Technical Report 1200

FOCUS: A Model of Sensemaking

Winston R. Sieck, Gary Klein, Deborah A. Peluso,
Jennifer L. Smith, and Danyele Harris-Thompson
Klein Associates, Inc.

May 2007

United States Army Research Institute
for the Behavioral and Social Sciences

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

A DIRECTORATE OF THE DEPARTMENT OF THE ARMY

Deputy Chief Of Staff, G1

Authorized and approved for distribution:
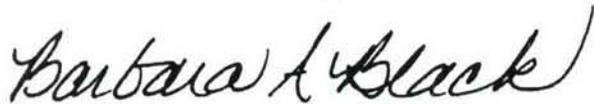
Barbara A. Black

BARBARA A. BLACK
Acting Technical Director

Michelle Sams

MICHELLE SAMS
Director

Research accomplished under contract
for the Department of the Army

Klein Associates Inc.

Technical reviews by

David P. Costanza, The George Washington University
Joseph Psotka, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

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

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
May 2007	Final	June 2001 – June 2004

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
	1435-01-01-CT-31161
FOCUS: A Model of Sensemaking	5b. PROGRAM ELEMENT NUMBER
	611102A

6. AUTHOR(S)	5c. PROJECT NUMBER
Winston R. Sieck, Gary A. Klein, Deborah A. Pelusa, Jennifer L. Smith and Danyele Harris-Thompson (Klein Associates Inc.)	B74F
	5d. TASK NUMBER
	4901
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Klein Associates Inc. 1750 Commerce Center Blvd. North Fairborn, OH 45324-6362	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. MONITOR ACRONYM
U. S. Army Research Institute for the Behavioral & Social Sciences	ARI
2511 Jefferson Davis Highway Arlington, VA 22202-3926	11. MONITOR REPORT NUMBER
	Technical Report 1200

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Contracting Officer's Representative and Subject Matter POC: Dr. Paul A. Gade

14. ABSTRACT *(Maximum 200 words)*:
Sensemaking is a relatively new concept that has largely been associated with Weick (1995) and his work in organizational behavior. Sensemaking refers to the set of processes involved in trying to improve one's understanding of a situation, often in response to surprise. The primary purpose of the current project was to unpack and develop the concept of sensemaking, principally by developing and testing a cognitive model of the processes involved. The resulting Data/Frame model posits a highly interactive relationship between data inputs and mental representations or "frames" for interpreting data. The Data/Frame model also proposes six key sensemaking activities for handling frames in light of (anomalous) data: Elaborating, Questioning, Comparing, Preserving, Re-framing, and Seeking. A secondary aim was to provide recommendations for training and other applications of the model that would be of direct benefit to the warfighter. To that end, several specific links to applied issues in domains such as information operations, intelligence analysis and combat systems design for UAV control have been developed and pursued. At this juncture, the concept of sensemaking and the Data/Frame model appear to be supported by the data, and also quite useful for military applications.

15. SUBJECT TERMS
Sensemaking, situation assessment, situation awareness, Cognitive Task Analysis, sensemaking training, scenario-based training, mental models, fixation

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			Ellen Kinzer Technical Publication Specialist 703-602-8047
Unclassified	Unclassified	Unclassified	Unlimited	47	

Technical Report 1200

FOCUS: A Model of Sensemaking

**Winston R. Sieck, Gary Klein, Deborah A. Peluso,
Jennifer L. Smith and Danyele Harris-Thompson**
Klein Associates, Inc.

**Research and Advanced Concepts Office
Paul A. Gade, Chief**

May 2007

ACKNOWLEDGMENTS

FOCUS: A MODEL OF SENSEMAKING

EXECUTIVE SUMMARY

Research Requirement:

Army leaders are often placed in situations that are either unfamiliar or filled with uncertainty. In order to be successful, Army leaders must be able to quickly and efficiently make sense out of rapidly changing situations. Without an understanding of the situation, they cannot take appropriate actions. Developing this understanding is difficult, especially when one is faced with conflicted or limited information. Sensemaking is the process by which people develop their understanding in the face of surprise information. Experts have demonstrated the ability to do this with almost uncanny skill as they put together pieces of a puzzle that novices do not even see. Research on the topic has helped define situation awareness and its different levels. However, we still do not understand the process by which experts make sense of situations and how this sensemaking ability develops over time. Understanding sensemaking is critical if the Army wants to train and develop this skill in its future leaders. The problem is that no one fully understands the process of sensemaking. The purpose of this project is to examine the process of sensemaking, how it develops, and ways to improve sensemaking through training interventions.

