WEATHER FORECASTING AND THE PRINCIPLES OF COMPLEX COGNITIVE SYSTEMS

Robert R. Hoffman, Ph.D.
John W. Coffey, Ed.D.
Institute for Human and Machine Cognition
Pensacola, Florida

This presentation will illustrate some of the principles for the design of complex cognitive systems that were manifested in a recent project on weather forecasting (Hoffman, et al., 2000). One focus here is on how technologies and workspaces can facilitate or interfere with knowledge-sharing. A second focus is the notion of methodological opportunism. This is illustrated by a discussion of a new knowledge elicitation method, the Cognitive Modeling Procedure, which is intended to support the rapid refinement and behavioral validation (in the field setting) of macrocognitive models of practitioner reasoning.

INTRODUCTION

As we have all experienced when using new technologies ranging from VCR controllers to "advanced" decision aids, machines often force users to adapt. Weather forecasting has its share of information processing and display systems that are user-hostile. In order to generate specifications for good decision aids ("goodness" hinging on a systems approach to cognitive technologies), one must have a rich understanding of the knowledge, perceptual skills, reasoning strategies, and decision requirements of the end-users. Only on the basis of such understanding can one make decision aids that are both usable and useful.

An attempt was made to reach such an understanding in a two-year research project (Hoffman, et al., 2001) that relied on the participation of 22 apprentice, journeyman, and expert forecasters at the Naval Training Meteorology and Oceanography Facility at Pensacola Naval Air Station. A variety of methods of Cognitive Task Analysis were brought to bear in this project—one goal of the project was a systematic comparison of alternative methods of cognitive work analysis/cognitive field research (Hoffman, 2002). Methods employed included workplace analyses, proficiency scaling, the Critical Decision Method, protocol analysis, Knowledge Audit, and Concept Mapping. In addition, a new procedure was created, which we call the Cognitive Modeling Procedure (CMP) (Hoffman, et al., 2000). This procedure, discussed in the final section of this paper, evolved from and illustrates the application of principles of complex cognitive systems.

SOME DESIGN PRINCIPLES

In the usual human factors sense, design principles can take the form of fairly specific guidelines. An instance might be "To make display elements legible, do not use blue to colorize thin lines in graphical objects." Design principles for complex cognitive systems need to be couched in terms of higher-level, interacting, continuous, parallel processes such as sensemaking and situation awareness. Taken at this macrocognitive level of analysis (Klein, et al., 2003), "principles" cannot be cookbook-style specifications. Rather, they can be construed as challenges for design: In order for a system to be human-centered it must fall within a design space that is defined by the principles. If a system falls outside of that constraint space, it will not be human-centered and will likely result in technology that is brittle, that triggers automation surprises, and has user-hostile features.

This presentation cannot cover all of the more than two dozen principles that were manifested in the research, and so I will discuss just a few examples.

The Lewis and Clark Principle states: The human user of the guidance needs to be shown the guidance in a way that is organized in terms of their major goals. Information needed for each particular goal should be shown in a meaningful form, and should allow the human to directly comprehend the major decisions associated with each goal (Endsley & Hoffman, 2000). In the weather forecasting operation we saw what happens when this principle is violated. When pilots would file their flight plans, forecasters would prepare for them a certain weather briefing form to describe the forecast for the
flight. The form is a single page crammed with boxes in which graphical symbols and cryptic acronyms are used to code weather (e.g., severe weather, turbulence, etc.) for each leg of the flight. An examination of dozens of the forms, for easy flights and simple ones, easy weather and rough weather, revealed that in most cases more of the boxes were unfilled than filled. The intent of the form is "one size fits all," but the effect is one size fits none. The form is obviously a remnant of an older technology and mind set, intent on cramming as much as possible onto a single side of a piece of paper. Experienced pilots are accustomed to the form, and even say they like it. What we observed repeatedly was that pilots and pilot trainers would come to the facility to get their forms, glance at them, and then go to that one expert forecaster whom they trusted, and ask things such as, "So how bad will that chop really be over the Rockies?" The form did not conform to the Lewis and Clark Principle.

What we also see here is a manifestation of the True Work Principle: True Work involves acquiring and acting on the basis of knowledge or a certain state of situation awareness. As Vicente (1999) has pointed out, True Work is often different from the actual work, which is shaped by the technology ("environmental constraints"). Our appreciation of this principle led to a series of interviews about things pilots really wanted to see, which in turn led to a re-conceptualization of the format for the weather briefing form. For instance, pilots wanted to see a satellite image that would provide them the "big picture" of the weather.

This story relates to yet another of the principles, the Principle of Client Participation: Clients of the users, as well as the users, need to be a part of the design team. One might create a fantastic system for helping forecasters predict the weather, but unless that system helps them provide weather products to their clients (pilots, in this case), it may not be either as effective or useful as one might hope. This principle encourages one to look beyond the misleading notion of the "end-user."

