# 2 Competence in Weather Forecasting

*Rebecca M. Pliske, Beth Crandall, and Gary Klein*

In recent years technological innovation has dramatically transformed the tools available to weather forecasters. Today's forecasters often find themselves overwhelmed by the sheer volume of computer products and displays that are now available (e.g., radar images, satellite pictures, computer-generated models). In this chapter, we describe a study we completed with weather forecasters that examined the development of competent performance in this context.

Weather forecasters have been studied for many years by both judgment and decision researchers and cognitive science researchers. Although a review of all this work is beyond the scope of this chapter, before describing our own research we briefly summarize some of the previous research conducted with weather forecasters. According to Stewart and Lusk (1994), the performance of a forecaster will depend on three factors: the environment about which the forecasts are being made (how predictable is it?), the information system that brings data about the environment to the forecaster (how reliable is the information provided?), and the forecaster's cognitive system (the forecaster's perceptual and judgmental processes). They point out that some weather phenomena are much more predictable than others (e.g., tomorrow's high temperature versus the size of the hailstones that will be produced by an approaching severe storm) and that some sources of information are more reliable than others (e.g., human observers are more

This research was supported by a contract from Armstrong Research Laboratory, Brooks Air Force Base, Texas. It was sponsored by the Air Weather Service, Scott Air Force Base, Illinois. We thank Anna Rowe and Carol Weaver for their help in obtaining the cooperation of the participants for this project. We also thank David Klinger, Robert Hutton, and Betsy Knight for their help in data collection and analysis.





reliable than electronic sensors for describing cloud cover). Stewart and his colleagues (Stewart, Heideman, Moninger, & Reagan-Cirincione, 1992) have also shown that forecasters use only a subset of the available information, and as the amount of information used increases, so does the unreliability of the forecasts.

Other research with forecasters described in the judgment literature has focused on *calibration*, which is defined as the extent to which decision makers' confidence matches their accuracy when making a set of judgments. Although many laypeople might disagree, judgment researchers have demonstrated that weather forecasters are well calibrated (Murphy & Winkler, 1977). For example, if a forecaster predicts a 60% chance of rain for a city, then rain is typically reported about 60% of the time.

Cognitive scientists have also studied weather forecasters. For example, Hoffman (1991) conducted interviews with 10 meteorologists in order to develop recommendations for the design of expert systems to support weather forecasters. Hoffman concluded that forecasters develop initial mental models based on the information presented to them on their various displays. These initial mental models imply various hypotheses, which the meteorologists then proceed to test. The particular type of information sought out by a particular meteorologist depends on the weather scenario for a given day.

The previous research conducted with weather forecasters gave us a starting point for our work with weather forecasters. We were funded by the U.S. Air Force (USAF) to conduct a cognitive task analysis of weather forecasters in order to provide recommendations on how to improve forecasting performance. Cognitive task analysis attempts to describe the cognitive activities that underlie an experienced person's job or task performance. In the past decade, the USAF has made an enormous investment in the new technologies available to support weather forecasting in an attempt to offset the effects of reductions in senior-level personnel. Unfortunately, weather forecasting performance has declined rather than improved since the introduction of many of these new technologies.

Our USAF sponsors were concerned that forecasters may have become overly reliant on computer models when developing their forecasts. These models are meant to be only one of many sources of information considered by the forecaster. A similar finding, called the *automation bias*, has been documented in other high-tech domains (Mosier & Skitka, 1996; Mosier, Skitka, & Heers, 1996). Mosier and her colleagues have found that some pilot errors result from overreliance on automated systems.

The majority of USAF weather personnel work in base weather stations that are equipped with all the latest technology available to support weather forecasters (e.g., Doppler radar, satellite imagery, and state-of-the-art decision support systems). These forecasters are referred to as *in-garrison forecasters*, and their job is to provide weather information to support air operations. Other USAF weather personnel work on cadre weather teams that are dedicated to supporting operational U.S. Army units (artillery, tanks, etc.), Special Operation units, and Army Ranger units. These forecasters work solely with tactical weather equipment and are referred to as *tactical forecasters*. We studied both types of forecasters for this project, which is described in detail in Pliske et al. (1997). In this chapter, we describe the results of our research with in-garrison forecasters and with National Weather Service (NWS) forecasters.

This research was guided by the Naturalistic Decision Making (NDM) theoretical perspective, which studies how people use their experience to make decisions in field settings. The NDM perspective investigates the strategies people use in performing complex, ill-structured, and high-stakes tasks under time pressure and uncertainty, and in the context of team and organizational constraints (Klein, Orasanu, Calderwood, & Zsambok, 1993; Zsambok & Klein, 1997). In many dynamic, uncertain, and fast-paced environments, there is no single right way to make decisions. Thus, NDM researchers typically study experts to define quality decision making and describe good decision-making processes. Researchers using the NDM framework have examined expert performance with a wide variety of professionals such as firefighters (Klein, Calderwood, & Clinton-Cirocco, 1986), critical care nurses (Crandall & Getchell-Reiter, 1993), weapons directors (Klinger et al., 1993), anti-air warfare command and control officers (Kaempf, Klein, Thordsen, & Wolf, 1996), pilots (Orasanu & Fischer, 1997), and electronic warfare officers (Randel, Pugh, & Reed, 1996). By studying the cognitive aspects of expert performance in these domains, NDM researchers have been able to make recommendations on how to improve training and system support to facilitate the performance of nonexperts.

When we began to study the weather forecasting domain, we were surprised to discover that it was very difficult to identify true experts due to the lack of operationally defined criteria for optimal performance. Prior to conducting our study, we believed that determining the accuracy of a weather forecast would be very straightforward (i.e., was the forecast correct or not?). However, we learned that validating the accuracy of a forecast is not as simple as it seems. For example, a forecaster may forecast rain in a certain city; if it rains in only half of the city, was the forecaster right or wrong? Furthermore, we learned that forecasters are not judged, by their peers or supervisors, to have made a bad forecast as long as their rationale for the forecast was sound. There is a widely shared belief that the weather is unpredictable given our current knowledge of the atmosphere and current technology. Therefore, a given forecast can be inaccurate, but the forecaster is still considered to have been correct in his or her prediction. Because the weather forecasting domain lacks clear criteria for defining optimal performance, we believe it fits the definition of competency-based performance as described by Smith, Shanteau, and Johnson (this volume).



We started our search for expert weather forecasters by interviewing forecasters with many years of experience. We quickly learned that expertise in weather forecasting is not purely a function of years on the job. This was not a surprise. In other domains we have studied, the opportunity for the decision makers to experience many different scenarios and situations is often a more important determinant of expertise than the number of years they have held their current positions. For example, an urban firefighter may attend several calls a day and become an expert in only a few years, whereas a volunteer rural firefighter may only attend one call a week and may never develop a high level of expertise.