Procedure:

This research was conducted over the course of three years and used different procedures across the years. In Year 1, experienced and novice Information Operations (IO) officers completed a series of scenarios designed to challenge the participants' sensemaking. Participant comments during the scenarios were categorized and coded based on the types of inferences, speculations, and explanations made. In addition, researchers collected real-life incidents of navigational sensemaking in which participants became lost while driving and had to reorient themselves. In Year 2, new Cognitive Task Analysis data collection methods were developed and used to further test the model. Cognitive Task Analysis is a series of methods and tools used to gain in-depth access to the mental processes that underlie performance of tasks. The researchers used these methods during interviews with IO officers to elicit and explore incidents in which the participants faced challenges to sensemaking. In Year 3, researchers recoded and reanalyzed the IO officer comments from Year 1 to further examine expert/novice differences. Additional interviews were conducted with IO officers at 1st IO Command in which participants responded to a series of scenarios. These data were coded and analyzed for specific expert/novice differences.

Findings:

This research effort has produced several significant findings about the nature of sensemaking. First, we have defined sensemaking as the process of fitting data into a frame, and fitting a frame around the data. People will try to make sense of data inputs they receive by finding or constructing a story to account for the data. At the same time, their repertoire of stories will affect which data elements they consider and how they will interpret these data. Thus, the frame

and the data work in concert to generate an explanation. Based on the results, we have differentiated and described six activities, or building blocks, of sensemaking: elaborating the frame, questioning the frame, preserving the frame, comparing frames, seeking a frame, and reframing. Our research also suggests that developing a comprehensive mental model for a complex, open system is unrealistic. Instead most people, and even most experts, rely on fragments of local cause-effect connections, rules of thumb, patterns of cues, and other linkages and relationships between cues and information to guide the sensemaking process (and indeed other high-level cognitive processes). We believe that a set of fragmentary mental models contribute to the frame that is constructed by the sensemaker, therefore guiding the selection and interpretation of data. We found that IO experts exhibit a stronger delineation between sensemaking and decision-making processes. Experts attempt to first clearly understand the situation they face, and then proceed to formulate courses of action and other decision-making activities. Furthermore, IO experts are much more likely than others to formulate high-quality questions geared towards understanding anomalous evidence. The kinds of questions IO experts asked varied in terms of generality.

Utilization and Dissemination of Findings:

Results of this research have been the subject of many invited briefings to a wide range of government and academic audiences. An understanding of sensemaking is relevant to Army efforts to design various types of decision support systems. By describing how humans actually use information to develop and revise interpretations of a situation, we can develop command and control and other technologies that work in concert with the human and support his/her natural process for filtering information and building and maintaining situation awareness. Our findings-to-date on the nature of the sensemaking process offer initial guidance on several important issues in the field of decision support, such as the effect of information rate on skilled performance and possible evaluation metrics for systems intended to bolster sensemaking activities. Defining the process of sensemaking also appears to be relevant to Army training needs. Our findings illustrate areas in which sensemaking can break down and even fail. Training programs can be structured to enable practice and feedback in key sensemaking activities. Our findings also highlight areas of expert/novice differences. Training scenarios can be structured to deliberately build sensemaking expertise in non-experts through the development of mental models, fragmentary mental models, and anticipatory thinking skills.

FOCUS: A MODEL OF SENSEMAKING

CONTENTS

LIST OF TABLES

LIST OF FIGURES

OVERVIEW

This report describes Klein Associates' research activities and progress over a three-year period (June 2001 to June 2004) to examine the phenomenon of "sensemaking," a concept initially formulated by Weick (1995), but not developed from a cognitive perspective. This project is sponsored by the U.S. Army Research Institute as a basic research activity to unpack and develop the sensemaking concept. The focus in Year 1 was on testing an initial model of sensemaking, the Framework for Observing, Comprehending, and Understanding Situations (FOCUS). As described below, the initial FOCUS model was found to miss important elements of the process, and a new Data/Frame model was specified that detailed four distinct sensemaking activities or functions.

In Year 2, new cognitive task analysis (CTA) data collection methods were developed and used to further test the model. Based on these data, the Data/Frame model was further refined and extended to include six distinct sensemaking functions. In addition, a number of assertions regarding the nature of sensemaking were proposed.

