, Gordon Hammond, Arthur Jackson, Thomas Kleespies, Stuart Muench, Keith Roberts, and Larry Thomason.

The author would also like to thank the seven anonymous reviewers of previous versions of this article. They made many helpful suggestions, and clarified the differences of opinion and contrasts between the views of meteorological researchers and forecasters. Especially supportive and helpful was William Moninger of the NOAA Environmental Research Laboratory. Finally, thanks to Robin Akersrom for assistance in preparing the final manuscript.

The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes not withstanding any copyright notation herein.

REFERENCES


factors, Committee on Human Factors, Commission on Behavioral and Social Sciences and Education, National Research Council, Washington, DC.


Miller-Jacobs, H. H., 1984: Color displays and the color deficient
operator. Advances in Display Technology V. E. Schlar, Ed.,

Moninger, W. R., 1985: ARCHER: A prototype expert system for
identifying some meteorological phenomena. Report, Environ-
m ental Research Laboratories, National Oceanic and Atmo-
spheric Administration, Boulder, Colorado.

human information processing in weather forecasting. Report,
Environmental Sciences Group, National Oceanic and Atmo-
spheric Administration, Boulder, Colorado.


the requirements for the Military Airlift Command Crisis Action
Team Workstations. Air Force Aerospace Medical Research
Laboratory, Wright-Patterson AFB, Ohio.

Morgan, C. T., A. Chapnis, J. S. Cook and M. Lund, Eds., 1963:
Human Engineering Guide to Equipment Design. McGraw-Hill,
51–120.

Murphy, A. H., and B. G. Brown, 1984: A comparative evaluation
of objective and subjective probability forecasters in the United

Murphy, A. H., and R. L. Winkler, 1974a: Probability forecasts: A

Murphy, A. H., and R. L. Winkler, 1974b: Credible interval tem-
102, 784–794.


Office of Technical Services, National Oceanic and Atmospheric
Administration, U.S. Department of Commerce.

Norman, D. A., 1984: Cognitive engineering principles in the design
of human-computer interfaces. Human-Computer Interaction,


Perby, M.-L., 1989: Computerization and skill in local weather fore-
casting. Knowledge, Skill, and Artificial Intelligence, B. Göranzon

Pettersen, S., 1940: Weather Analysis and Forecasting. McGraw-
Hill, 411–490.

& Psychophysics, 37, 543–548.

Human Factors, 30, 143–151.

Reddy, R., 1988: Foundations and grand challenges of artificial in-
telligence. AI Magazine, 19–21.

Remington, R., and D. Williams, 1986: On the selection and eval-
uation of visual display symbology: Factors influencing search

Sanders, M. S., and E. J. McCormick, 1987: Human Factors in En-

Schlatter, T. W., 1985: A day in the life of a modern mesoscale
forecaster. ESA Journal, 9, 235–256.

——, P. Schultz, and J. M. Brown, 1985: Forecasting convection
with the PROFS system: Comments on the summer 1983 ex-

Seguin, W. R., 1989: Preparing for AWIPS-90 applications. Report,
Techniques Development Laboratory, National Oceanic and Atmo-
spheric Administration, Rockville, Maryland.

Smith, S. L., 1962: Color coding and visual search. Journal of Ex-
perimental Psychology, 64, 434–440.

knowledge-based expert system that assists in predicting visibility
at airbases. Report TR-87-0019, Air Force Geophysics Lab-
ratories, Hanscom AFB, Massachusetts.

Technology Systems, Inc., 1984: Meteorological data acquisition,
management, and processing system. Technology Systems, Inc.,
Lexington, Massachusetts.

Thorndyke, P. W., and S. E. Goldin, 1983: Spatial learning and rea-


Van Cott, H. P., and R. G. Kinkade, 1972: Human Engineering
Printing Office.

Ware, C., and J. C. Beatty, 1988: Using color dimensions to display

Wickens, C. D., 1984: Processing resources in attention, Varieties of

Wilkins, E. M., and R. E. Johnson, 1975: The AFOS experimental
system. IEEE Transactions on Geoscience Electronics, GE-13,

Hill, 1049 pp.

Zubrick, S., 1984: An expert system to aid in severe thunderstorm
forecasting. Report ZU-RS-00025, Radian Corporation, Austin,
Texas.