Because there was no objective criterion for identifying true experts in the weather forecasting domain and because we could not equate years of forecasting experience to expertise, we developed a description of expertise in this domain based on the results of our own observations and interviews with weather forecasters. In the following sections, we describe the methods we used to collect and analyze our data and the conclusions we drew. We conclude with a discussion of the importance of metacognitive skills in the development of forecasting expertise.

## Methods

### *Sample Description*

Our research team interviewed USAF weather forecasters stationed in Ohio, Florida, Alabama, Texas, and Colorado. We conducted a total of 22 in-depth interviews with USAF in-garrison weather forecasters.

Our sample included 5 civilians (all of whom had previous active duty experience in the Air Force), 1 forecaster who was currently serving in the Air Force Reserves, and 16 active duty enlisted personnel. The number of years of experience as a forecaster varied from 4 months to 21 years, with a median of 11 years.

In order to gain a broader perspective on the weather forecasting domain, we interviewed two NWS forecasters at an NWS office in the Midwest, and we interviewed 13 forecasters who had worked as part of the 1996 Olympic weather support forecasting team. The number of years of forecasting experience reported by the NWS forecasters ranged from 3 to 25, with a median of 7 years. Several NWS participants came from forecasting jobs that involved producing regional aviation forecasts or national forecasts for severe storms, hurricanes, or heavy precipitation. Others came from NWS forecasting offices in California, Florida, Georgia, Texas, Arizona, and Alabama.

### *Procedure*

We used a semistructured interview procedure that included two knowledge elicitation methods: the critical decision method (CDM) and the knowledge audit. The CDM uses an event-based approach and is organized around an account of a specific incident. Interviewees are asked to describe an incident in which their expertise made a difference. For the weather forecasting domain, we found that a useful opening query was to ask forecasters to describe a time when they made a forecast and other people thought they were wrong, but it turned out that they were right. Once a relevant incident had been described by the interviewee, the interviewer led the interviewee back over the incident to solicit specific information. Elicited information included the presence or absence of salient cues and the nature of those cues, assessment of the situation and the basis for that assessment, expectations about how the situation might evolve, goals considered, and options evaluated and chosen. For a more complete description of the CDM, see Hoffman, Crandall, and Shadbolt (1998).

The knowledge audit includes a series of questions (or probes) that ask interviewees to provide specific examples of expertise in a domain with which they are familiar. This method is typically used with interviewees who have a high level of expertise in the domain. However, it can also be used with nonexperts to provide concrete, domain-relevant examples that reflect a wide range of proficiency. The knowledge audit elicitation technique was developed from the literature on expert–novice differences (Chi, Feltovich, & Glaser, 1981; Dreyfus, 1972; Dreyfus & Dreyfus, 1986; Klein & Hoffman, 1993; Shanteau, 1988). The set of knowledge audit probes was designed to help subject matter experts articulate their expertise. Each probe addresses an area identified in the literature as representing key areas of distinction between the performance of experts versus that of novices. These areas include diagnosing and predicting, situation awareness, perceptual skills, developing and knowing when to apply tricks of the trade, improvising, metacognitive skills, recognizing anomalies, and compensating for equipment limitations. For a more complete description of the knowledge audit method see Militello and Hutton (1998). An example of the types of probe used for this study is as follows:

> Equipment can sometimes mislead. Novices usually believe whatever the equipment tells them; they do not know when to be skeptical. Have there been times when the equipment pointed in one direction but your own judgment told you to do something else? Were there times when you had to rely on experience to avoid being led astray by the equipment?

### *Interview Process*

Interviews took place in private offices or conference rooms, depending on what was available at the different locations. Interviewees were told that the content of the interviews would be confidential. We requested permission to audiotape the interviews for subsequent transcription for verification of our notes. All but one of the forecasters agreed to this procedure. For most of the interviews, two researchers were present; for a few of the interviews, three researchers attended. Interviews lasted for 1 to 2 hours. Due to time constraints, the interviews conducted with the Olympic weather forecasters were conducted in groups of three to four forecasters, whereas all the other interviews were conducted with individual forecasters.

In addition to conducting in-depth interviews, researchers observed the forecaster's work environment and made note of the types of equipment available, the number of personnel on duty, the types of customer requests for weather information, and other relevant incidents. Researchers watched forecasters carry out their duties during routine work shifts and talked briefly with the forecasters on duty about what they were doing.

## Results and Discussion

### *Data Summary*

Analyses were based on notes of interview and observation sessions. Typically, one researcher prepared a detailed summary of the interview content. A second researcher, who had also been present for the interview, reviewed the summary and made additions or revisions. We transcribed sections of interviews that contained critical incidents or clear descriptions of the forecasting process. The interview notes and transcripts were systematically reviewed for specific types of information relevant to our goal of developing a better understanding of the knowledge and skills underlying expert weather forecasters. For example, we made lists of all the comments related to good versus bad forecaster characteristics, examples of information overload, the types of information forecasters used, and so on. In addition, we conducted the two analyses described in this section.

*Sorting Forecasters into Similar Categories.* We used a card-sorting process to determine if we could identify different types of USAF forecasters. Each of the 22 in-garrison forecasters' names were written on a file card.[1] Working as a group, the four researchers who had interviewed these forecasters sorted the cards into categories of individuals who were similar to each other. Once all four researchers agreed that the forecasters within each category were similar to each other and different from those in the other categories, we identified the common elements that defined each category. This sort resulted in a set of five forecaster groups and a sixth group of forecasters who could not be reliably classified into any of the other five groups. Due to the subjective nature of these classifications, we decided that it was preferable not to categorize forecasters into any category unless all four researchers were confident of the classification. The final step in the categorization task was to develop a definition for each category and a label that captured the central meaning of the category. We agreed

[1] This analysis was done before we conducted the in-depth interviews with the tactical forecasters, so they were not included. The NWS forecasters were not included because most of these interviews were conducted with small groups and did not elicit sufficient detail on individual forecasters' processes. However, what we learned from these interviews with "true" expert forecasters undoubtedly had an impact on how we sorted the USAF forecasters into categories for this analysis.



on the following labels for defining the five categories of forecasters: Intuitive-based Scientists, Rule-based Scientists, Procedure-based Observers, Procedure-based Mechanics, and Disengaged Forecasters. Attributes and forecasting processes that characterize each group will now be described.