Efforts in Year 3 were devoted to accomplishing two aims. First, tests of critical assumptions of the model were sought. Second, a deeper understanding of expert/novice differences was sought in order to further expand on the training recommendations offered in Year 2. The report is organized as follows: First, the research and findings from Years 1 and 2 are summarized, including the state of the model as developed to that point in time. Then, three new experiments conducted in Year 3 are described in greater detail.

YEAR 1 SUMMARY

During the first year of the FOCUS project, Klein Associates' researchers conducted an exploratory experiment of sensemaking (Klein, Phillips, Battaglia, Wiggins, & Ross, 2002). The Year 1 experiment consisted of three primary tasks. Task 1 was an exploration of expert/novice differences in the sensemaking of Information Operations (IO) specialists at the Land Information Warfare Activity (LIWA), now reorganized under the 1st Information Operations Command. Task 2 was an analysis of real-life incidents related to automobile navigation in which understanding had broken down. Using the results of Tasks 1 and 2, Task 3 was the development of an initial Data/Frame Model of Sensemaking.

Task 1–Expert/Novice Differences in Sensemaking

For Task 1, Klein Associates' researchers developed three different scenarios based on actual incidents from peacekeeping operations in Bosnia. Each scenario consisted of a string of 11 to 17 situation reports (SitReps). Participants read these SitReps one at a time, thereby enabling the researchers to observe how the participants' understanding of the situations developed and changed as more information was provided. The SitReps were designed to expose the participant to multiple themes or storylines simultaneously. In addition, the SitReps also contained messages that were noise and did not pertain to the major themes of the scenario. The multiple themes and noise caused initial uncertainty about the nature of the problem that the U.S.

peacekeeping forces were facing. As the scenario progressed, more information was presented to reduce the uncertainty. The issues were much clearer by the end of the scenario.

Twelve IO specialists from the U.S. Army's LIWA participated in this experiment. Six were experts and six were novices. Each participant was interviewed individually for two hours by one to three researchers. The participants were instructed to read the SitReps and to comment on what was going on in the situation, especially issues that required the attention of higher headquarters. In addition, participants were provided with maps to familiarize them with the villages and refugee camps mentioned in the scenarios. The researchers asked the participants if they had any observations or interpretations after each SitRep. Researchers also interrupted the protocol at places in the scenario where major themes and issues were developing. They used this opportunity to elicit more detail from the participants surrounding key events.

An effort was made to counterbalance the scenarios across levels of experience. However, this was complicated by the difference in the amount of time it took for experienced versus novice participants to complete the scenario. In general, experienced participants had more to say and were only able to complete one scenario in two hours, while the novices were able to complete two scenarios in the same time frame.

Scenario 1, *Preparing for Winter*, was about Bosniacs living in refugee camps and the problems they faced. The major theme was that an unresolved sewage problem in one of the camps led to a cholera outbreak, and that the refugees leaving this camp for another resulted in the spread of the disease to a second camp. Four experts and four novices completed Scenario 1. Scenario 2, *Rebuilding the Schools*, was about an attempt to integrate Serb and Bosniac students in a high school in Brcko. Two experts and three novices completed Scenario 2. Scenario 3, *Activity in the Republica Srpska*, was about an influx of Bosniac refugees to an area that was populated by Serbs. The problems between the two groups resulted in the death of several Serbs, and some blamed this on the U.S. peacekeepers for siding with the Bosniacs. One expert and three novices completed this scenario.

The data were analyzed by categorizing participant comments using a coding scheme that was designed to explore the types of inferences, speculations, and explanations the participants made as they read the SitReps. The data were analyzed by two coders. First, the coders developed and defined the categories. Second, the coders independently coded portions of the data and then met to determine the level of agreement between their analyses. When the coders agreed on a code, they continued to the next code. When they did not agree, the coders reached consensus by discussing the rationale behind each coding choice.

The results of this analysis showed that all of the participants were actively seeking connections between messages, including messages designed to be noise. However, the experts' experience and richer mental models enabled them to generate more connections than the novices did. Both groups tended to use the same sensemaking strategies, including inferring cause/effect connections, inferring causes, inferring effects, inferring causes from effects, awareness of multiple causes, and identifying instances when the cause resulted in an unexpected effect. An implication of these results is that little would be gained by trying to teach novices to

think like experts, because they are already trying to make the appropriate connections between data.