Another principle we saw in operation is the Principle of Trickle-Off Ergonomics: Prototyping never ceases, it just trickles off. This was manifest in a number of ways: New software builds that altered the work patterns, new workstation platforms, ever-changing local solutions (e.g., Post-its), and lots of kluges. We believe this principle is an empirical fact, if not a defining feature of the sociotechnical workplace. The fact that it is always a moving target complicates any field research aimed at investigating the envisioned world (Woods & Dekker, 2001). The principle also has implications for systems design, especially the idea that the "deliverable" has to be a finished (or nearly finished) product, as opposed to a system that is deliberately designed such that the (end?) users can adapt it to their individual and local needs.

Another of the principles that was manifested in the forecasting study was The Fort Knox Principle: The knowledge and skills of senior workers (the "gold") must be elicited and preserved in such a way that it can be shared. The forecasting facility preserved its knowledge in two forms: Videotapes of training briefings that would be presented to the staff by the forecasters, and files of data (images, observations) that a few individual forecasters kept concerning interesting or tough cases that they had experienced. Both were rich sources of information, but were kept cloistered. Furthermore, most of the organization's knowledge and forecasting heuristics resided in the heads of the senior forecasters, and when they retired, it went out the door with them. It was for this reason that knowledge elicitation, preservation, and sharing was identified as a leverage point and was selected for our prototyping effort. (For more details see Hoffman et al, 2000.)

Another noteworthy principle that was manifested in the forecasting case study was The Principle of Methodological Opportunism: Procedures used for data collection, data analysis, and system development can be developed and applied opportunistically. A traditional process for data collection and analysis attempts to follow a preplanned agenda. When it comes to the development of complex cognitive systems, cognitive field research should involve methodological innovation and even ravenous opportunism. If one is not always open to innovating the methods, one may not be primed to take advantage of the unique or serendipitous aspects of the immediate context. Instances of this Principle were:

- During a pre-planned workspace analysis, an important client (a squadron-level pilot trainer) arrived at the workplace. There was an opportunity to conduct an informal and semi-structured interview with this client. The Workspace Analysis was put on hold, a series of potentially useful probe questions was quickly conceived on the fly, and a valuable interview with the client was conducted.

- In some domains of practice, performance measurement is routine. An example would be the "skill scores" of weather forecast accuracy. Such measures can be integrated into the work analysis methodology as a ready-made metric for proficiency scaling. In our case, it enabled us to compare actual performance to a proficiency scale (based on career interviews) that distinguished apprentices, journeymen, and experts (and levels within each category) in terms of depth and breadth of experience. One of the interesting findings was that expertise in this domain applies to individuals having many tens of thousands of hours of experience, well
beyond the 10,000 hr. rule of thumb that is often cited in cognitive psychology.

- A recommendation from ethnomethodologists is that work analysis should include extensive observations of worker activities, often based on analysis of videotapes. An argument for doing this is that one never knows beforehand where one might discover important collaboration patterns or leverage points for effective change. In the weather forecasting case study, we learned that apprentice forecasters gave weekly weather briefings to the senior forecasters, who would probe the apprentices concerning their reasoning. These briefings were always videotaped. Analysis of those tapes afforded insights into forecaster reasoning that enabled us to forego videotaping of weather discussions and other operations floor activities.

A final instance of the Principle methodological opportunism involved the creation of an entirely new method, the Cognitive Modeling Procedure (CMP). We conclude with a discussion of this procedure, as an invitation to others to adapt and test it further.

THE COGNITIVE MODELING PROCEDURE

Weather forecasters have been comfortable with the notion of a "mental model" for many years (see Hoffman, Trafton, & Roeber, in preparation). They refer, for example, to their process of inspecting satellite imagery in order to "get the big picture" of what's going on in the weather. Their descriptions, many of them in the published literature of meteorology, suggested that the forecasting process involves: (1) The generation of a mental model that represents atmospheric dynamics in the medium of mental imagery, shaped by a knowledge of hydrodynamical principles and concepts, and (2) The refinement of the model in successive waves of data inspection and hypothesis testing. Both of these notions, attributable to Karl Duncker, are central to our current understanding of the nature of expertise. Furthermore, meteorologists are accustomed to the notion of computational forecasting models, and thus have found themselves needing to distinguish such computer models from their own understanding. Hence, meteorologists themselves have referred to the notion of a "mental model" as a part of the diagnosis-prognosis sequence in weather forecasting (e.g., Doswell, 1986).

During our cognitive field research we had occasion to discuss this ideas with a journeyman forecaster who was participating in one of our workspace analyses. In the discussion, the interviewer sketched a general model of expertise, depicting a sequence that began with data examination, which led to the formation of a mental model, which in turn would suggest hypotheses that would be tested against other data. The Participant responded that it seemed somewhat reasonable but that forecasters do other things as well:

- They engage in (what we in human factors call) recognition-primed decision making,
- They vigilantly maintain situational awareness, and
- They actively maintain (what we call) an action queue.