We called one group of forecasters Intuitive-based Scientists. These forecasters seemed to love the weather at some basic affective level. They could talk in vivid detail about past weather events. Their mental representations of the weather were highly visual, at times even tactile in nature. For example, these forecasters recalled incidents in which they described feeling the amount of humidity in the air. Their descriptions of the weather forecasting process included rich visual imagery and the use of dynamic mental models and simulations that allowed them to mentally construct weather systems and "observe" changes in these systems over time. These forecasters evidenced high-level skills for pattern recognition and flexible use of various sources of information as they attempted to solve the problem of the day. Although these forecasters seemed to have a highly developed intuitive understanding of weather dynamics, they did not seem to think in terms of weather rules. Only two of the USAF forecasters included in our sort fell into this category; they had 11 and 15 years of forecasting experience.

A second group of forecasters we labeled the Rule-based Scientists. These forecasters were characterized by an extensive knowledge base of meteorological rules. They used these rules to construct a complete understanding of the current weather situation. They knew how to use a wide variety of tools to obtain the information they needed. They were able to detect patterns of cues presented on a variety of information sources (satellite, radar, etc.) and integrate this information into a useful mental representation. These forecasters seemed to have a good sense of the physical dynamics of various weather systems. Their self-reports of the forecasting process reflected an analytic reasoning style characterized by use of critical thinking and reasoning skills. These forecasters typically had experience forecasting at a variety of geographic locations, and this seems to have contributed substantially to their development of expertise. Four of the USAF forecasters included in our sort fell into this category; they had between 12 and 21 years of experience.

We labeled another group of forecasters Procedure-based Observers. They approached the weather forecasting task as a rule-based, procedural task. Unlike the Rule-based Scientists, however, the forecasters in this group could not seem to use their rule-based knowledge to construct a detailed understanding of the current weather situation. This group was also characterized by their love of the weather and their keen observation skills. Their verbal reports contained occasional glimmers of the higher-level understanding that characterized the Rule-based Scientist and Intuitive-based Scientist groups, but this higher level of understanding disintegrated with further probing by the interviewer. Much of their knowledge base was limited to the types of weather patterns they had observed at a specific location. They seemed to lack the understanding of weather as a global system that characterized both of the scientist groups previously described. Three of the USAF forecasters included in our sort fell into this category. This group of forecasters primarily included younger forecasters with relatively few years of forecasting experience (2–8 years). It is our belief that with more experience and training opportunities, they would develop into weather scientists.

We also identified a group of forecasters whom we called Procedure-based Mechanics. Their approach to the weather forecasting task was to complete a relatively fixed set of procedures. These forecasters had a limited knowledge base of meteorological rules. They appeared to look at the same sources of information, in the same sequence, every day. (We interviewed forecasters only during the summer months, but we believe that this group of forecasters varied their standard set of procedures according to the season of the year.) When asked to describe the thinking behind their weather forecasting process, these forecasters got to a point where they verbally faded out, wandering to the end of a sentence and just stopping. We considered these forecasters to be "locally proficient" because they knew enough to produce a reasonable forecast and give an acceptable pilot briefing, but they did not appear to be motivated to improve their forecasting skill. Three of the USAF forecasters included in our sort fell into this category; they had between 5 and 11 years of forecasting experience.

A fifth category we described as Disengaged. These forecasters had a very limited knowledge base of meteorological rules specific to their current location. Our perception was that they could use these rules to produce a marginal forecast. They did not like being weather forecasters, and they did not seem to like to think about the weather. One of the forecasters in this group had only a few months of forecasting experience, but the others had sufficient experience to have developed a more advanced level of expertise (2–12 years). Four of the USAF forecasters included in our sort fell into this category.



### *Levels of Expertise*

After completing our descriptions of the five categories of forecasters, we were struck by the similarity of the ordering of our forecaster categories to the five levels of expertise described by Dreyfus and Dreyfus (1986). Their model claims that novices initially make decisions and take action by learning a set of rules. As they gain competency, they compile these rules into more and more comprehensive and abstract rules that facilitate more efficient actions. However, Dreyfus and Dreyfus claim that there is a qualitative difference that separates individuals performing at the highest level of competency from nonexperts who have task experience. Experts do not simply rely on more highly developed rules or proceduralized knowledge; they possess a qualitatively different type of knowledge representation.

Dreyfus and Dreyfus have described the development of expertise as proceeding through the five levels described in Table 2.1. The forecasters whom we labeled Intuitive-based Scientists were most similar to the *expert* level described by Dreyfus and Dreyfus. These individuals had an intuitive grasp of the current situation, and their performance was characterized as fluid, flexible, and highly proficient. Dreyfus and Dreyfus label the next level of expertise *proficient*, and this category corresponds to the category of forecasters we labeled Rule-based Scientists. These individuals formed a holistic representation of the situation, had keen perceptual skills, and had a large experience base that they used to determine the typicality of the present situation. Although these individuals were very flexible, they did not rely on their intuition, but instead used their rule-based knowledge of meteorology. Dreyfus and Dreyfus label the next level of expertise as *competent*, and this label relates to the categories of forecasters we called Procedure-based Observers and Procedure-based Mechanics. These individuals were able to perform well given a limited range of situations, but they lacked the flexibility and knowledge that would allow them to perform well in a wider variety of situations. The next level of expertise is labeled as *advanced beginner* by Dreyfus and Dreyfus, and this level corresponds in some ways to the category of forecasters we called Disengaged. These individuals were able to recognize recurring patterns of features in their environment and operate using general procedures. The *novice* category described by Dreyfus and Dreyfus does not seem to relate directly to any of the forecaster categories we described. This is probably due to the fact that we interviewed only one forecaster who had less than 2 years of forecasting experience.

Table 2.1. *Levels of Expertise*

*Novice*

Beginners have had little experience in the situation in which they are expected to perform. Their initial learning about the situation is in terms of objective, measurable attributes. These attributes can be recognized without situational experiences because novices are very limited in their understanding; their behavior is limited and inflexible.

*Advanced Beginner*

Advanced beginners have coped with enough real situations to recognize recurring, meaningful situational components. At this level, understanding the aspects of the situation is limited to global characteristics that reflect prior experience in actual situations. Advanced beginners need help setting priorities, because they operate on general guidelines and are only beginning to perceive recurrent, meaningful patterns.

*Competent*

Performers at a competent or journeyman's level can see their actions in terms of long-range goals or plans. They are able to formulate, evaluate, and modify goals and plans. These plans are generated in terms of the current and future aspects that are most important. The competent performer lacks the speed and flexibility that emerges at higher levels of expertise, but has a sense of mastery and the ability to cope with a variety of situations.

*Proficient*

Proficient performers perceive situations as wholes rather than in terms of components. Their performance is rule-based and aided by well-developed perceptual skills. Proficient performers have learned what typical events to expect in a given situation and how to modify plans in accord with these events. They also recognize when the expected typical picture does not materialize and modify their plans and goals accordingly.