A few expert/novice differences were identified by the coders. There were instances in which novices identified multiple causes and effects but were uncertain on how to link them; this did not occur with the experts. Experts identified associations (in which a message made the participant think of something that he or she needed to keep in mind), opportunities to take IO actions, and action items (including anticipation of problems) significantly more often than the novices did. The experts' comments displayed a deeper, richer understanding of the situation and a greater sensitivity to the context. The two groups generated inferences in the same way, but the nature of these inferences was different. Novices were less certain about the relevance of messages and were more likely to treat important signals as noise. Experts were more likely to question the quality of the data, while the novices tended to take them at face value. Throughout this experiment, the experts displayed a "functional understanding" of the situation. That is, their responses were framed in terms of IO actions and means to advance SFOR agenda. The novices, however, showed an "abstract understanding"—they viewed the situations as a connected series of events that needed to be untangled, but without a clear consideration of ways to influence the situation.

Task 2–Navigational Sensemaking

For Task 2, Klein Associates' researchers gathered information and ideas from a variety of data sources, including incidents elicited for past projects and data that were collected for a parallel experiment of sensemaking. In addition, an informal pilot experiment was conducted with a self-selected group of researchers from Klein Associates. Participants submitted a total of 14 incidents in which they were lost while driving. They were asked to describe how they became lost, how they realized they were lost, and how they "got found" again. The ideas and hypotheses in these incidents were examined individually. These incidents contributed to the development of the model of sensemaking, particularly concerning the corruption and recovery of sensemaking.

Task 3–Development of the Data/Frame Model of Sensemaking

Task 3 was the development of the Data/Frame Model of Sensemaking based on Year 1 research. Revisions and improvements were made on this model in Year 2 as theories about sensemaking changed. The following is a description of sensemaking as it was described in Year 1 of this program.

In Year 1, Klein Associates' researchers developed a list of assertions about sensemaking, based on the Data/Frame Model:

1. The definition of sensemaking is fitting data into a frame.
2. Sensemaking is a process, not a state.
3. Sensemaking is a deliberate and conscious process.
4. The situation is constructed out of a noisy background.

5. There is no primitive data (cues are not predefined but depend on the conceptual events).
6. There are different sensemaking functions.
7. Sensemaking requires a reciprocal connection between the data and the frame.
8. Sensemaking requires judging the credibility of the data, or the story of the data itself.
9. We can describe a set of triggers for sensemaking.
10. Sensemaking can be an intellectual exercise or serve a functional purpose.
11. Experts do not differ from novices in their sensemaking strategies.
12. Experts use fragmentary mental models.
13. Sensemaking has four different facets.

Sensemaking can be divided into four different facets: 1) Representing the Situation, 2) Elaborating the Frame, 3) Corrupting the Frame, and 4) Recovering From a Corrupted Understanding. Each of these facets is a distinct form of sensemaking. These facets are not presented in a definite sequence that all sensemaking will follow. For example, sensemaking does not always result in corrupted frames. Sensemaking is often successful at the Elaborating the Frame facet. Other times, the frame becomes so corrupted that recovery is not possible.

In *Representing the Situation*, Facet 1 of the Data/Frame Model, the individual prepares him or herself to construct meaning in order to make sense of the situation. Facet 1 consists of five components: 1) extracting cues from the signal stream (selecting and discarding data); 2) selecting a frame; 3) the noise in the signal stream (the frame for the situation is used to differentiate the signals from the noise); 4) using categories within the frame; and 5) evaluating cues.

In *Elaborating the Frame*, Facet 2, the individual deepens the initial representation of the situation. This elaborated frame is ready to test against real-world situations. Facet 2 consists of four components: 1) seeking connections between cues; 2) conducting quasi-logical operations; 3) utilizing anchors (certain cues are used as a foundation for constructing the frame); and 4) constructing the story/script/map.

In *Corrupting the Frame*, Facet 3, the individual's sensemaking has broken down and the sensemaking processes may be increasing uncertainty. Facet 3 consists of eight components: 1) managing corrupted cues and anchors; 2) spreading corruption; 3) bending the map (explaining away discrepancies); 4) levels of being lost (amount of uncertainty); 5) accepting the loss of adequate sensemaking; 6) knowledge shields (arguments used to explain away discrepant new data); 7) fixating on a hypothesis; and 8) defending against knowledge shields.

In *Recovering From a Corrupted Understanding*, Facet 4, once people recognize the flaws in their understanding, they can recover and make sense of the situation. Recovering involves making a shift in the frame used to understand the situation. Facet 4 consists of three components: 1) using recovery strategies; 2) reframing; and 3) benefiting from corruption and recovery (the development of richer mental models).