Our Participant also asserted that some forecasters do not reason quite in the way the initial informal sketch suggested—some forecasters rely more heavily, and in different ways, on the outputs of the various computer-based forecasting models. Our Participant was even able to suggest names of colleagues who reason in ways different from that depicted in the sketch. This suggested that we had here the germ of a method for conducting sociogrammetry as well as a method for developing models of reasoning.

The CMP is intended to provide a "fast track into the black box," in that it might be used to generate, refine, and behaviorally validate models of practitioner reasoning, and do so more efficiently than, for example, by the method of think aloud problem solving (Hoffman, Coffey, & Carnot, 2000).

The CMP begins with what we call the Base Model of Expertise, shown in Figure 1. This conforms to what is known from studies of expertise in diverse domains.

![Figure 1. A model of forecaster reasoning.](image)

From the base model we created "bogus models." These are shown in Figure 2. The bogus models involved adapting the Base Model to the domain,
and were intended to include some of the basic strategic elements, but do so in a way that is not-quite-right but is also not totally unrealistic. In the first step of the CMP, the Participant is invited to choose the one bogus model that they feel best fits their own general forecasting strategy.

![Image](36x82 to 562x543)

**Figure 2.** The “bogus models” of forecaster reasoning. “MOS” stands for model output statistics (i.e., the outputs of computer forecasting models).

Next, the Participant and the interviewer work together, using the elements within the bogus models as a scaffold to create a diagram that the Participant feels is a better representation of their overall or general strategy. After the sketch is rendered as a graphic image, it is verified in a follow-up discussion with the Participant.

In the sociogram step, deliberately conducted after some weeks' delay, all of the Participants' models (and additional bogus models) are presented to each of the Participants, with a guessing game task: Who does this model relate? This step allows one to probe hypotheses about expertise (Shouldn’t experts have a common strategy?) and hypotheses about knowledge sharing and mentoring within the organization.

In the next step, each of the Participants’ models is validated by workplace observation. Some elements of models can be validated by behavioral observation (e.g., the expert had asserted that he always first inspects a certain data type on his displays). Model elements that could not be directly observed were the subject of appropriately-timed probe questions (e.g., when the forecaster called up a data field, s/he was asked about their reason or goal). Figure 3 shows one of the Participant’s models, with annotations from the validation phase of the procedure.

**Figure 3.** An example Participant model, with annotations from the behavioral validation.

Results

Results from the procedure included refined models of the reasoning of seven forecasters, five of which underwent the behavioral validation step. (Two journeyman participants were reassigned to other duty stations after the first CMP step had been conducted.) The models were drafted, refined, and validated in an average total task time of 52 min. This is without doubt considerably less than time that taken to reveal and verify reasoning models in experiments using think aloud problem solving and protocol analysis. (Transcription alone takes many hours.)

In the sociogram, three of the Participants correctly identified the bogus models, but others saw those models as being apprentice models. (So the bogus models were not-too-unrealistic.) Most of the Participants correctly identified only one or two of the models. These results revealed an important leverage point for this organization: Except for scattered chance experiences, they did not really know much about how each other reasoned and strategized. The forecasters were not sharing information about their reasoning.

The sociogram results also yielded an interesting cautionary tale about modeling in this domain. One of the five Participants who did not correctly identify his own model asserted that this was no longer a good
depiction of how he worked (especially in terms of his choice of computer model outputs). After his “miss” was pointed out and he retrospected about the first step in this procedure, he explained that the model had been made during a weather regime that was no longer in effect. Any intelligent decision aid for forecasting is bound to be of limited usefulness if it is based on assumption that forecasters engage in one or a few reasoning sequences, even in a single region or climate.

The CMP results also yielded a strong consensus validation of the features of the base model (i.e., the importance of mental model formation, the mental model refinement cycle, the situation awareness cycle, etc.). This shows clearly that a major goal for forecasting decision-aiding should be to provide a system that supports the forecaster in generating, manipulating, and verifying a graphical 4-D representation of their mental models of atmospheric dynamics (Hoffman, 1991). Current systems are only beginning to approximate such a capability. Meeting this goal will require the judicious application of all of the principles of human-centered computing that govern complex cognitive systems.

ACKNOWLEDGEMENTS
The work reported here was supported by a contract from the National Technology Alliance.

The Authors would like to thank the Command, Officers, aerographers, and forecasters at the Naval Training Meteorology and Oceanography Facility at Naval Air Station Pensacola. Fair winds and following seas.

He senior author would like to express his appreciation to David D. Woods for his many contributions.

REFERENCES