*Expert*

Expert performers no longer rely on analytic principles (rules, guidelines) to develop their understanding of the situation to select an appropriate action. The expert, with an enormous background of experience, has an intuitive grasp of each situation and focuses on the accurate region of the problem without consideration of a large range of irrelevant, alternative diagnoses and solutions. Expert performers are no longer aware of features and rules, and their performance becomes fluid, flexible, and highly proficient.

*Source:* Adapted from Dreyfus and Dreyfus (1986).



## *Content Analysis*

In order to examine the data for the cognitive activities involved in the forecasting process, we generated a list of cognitive activities that had been described in the interviews. Two researchers then reviewed the notes from the 22 interviews with in-garrison forecasters to identify which of these cognitive activities occurred with sufficient frequency to allow us to code these activities reliably across the different forecasters. The following activities, which are defined later, were identified for subsequent analysis: noticing patterns, seeking information, meaning making, use of visual mental representations, and metacognitive processes. The initial review of the interview notes also identified five interviews with in-garrison forecasters that seemed to lack sufficient detail for the content analysis. The participants in these five interviews were not currently involved in producing forecasts because they had been promoted to management positions, and they did not describe their forecasting process in a very detailed manner.

The coding of the cognitive activities was carried out by a pair of coders as follows. Working with interview notes and preselected portions of transcripts from interviews with 17 forecasters, one coder made an initial pass through the protocol data, highlighting any portion of the data that contained information about self-reported cognitive activity. No attempt was made at this point to determine the nature of the cognitive activity, just that it occurred. Working with an initial set of five protocols, two coders worked independently to evaluate the highlighted portions of the protocols, assigning codes to any cognitive activity category that appeared to apply. The two coders then compared category codings, discussed disagreements, and refined the category definitions. After the practice coding was completed, a second set of five protocols was coded using this same process. The two coders agreed on 69% of the codes; differences were resolved by consensus. The remaining data were divided between the two coders to be coded separately.

The results of the content coding are summarized in Table 2.2. Note that the numbers shown in this table are the average number of cognitive activities coded for a particular cognitive activity for each skill level, collapsed across individual forecasters. We used a coding scheme that allowed an individual forecaster to receive more than one code per category if the forecaster described multiple instances of those cognitive activities during the interview. In order to develop the skill level classification used for this analysis, a second category sort was conducted.

Table 2.2. *Mean Number of Cognitive Activities by Forecast Skill Level*

| Skill Level | Cognitive Skill: Noticing Patterns | Seeking Information | Meaning Making | Visual Mental Representation | Metacognitive Processes | Total |
|---|---|---|---|---|---|---|
| Highest ($N = 5$) | 3.8 | 3.8 | 5.0 | 1.2 | 1.4 | 15.2 |
| Medium ($N = 6$) | 3.5 | 2.7 | 2.8 | 1.2 | 1.2 | 12.6 |
| Lowest ($N = 6$) | 1.5 | 3.0 | 1.5 | .2 | .8 | 7.5 |
| TOTAL ($N = 17$) | 8.8 | 9.5 | 9.3 | 2.6 | 3.4 | 35.3 |

Whereas the sort described previously was based on similarity in forecasting style, this sort was based on the researchers' subjective judgments of skill level. An initial sort produced six levels of forecasting skill, which were then combined to produce three categories of skill level due to the small number of forecasters within each category.[2]

Several general trends can be seen in Table 2.2. First, the forecasters at the higher skill levels (highest: $M = 15.2$; medium: $M = 12.6$) reported more frequent use of all of the cognitive elements than the forecasters in the lowest skill category ($M = 7.5$). Second, forecasters at all skill levels were more likely to describe the cognitive skills of Noticing Patterns ($M = 8.8$), Seeking Information ($M = 9.5$), and Meaning Making ($M = 9.3$) than they were to describe the use of Visual Mental Representations ($M = 2.6$) or Metacognitive Skills ($M = 3.4$). More important than the observed quantitative trends are the qualitative differences observed for these cognitive activities for forecasters of the different skill levels. We now describe these differences.

2 Although the two sorts focused on different criteria (forecasting style versus forecasting skill level), the results were very similar. All but one of the forecasters classified in the Scientist categories were categorized as having a high skill level (the exception was categorized as having a medium skill level). The forecasters classified in the Disengaged category were categorized as having low forecasting skill, and the remaining forecasters were all categorized as having a medium skill level. The results of the two sorts cannot be considered independent analyses because the same researchers conducted both sorts and the results of the first sort may have influenced the results of the second sort. We believed both sorts were necessary because it was theoretically possible that forecasting style would not map directly onto forecasting skill level.



*Noticing Patterns.* This category included instances in which the forecasters described an awareness of typical co-occurrences, deviations, and/or patterns in the weather information available to them. All but two of the forecasters (both from the lowest skill category) described some cognitive activity that we coded as noticing. The forecasters at the medium and highest skill levels described about twice as many instances of noticing as did forecasters at the lowest skill level. When we examined the content of the forecasters' comments for this category, we saw a great deal of similarity across the different skill levels. All of the groups talked about the importance of local effects, their knowledge of typical deviations caused by local terrain features, and seasonal variations at their particular location.

The nature of the subtle cues and patterns varied according to skill level. Many of the cues discussed by the most skilled forecasters could not be listed in a catalog. They include inconsistencies (which can only be spotted from a coherent sense of the overall pattern). They include violations of expectancies in which something happened that was not anticipated or something that was supposed to happen did not (both of which require a comprehensive mental simulation). They include interesting places on the map (which are interesting only in the context of the overall weather pattern). They include the ease with which inconsistencies are explained away, and how this explanation makes other facets of the weather settle into place or build up strain that leads to a suspicion that the obvious forecast may not be holding. They include acceleration cues – indicators showing rapid change that can signal dynamic shifts (which can only be spotted if these indicators are being studied over time). They include areas of instability that are often smoothed by computer graphics and can indicate turning points in the weather pattern (which can only be spotted by hand plotting the appropriate data once the forecaster knows what these are). For less skilled forecasters, their attention is on the centers of the masses (e.g., high-pressure areas) as they move. For the competent forecasters, their attention is on the edges, looking for areas of instability.

*Seeking Information.* This category included instances in which the forecaster described the collection of information to produce a forecast. All but one of the forecasters (from the lowest skill category) described some cognitive activity that we coded as Seeking Information. Thus, it appears that almost all forecasters are involved in data-seeking/gathering activities as part of their forecasting process. The one forecaster who did not report any cognitive activity that we coded as Seeking Information claimed that "he sometimes can skip analyzing the charts" while formulating his forecast. Even though the rest of the forecasters we interviewed all described actively seeking out weather information, the types of data they sought and how they used these data varied across the skill levels.