A full report of Year 1 research activities is documented in Klein et al., 2002.

YEAR 2 SUMMARY

During Year 1 an initial model of sensemaking was developed based on simulation interviews conducted with IO specialists at the LIWA. The analysis of these data led to a Data/Frame description of sensemaking. In the second year of this effort, researchers examined and modified the Data/Frame account of the sensemaking process to produce a version that is more comprehensive and testable.[1]

In Year 2 we conducted two primary activities. First, we developed sensemaking training scenarios for the IO novices and provided them to the LIWA field office. The simulations developed and used for data collection in Year 1 were revised to incorporate edits suggested by a LIWA subject-matter expert (SME). They were then annotated with the CTA data from LIWA experts to serve as a comparison case with which trainees could assess the accuracy and thoroughness of their sensemaking. The training scenarios were then pilot tested by presenting them to two IO novices to determine the usefulness of the experts' analysis. The finalized scenarios with annotations were provided to LIWA to be utilized in their next round of IO training. In addition to developing the training scenarios, the majority of our Year 2 effort involved conducting various activities to investigate the quality of our Data/Frame model and to refine it accordingly. We reviewed, collected, and analyzed additional incident data. To support the new directions of our research, we conducted additional reviews of sensemaking literature and elicited feedback about the Data/Frame Model from leading researchers in the community.

Our primary goals for the model of sensemaking are to describe the cognitive processes people follow in order to make sense of a situation, the ways in which sensemaking can break down and result in failure, and the implications for improving individuals' sensemaking through training. We made great strides in Year 2 by developing a broader framework of different sensemaking activities. The data collection and data review activities for Year 2 will be discussed briefly.

Data Collection

Year 2 data collection served two purposes: to further our understanding of the nature of sensemaking in real-world ambiguous situations, and to refine our Data/Frame Model of Sensemaking (Klein et al., 2002).

Two data collection trips were made to LIWA. Two interviews were conducted during the first visit and five were conducted during the second visit. All participants were designated by the office as experienced in the domain of IO and were military officers or former officers currently working for LIWA as civilians. These interviews were conducted using three knowledge elicitation techniques. Each technique provided the ability to elicit rich sensemaking incidents and explore, in depth, the cognition underlying this process.

---

[1] For a complete description of the Year 2 effort, see Klein et al., 2003.

Method 1: Sensemaking Knowledge Audit

The first method, the Sensemaking Knowledge Audit, consisted of a series of questions designed to help SMEs articulate their experience. Each question focused on the distinction between the performance of experts and the performance of novices. For this data collection, we wanted to elicit incidents that challenged the participants' abilities to makes sense out of an ambiguous situation. Questions were crafted around phases of sensemaking expressed in Year 1 and around expert/novice differences discovered during Year 1 data analysis (Klein et al., 2002). The resulting probes addressed selecting and elaborating the frame, corrupting and recovering the frame, inferences, associations vs. noise, and spotting opportunities. These probes were used with all seven SMEs. The Critical Decision Method was used to elicit the rest of the incident and probe the process the participant used in making sense of the situation.

Method 2: Critical Decision Method (CDM)

The second interview technique, the Critical Decision Method (CDM), was originally created to unpack the cognitive elements of a task in the context of a specific incident (Klein, Calderwood, & MacGregor, 1989). It is structured around four interview phases, called sweeps, that successively examine an incident in greater detail: 1) identification of a complex incident showing some cognitive skill and expertise, 2) the creation of an incident timeline to show the sequence of events, 3) deepening on strategies used for managing the decision points represented on the timeline, and 4) probing with what-if questions that explore potential expert/novice differences (Klein et al., 1989).

To support this experiment, the CDM interview was tailored to probe more specifically on the cognition underlying sensemaking. This was done by structuring the probes in the third sweep around the SME's understanding of the developing situation, instead of around decision points. The focus of these probes included: pre-incident goals and mission, initiation of the sensemaking process, interpretation of information throughout the evolving incident, and discarding data. This method was used in the five interviews conducted during the second trip to LIWA.

Method 3: Sensemaking Interview (SI)

The Sensemaking Interview (SI) is another variation of the CDM because it focuses on specific incidents from the interviewee's experience and is based on several sweeps through the same incident. It also utilizes a recounting of the events on a timeline. However, instead of focusing on drawing out the critical judgments, cues, and strategies that support decision making, the SI unpacks how situation awareness is elaborated, shifted or discarded and replaced by a new understanding (Ross, Thunholm et al., 2003).