Forecasters at the highest skill level drew on a wider variety of technologies than forecasters at the lower skill levels. Their data-gathering strategies were fluid and flexible, guided by their sense of what was important for today's problem. The most highly skilled forecasters shared an information-seeking strategy in which they continuously shifted their perspectives. For example, these forecasters reported shifting their perspective from examining the satellite imagery down through the cloud layers and then examining the local surface observations and relating this information back to the satellite image. They were also keenly aware of the need to adjust observations and other products (e.g., radar, satellite images) in accord with the amount of time that has passed since these data were gathered and to consider the timing and location of data sources in order to recalibrate and adjust their understanding of what was going on. The most highly skilled forecasters considered computer models as an important source of data but not as the only source. They tended to examine *subelements* of the computer models for particular types of data they needed (such as wind speeds); they did not talk about the models in general terms of being "right" or "wrong."

Forecasters categorized as having a medium level of skill tended to rely heavily on the continuity principle. They placed a great deal of emphasis on looking upstream to obtain data. Their information seeking focused on trying to determine how fast the approaching weather pattern was moving toward them. It also appeared to be much more proceduralized; they typically followed the same steps every day (although these steps varied among the different forecasters). They placed greater emphasis on computer models than did the more highly skilled forecasters, and they referred to the models as intact representations rather than referring to subelements of the models. Many of the forecasters in the medium skill level category focused their information-seeking activity on determining whether or not "the model is right today."

Forecasters in the lowest skill category described their information seeking as highly proceduralized and static. There was no evidence that



their strategies varied in accord with local effects or weather events of the particular day. Some of these forecasters described a very limited data-gathering strategy that involved few data points and few data sources; other forecasters in this category described data-gathering strategies that involved looking at everything, although they did not attempt to link these data together in a meaningful way.

It appears that the information seeking of the more skilled forecasters was structured around their ability to detect problems and to identify the problem of the day (e.g., will this front move east or southeast?). The problem of the day was the instability or complication that required close attention, particularly if it could have major implications for severe weather. The problem of the day anchored their situation awareness. In addition, the act of information seeking seemed to provide ownership of the data, particularly if it required some sort of hand plotting.

*Meaning Making.* This category referred to a process by which forecasters organized or explained the information they had previously gathered in an attempt to make sense of it. All of the forecasters described some cognitive activity that we coded as Meaning Making, but the more highly skilled forecasters ($M = 5.0$) reported more incidents of meaning making than did the medium-skill ($M = 2.8$) and lower-skill forecasters ($M = 1.5$). An examination of the protocols indicated that the most highly skilled forecasters were more likely to describe the forecasting process in terms of causal reasoning than were forecasters in the other skill categories. All five of the forecasters in the highest skill category reported trying to understand the causal connections among various data elements and trying to construct an understanding of weather events. Their attempts to make meaning of the data involved using multiple weather events and multiple potential causes of these events; they anticipated interactions among complex sets of factors.

In contrast, the forecasters in the medium-skill category rarely talked about trying to determine the causes of the weather. Only two of the six forecasters in this group talked about causal connections. Most of this group's efforts at meaning making focused on determining how quickly the weather would change rather than trying to figure out how complex weather systems were likely to develop across the Earth's surface. Although all of the forecasters in the lowest skill category described some cognitive activity coded as Meaning Making, in general these remarks reflected very little depth of understanding of meteorology. For example, one forecaster's comment that "If the K Index is in the thirties, then there is a chance of thunderstorms" was coded as Meaning Making because the forecaster was going beyond Seeking Information and was making sense of the data. However, this remark does not reflect an in-depth understanding of the forces of the atmosphere; it is simply a rule that the forecaster had memorized.

We concluded that the skilled forecasters were not building up meaning from basic data elements. Rather, the act of interpretation also determined what counted as data. These forecasters have no standard "grain" for analyzing events. One forecaster described a case where he believed there was a second front. However, the computer-generated picture was plotted at the 500 millibar grain. Therefore, he hand plotted the data at the 200 millibar grain for the area he believed was critical, and he found the evidence for the second front. He would not have wanted an entire map at the 200 millibar grain.

*Visual Mental Representation.* This category included instances in which forecasters described the use of visual mental representations when they discussed their forecasting processes. If forecasters reported making a "picture" in their minds as they tried to understand the weather, we coded their representation as Visual Mental Representation. For example, one forecaster described his use of visual mental representation when he stated that he "needs to build a big picture of the atmosphere. The atmosphere is fluid; like a rock in the pond, there are ripple effects." Another forecaster described how he forms a picture of the weather: "I stack the upper levels of the atmosphere over the surface features and think about how these features will change the surface features and when these changes will occur."

Four of the five forecasters in the highest skill group reported using a visual mental representation, while the other highly skilled forecaster specifically stated that he did not form a visual representation of the weather. All but one of the forecasters in the medium-skill group described the use of visual mental representation; however, only one of the forecasters in the lowest skill group reported using a visual representation, and three of these forecasters failed to describe any type of mental representation.

Although there was no difference between the highly skilled and the medium-skilled forecasters in terms of the frequency with which they reported using visual mental models, an examination of their protocols indicated a qualitative difference in many of the instances that were reported. In general, the descriptions of the visual representations used by



the highly skilled forecasters included more use of vivid visual imagery. Their protocols were more likely to describe three-dimensional, dynamic representations, whereas the medium-skilled group were more likely to describe more static images. For example, one highly skilled forecaster stated that "The central analysis smoothes the fronts, how consistent they are, and how smoothly they move. In reality, the fronts wiggle as they move across the land. And it's at those whorls and wiggles that weather happens." In contrast, only one medium-skilled forecaster stated that he pictures in his mind "what's happening at each of the [atmospheric] levels and then he begins to stack the features on top of one another."

The type of visual imagery reported by the forecasters focused on three aspects of the weather. Some forecasters reported visualizing the atmosphere as a fluid and picturing atmospheric dynamics within that fluid space. Others reported visualizing the interface between the atmosphere and the surface of the earth (land and water). Forecasters also reported visualizing the dynamics of the weather systems as they move across the Earth's surface. It appears that these forecasters used visual imagery and mental simulations to explore weather phenomena that they could not experience in any direct physical way.

*Metacognitive Processes.* This category included instances in which forecasters described cognitive activities that were self-reflective. Approximately two-thirds of the forecasters described some cognitive activity that we coded as Metacognitive. Although there were no quantitative differences across the different categories of skill level, there were several interesting qualitative differences that we found in the protocols.

Protocol data from more highly skilled forecasters suggested that they consciously managed their cognitive approach to the forecasting task, thought about what worked best, and were capable of stepping away from their own cognitive processes and evaluating them. Several comments indicated strongly held opinions about the interplay between current weather technology and information management tasks. The more highly skilled forecasters indicated that data overload makes their job more difficult and that they had to develop methods for managing the information stream. For example, one forecaster said, "You have to analyze what you have and what you don't have. Then you have to figure out how you can fill in what you don't have."

The more highly skilled forecasters stated that from their point of view, current technology has made some of the cognitive elements of their job more difficult. They identified the "big picture" as being important but not well supported by current technology. One forecaster noted that he used to generate multiple observations over space and time. Because they were on paper, he could look across localities and time to get a sense of emerging patterns. He said:

> Now, you can get this data, but in single elements and separated by screens and menus. It's there, but not in front of you. The forecaster has to go and get the data, know how to put the big picture together, and hold various elements of it in his head. They have the detail but none of the big view. They have the data but have lost the ability to process and represent it.

Forecasters from the medium-level skill category also expressed an awareness of the importance of data management strategies, particularly in situations where time was short or data conflicted. How to set priorities for additional data seeking was noted as being an important skill. These types of comments were not found in the protocols of the lower-skilled forecasters.

### *Olympic Forecasters*

The results of the analyses just summarized increased our understanding of the characteristics of competent performers in the weather forecasting domain. However, because there were very few highly skilled USAF forecasters in our sample, we believed we needed to talk to some "true experts" before we could make our recommendations to the USAF. We learned that a team of expert NWS forecasters had been assembled to predict the weather for the 1996 Olympic games in Atlanta, and we arranged to conduct small-group interviews with most of these forecasters. Because these interviews were conducted in small groups, we did not obtain sufficiently detailed information on individual forecasters to allow us to include their data in the analyses described previously. However, the interviews we conducted with the Olympic forecasters influenced our thinking, so we provide a brief description of what we learned before presenting our conclusions.

The task of the Olympic forecasters was actually much more similar to the forecasting task faced by USAF forecasters, who have to make very specific predictions and warnings for an airfield, than the task faced by most NWS forecasters, who typically forecast by region and issue countywide warnings. For the Olympic Games, very specific warnings



had to be issued for the different venues for different sporting events (e.g., dew warnings for cycling events, wind warnings for divers). Like the USAF forecasters we interviewed, each of these forecasters seemed to have his or her own approach to the forecasting process. One forecaster described his process as follows: First, he looks at current observations (the surface data, the upper air, radar, and satellite images); then he looks back to see what was happening earlier in time to figure out what caused the current situation; then he tries to figure out what sequence of events will occur given the current situation; and only then does he look at the computer models. At this point, he tries to reconcile differences between his prediction and the computer models.

When asked to talk about the forecasting process, several Olympic forecasters mentioned that they used a *pattern recognition* process. In addition, almost all of them described their use of visual mental models in the forecasting process. Some forecasters described two-dimensional visual models that looked like the weather maps shown on TV, which have areas of high pressure and low pressure, with fronts marked. Other forecasters described four-dimensional (including time) models that allow them to construct an understanding of the weather over their geographic area of interest. For example, one forecaster described his visual model as follows:

> Then you get a picture in your mind and you can say "Okay, well, they've [thunderstorms] already developed in this area, and they're going to move in this direction over time, and that's going to have an effect on the clouds coming into Nashville, or which way the wind's coming from, or have an effect on temperatures." In my mind, I watch the problem whether it's going to evolve throughout the day.

Several of the Olympic forecasters mentioned that they suffered from information overload given all the new technology available to them. For example, one forecaster said, "Now, when you sit down on a shift, there is too much information. You cannot look at it all. You literally have to decide what you want to look at, what you think you need to look at, and use that information. There's too much data now."

Although we had only a limited amount of time to spend with these forecasters, most of them impressed us as true experts according to the levels of expertise described in Table 2.1. These forecasters all talked about identifying the "problem of the day" in order to focus their information-gathering activities. They seemed to have a fluid, flexible style that allowed them to quickly get up to speed in a new environment.

### *Characteristics of Competent Weather Forecasters*

After analyzing the information we collected by conducting observations and interviews with USAF and NWS forecasters, we concluded that the following characteristics differentiate competent forecasters from noncompetent forecasters. Competent forecasters are characterized as follows:

- They typically identify a specific weather problem of the day, which serves as an anchor for information seeking and interpretation.
- They form their mental model prior to seeing computerized forecasts.
- They look at the weather situation using a larger (more global) perspective and can switch easily between a global and a local perspective.
- They are flexible in their use of various tools and procedures.
- They have a mental representation that incorporates the dynamic causes underlying the current weather situation.
- They can use their mental representation to quickly provide whatever weather information they are asked for (a forecast, a pilot briefing).

In contrast, forecasters who do not display competency are characterized as follows:

- They rely too much on computer models.
- They use a fixed set of procedures to produce their forecast.
- They have a narrow focus and do not attempt to understand the relevant larger-scale weather features.
- They are reactive and end up "chasing the obs."[3]

### *The Development of Forecasting Expertise*

After we identified the characteristics of competent weather forecasters, we reviewed our data to determine what factors contribute to the

[3] *Chasing the obs* refers to a practice used by some forecasters in which they amend their current forecast only after the current observations prove the forecast to be wrong. For example, rather than predicting that winds may exceed the minimum required to issue a warning, the forecaster waits until the winds observed are over the minimum and then issues the warning.



development of forecasting expertise. We identified the following three factors as being important: formal training opportunities, on-the-job training opportunities, and opportunities for feedback on specific forecasts. We discuss each of these in turn.

*Formal Training.* A major difference between the USAF and NWS forecasters we interviewed was the amount of formal training they received in meteorology. NWS forecasters must have completed a bachelor's-level degree in meteorology before they can serve as intern forecasters in the NWS. In contrast, enlisted personnel in the USAF receive a total of 35 weeks of formal classroom training to prepare them to act as forecasters. It was clear to us that more extensive training in meteorology contributes significantly to a forecaster's ability to develop a detailed understanding of the current weather situation. In addition to obtaining an extensive knowledge base of the science of meteorology, forecasters who have completed college are more likely to have had practice using critical thinking skills to solve abstract problems and to have developed and practiced their formal communication skills.

*On-the-Job Training.* Many of the more highly skilled USAF forecasters we interviewed commented that they were fortunate to have had a mentor early in their career. That is, they were stationed at a location in which a senior-level person (an officer, senior enlisted person, or civilian) guided their on-the-job forecaster training. No matter how much formal scientific training forecasters receive prior to their first assignment, they need an extended period of on-the-job training in their new location. Based on our observations at USAF weather stations, it appears that there is no standard practice regarding on-the-job training. There are also very few senior-level people who have the expertise to share. In contrast, the NWS requires forecasters to serve as intern forecasters for at least 2 years before they issue forecasts on their own.

*Feedback Opportunities.* A fundamental law in the psychology of learning is that learners must have feedback on their performance if they are to improve their skills. Based on our observations and interviews at USAF weather stations, there is little opportunity for individual forecasters to obtain timely feedback on the accuracy of their forecasts, and there is even less opportunity to get feedback on the effectiveness of their forecasting process. This second point is an important distinction because it is possible for the forecaster to have done everything "right" but still to have produced an inaccurate forecast; conversely, the forecaster may have done everything "wrong" and still not "busted" the forecast. If forecasters are highly motivated, they can attempt to get feedback on their forecast by checking pilots' reports or examining observations from the period for which they produced their forecast. However, this did not occur routinely in the stations we visited.

When we visited the Olympic Weather Support Office, we were surprised to learn how little formal training the forecasters were given to learn how to use new technology to make precise predictions in an unfamiliar geographic region. How did they get up to speed so quickly? They were brought to Peachtree City, Georgia, about 12 months prior to the Olympics for 2 weeks of training. They had an additional 2 weeks of training immediately before the beginning of the Olympics that included site visits to the various locations for which they would be making forecasts. The site visits allowed them to identify local terrain features that might affect their forecasts. Two additional factors were also identified as being critical to their rapid learning. First, the forecasters got timely and detailed feedback on their forecasts for 2 weeks before they had to produce real forecasts. They developed forecasts each day for the specific venues but did not disseminate this information. They were able to get detailed feedback on the correctness of these predictions based on observations taken at each location. Second, because they worked in teams of four to nine forecasters per shift, they learned from critiques of each other's forecasts.

To summarize, we concluded that there are a number of factors that may work against the development of high levels of skill among USAF forecasters. We believe that abbreviated schoolhouse training, uneven on-the-job training programs, and feedback opportunities that are rare or missing altogether may all contribute to the extreme variability in skill levels we found among USAF forecasters. An additional factor that appears to bear directly on forecasters' performance is the type of technology made available to them. For a detailed discussion of the impact of technology on USAF weather forecaster performance, see Pliske et al. (1997). In brief, we concluded that USAF forecasters have become overly reliant on technology and insufficiently reflective and self-regulated. USAF forecasters have access to computer models developed by the NWS that allow them to produce a good enough forecast most of the time. They have advanced radar systems that alert them to times when they need to issue warnings. These technologies allow USAF forecasters to produce forecasts in the absence of a well-developed understanding



of basic principles of meteorology, or even of the technology itself. An important consequence of this overreliance on technology is that USAF forecasters are unable to function when that technological support is missing, as commonly occurs in tactical settings and frequently occurs at base weather stations when their technological tools malfunction.

## Conclusions

This chapter has described an exploratory investigation of the decision making of weather forecasters. We relied on qualitative data to form conclusions and hypotheses about the nature of competence in this domain. If we had used more rigorous methods, such as presenting a standardized weather scenario, we might have been able to conduct more objective analyses and obtained cleaner findings. We considered using standardized scenarios, but we rejected that strategy for two reasons. The first reason was a practical concern; our USAF sponsor needed this research project to be completed within 7 months. This constraint did not allow sufficient time for the enormous effort that would have been involved in assembling and pilot testing realistic scenario packages, including all of the different maps and readings from the large variety of available instruments. The second reason had to do with issues of representativeness and generalizability. Our initial interviews clearly indicated that the types of information utilized by particular forecasters varied greatly, depending on the nature of the weather problem, which in turn was strongly affected by their geographic location. We were told to make recommendations relevant to all USAF forecasters and not to focus on a specific weather problem (e.g., severe storms) or limit our study to a specific geographic region. We were concerned that if we focused our efforts on documenting the skills and knowledge underlying forecasts developed for one or two specific weather scenarios, the resulting recommendations would not address the most significant training and system design needs currently facing USAF forecasters.

Consequently, we chose to use actual incidents, and to accept the vagaries of the forecasters' memories and the lack of standardization across participants. This strategy allowed us to gain a broad understanding of what characterizes competence in the weather forecasting domain. We were then able to provide our sponsor with general recommendations for changes in current training programs and with specific recommendations about the types of future research that needs to be done in order to redesign the technological tools provided to USAF forecasters.

With our approach, we learned that there were no standard information-seeking strategies used by weather forecasters, but that the skilled performers could flexibly adapt their information seeking to their interpretation of the problem of the day. We learned about the importance of constructing an explanatory causal model and found that unskilled personnel can generate (mediocre) forecasts without having a causal model in mind. We learned how important ownership was for them – forming their own interpretation prior to seeing the forecast of another person or of a computer program.

In contrast to the earlier research on weather forecasters, which focused exclusively on highly skilled forecasters, our research examined forecasters with a wide variety of skill levels. This provided us with an opportunity to explore the characteristics of competent performance. Shanteau (1989) describes a theory of expert competence that assumes that competence depends on five components: the decision maker's competence in the domain in question, psychological traits of the decision maker, the cognitive skills of the decision maker, the strategies used by the decision maker, and the characteristics of the decision-making task. Shanteau goes on to describe the characteristics of tasks that should lead to good versus poor performance by experts. We would argue that the weather forecasting task has characteristics that could promote both good and poor performance. In some ways, forecasting the weather at a particular airfield is a highly repetitive task. A USAF forecaster has to predict ceiling and visibility for his or her airfield at least every 4 hours during a 12-hour shift. Furthermore, feedback can be made available to the forecaster (e.g., did the predicted visibility occur?). On the other hand, one of the important features of tasks that Shanteau claims leads to poor performance is the nature of the stimuli involved: Are they static or dynamic? In most cases, the stimuli in the weather forecasting task are highly dynamic: The atmosphere is constantly changing, sometimes in unpredictable ways. The dynamic nature of the weather forecasting task could hinder the development of competency.

One of the key observations we made in this research project was that USAF forecasters were operating with virtually no feedback on their performance. Although in most situations this information could be made readily accessible to these forecasters, it typically was not. The only time many of the forecasters we interviewed received feedback on their forecasts was when they had failed to issue a critical warning when appropriate. These feedback sessions were viewed more as a type of punishment than as a learning opportunity. Without the opportunity to learn from their experience, most of these forecasters seemed destined to stay at fairly low levels of proficiency rather than develop into competent performers.



The importance of feedback has been discussed extensively by researchers who have studied the development of expertise (Ericsson, 1996). For example, Ericsson, Krampe, and Tesch-Romer (1993) reviewed laboratory studies of learning and skill acquisition in order to identify training activities related to the development of expertise. They concluded that the most effective learning results from deliberate practice that involves a well-defined task at an appropriate level of difficulty, informative feedback, and opportunities for correcting errors. Repeated trials without informative feedback and without opportunities for correcting errors will not result in improved performance.

Glaser (1996) also discusses the necessary conditions for the development of expertise. He describes the importance of self-regulation in deliberate practice. According to Glaser, individual performers must eventually develop the ability to structure their own learning situations and provide their own performance feedback. Our results generally support his claim. As discussed previously in this chapter, the more highly skilled weather forecasters in our sample were more likely to report using metacognitive strategies in their forecasting process that involved the evaluation of their own forecasting processes. These forecasters were also more likely to report seeking out feedback on the correctness of their forecasts. Metacognition-self is a critical part of their expertise.

In addition to describing metacognitive activity involving self-monitoring, forecasters in our study also described metacognitive activity reflecting their ability to handle the high degree of uncertainty that is inherent in the forecasting task. Several of the forecasters from the medium-level skill group indicated that they had difficulty handling the uncertainty in the forecasting task. Although several forecasters mentioned their tendency to second-guess themselves, there was no clear consensus on whether second-guessing was a positive strategy. On the one hand, these forecasters appeared to recognize that seeking verification or disconfirmation of their forecasts can be an important check. For example, one forecaster noted that "You have to be willing to second-guess yourself. . . . [Novices] need to have checkpoints for their forecasts so they can revise, check, and verify." But another forecaster acknowledged that second-guessing often led him astray. "My first belief or forecast reasoning is usually correct. It's when I begin to second-guess myself that I miss them." We do not know what accounts for this variation in viewpoint across forecasters from the medium-level skill category, but we note that it was absent from the more highly skilled group, who did not mention second-guessing at all. Furthermore, we found no evidence that lower-level skill forecasters experience difficulty dealing with uncertainty, perhaps because their approach to forecasting avoids the issue altogether. Several forecasters in this group mentioned seeking disconfirming information or generating alternate scenarios. For example, "[We] have to be flexible, let [our] imaginations wander over different potential, hypothetical scenarios so that [we] can think of what might or could happen." But none of these forecasters offered information about how they handle situations when disconfirming information or puzzling scenarios occur. Perhaps most striking in the protocol data from the lower-skilled forecasters was the evidence of a "quick size-up" strategy, which allowed them to move quickly from initial diagnosis to forecast. They were more likely to skim the surface of the information stream, go with their first impression, and be satisfied with it. For example, one of these forecasters asserted, "[I] really trust [my] first impression of the weather and don't really try to figure out where I might be wrong."

How weather forecasters at various skill levels deal with uncertainty is an important topic for future research. In addition, research using more scientifically rigorous methods is needed to validate and extend the results of this exploratory study. However, based on our findings, we were able to make a number of recommendations to the USAF as to how they could improve their weather forecasting performance. Many of these recommendations addressed training issues, whereas others addressed needed changes to the existing interfaces for the technical tools currently used by USAF forecasters. Our USAF sponsors are currently using our findings as they attempt to reengineer Air Force Weather.

### References

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Crandall, B., & Getchell-Reiter, K. (1993). Critical decision method: A technique for eliciting concrete assessment indicators from the intuition of NICU nurses. *Advances in Nursing Sciences, 16*(1), 42–51.

Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. New York: Harper & Row.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuitive expertise in the era of the computer*. New York: Free Press.

Ericsson, K. A. (Ed.). (1996). *The road to excellence*. Mahwah, NJ: Erlbaum.



Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406.

Glaser, R. (1996). Changing the agency for learning: Acquiring expert performance. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 303–311). Mahwah, NJ: Erlbaum.

Hoffman, R. R. (1991). Human factors psychology in the support of forecasting: The design of advanced meteorological workstations. *Weather and Forecasting, 6*, 98–110.

Hoffman, R. R., Crandall, B. W., & Shadbolt, N. R. (1998). A case study in cognitive task analysis methodology: The Critical Decision Method for the elicitation of expert knowledge. *Human Factors, 40*, 254–276.

Kaempf, G. L., Klein, G. A., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex command-and-control environments. *Human Factors, 38*(2), 220–231.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. *Proceedings of the Human Factors Society 30th Annual Meeting, 1*, 576–580.

Klein, G. A., & Hoffman, R. (1993). Seeing the invisible: Perceptual/cognitive aspects of expertise. In M. Rabinowitz (Ed.), *Cognitive science foundations of instruction* (pp. 203–226). Hillsdale, NJ: Erlbaum.

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. E. (Eds.). (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Klinger, D. W., Andriole, S. J., Militello, L. G., Adelman, L., Klein, G., & Gomes, M. E. (1993). *Designing for performance: A cognitive systems engineering approach to modifying an AWACS human–computer interface* (AL/CF-TR-1993-0093 for the Department of the Air Force, Armstrong Laboratory, Air Force Materiel Command, Wright-Patterson AFB, OH). Yellow Springs, OH: Klein Associates, Inc.

Militello, L. G., & Hutton, R. J. B. (1998). Applied Cognitive Task Analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 42*(11), 1618–1641.

Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 201–220). Mahwah, NJ: Erlbaum.

Mosier, K. L., Skitka, L. J., & Heers, S. T. (1996). Automation bias, accountability, and verification behaviors. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 204–208).

Murphy, A. H., & Winkler, R. L. (1977) Can weather forecasters formulate reliable forecasts of precipitation and temperature? *National Weather Digest, 2*, 2–9.

Orasanu, J., & Fischer, U. (1997). Finding decisions in natural environments: The view from the cockpit. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 343–358). Mahwah, NJ: Erlbaum.

Pliske, R., Klinger, D., Hutton, R., Crandall, B., Knight, B., & Klein, G. (1997). *Understanding skilled weather forecasting: Implications for training and the design of forecasting tools* (Technical Report No. 1L/HR-CR-1997-0003 for

Armstrong Laboratory: Brooks AFB, TX). Fairborn, OH: Klein Associates, Inc.

Randel, J. M., Pugh, H. L., & Reed, S. K. (1996). Methods for analyzing cognitive skills for a technical task. *International Journal of Human-Computer Studies, 45*, 579–597.

Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica, 68*, 203–215.

Shanteau, J. (1989) Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252–266.

Stewart, T. R., Heideman, K. F., Moninger, W. R., & Reagan-Cirincione, P. (1992). Effects of improved information on the components of skill in weather forecasting. *Organizational Behavior and Human Decision Processes, 53*, 107–134.

Stewart, T. R., & Lusk, C. M. (1994). Seven components of judgmental forecasting skill: Implications for research and the improvement of forecasts. *Journal of Forecasting, 13*, 579–599.

Zsambok, C. E., & Klein, G. (Eds.). (1997). *Naturalistic decision making*. Mahwah, NJ: Erlbaum.

# Part II

# Metacognition-Others