While other methods focus on incidents within the interviewee's recent memory, the SI begins with an incident that just occurred. The opening query for each incident is: "Considering all that has happened in the past four hours, what were the top challenges or toughest issues you had to figure out? We are looking for a short list of things of which you need to make sense."

The research team utilized this technique during interviews with participants of a week long battle command simulation exercise.

Two of the three incidents collected using this method were considered rich examples of sensemaking. These two, along with the incidents collected using the other two methods, were incorporated into the data review.

Archival Data Review

The purpose of the data review was to determine how descriptive the Data/Frame Model is of how people make sense of ambiguous real-world situations across several domains. The team selected incidents to review from the following sources:

- A study of the decision making of fire ground commanders (Klein, Calderwood, & Clinton-Cirocco, 1988)

- An examination of the expertise of Neonatal Intensive Care Unit (NICU) nurses (Crandall & Gamblian, 1991)

- A study of decision strategies of Naval Command and Control, AEGIS-class Combat Information Center teams (Kaempf, Wolf, Thordsen, & Klein, 1992)

- A study of sensemaking of small unit Army infantry leaders in the context of a counter-reconnaissance ambush (Phillips, Baxter, & Harris, 2003)

- A project identifying facilitation and instructional strategies for teaching tactical thinking skills to Army Captains (Ross, Battaglia, Hutton, & Crandall, 2003)

- The format of our data varied based on the nature of the project. Some sets of data had concise incident or case accounts, while others consisted of interview notes and transcripts, transcripts of instructional sessions, or facilitator descriptions of trainees' responses to training scenarios. For each of these data sets, a researcher familiar with the original work created summaries of the complete incidents, interviews, and sessions.

Data review and model development took place iteratively. Each member of the project team reviewed one or two of the data sets in full (a "set" consisted of data from a single domain or project). All members of the project team reviewed the data summaries from all the incidents. Each team member was tasked to identify aspects of sensemaking that were consistent with, inconsistent with, or not well represented by the Data/Frame Model. Insights about sensemaking that emerged from this individual review were articulated and discussed in a team forum and areas were identified for further exploration.

We revisited the incidents several times to clarify which phases of the Data/Frame Model were useful for coding key phenomena within the incidents, and to see whether additional phases emerged from examination of the data. We also looked to see if different overall phases or activities for sensemaking emerged, and attempted to classify and represent those.

Literature Review

One of the initial Year 2 activities was to gather critiques of our initial Data/Frame Model. To achieve this, we reviewed relevant literature and requested feedback from researchers and practitioners in sensemaking. The goal of this literature review was to identify linkages and disconnects between the Data/Frame Model and existing research on sensemaking. This review allowed us to better understand the boundary conditions of deliberate sensemaking and how and where conscious sensemaking activities fit into the larger picture of cognition (for a description of this review, see Klein et al., 2003).

Year 2 Findings and Implications

Our work during Year 2 allowed us to expand the set of findings and conclusions about the nature of sensemaking. These key findings were:

- *Data/frame interaction, rather than information processing stages*: Sensemaking is a process of framing, or fitting data into a frame, which helps us filter and interpret the data while testing and improving the frame. The purpose of the frame is to a) define the elements of the situation, b) describe the significance of these elements, c) describe their relationship to each other, d) filter out irrelevant messages while highlighting relevant messages, and e) reflect the context of the situation, not just the data.

- *Construction of data*: The Data/Frame Model states that data elements are not perfect representations of the world, but are constructed. They are sampled from the available information in the environment and defined in terms of available frames.

- *The use of anchors*: We assert that the initial one or two key data elements serve as anchors. These initial anchors affect the frame that is adopted, and that frame guides information seeking.

- *Sensemaking cycles*: We differentiated six activities: elaborating the frame, questioning the frame, preserving the frame, comparing frames, seeking a frame, and reframing. In relating the current work to earlier concepts by Piaget, we proposed ways in which cycles of assimilation and accommodation encompass the six activities.

- *Possibility theory*: Klein et al. (2002) referred to quasi-logical reasoning, which we extended into the notion of "possibility theory." This theory describes those sensemaking activities that rely on identifying possibilities, plausible inferences, and justifiable guesses.

- *Fragmentary mental models rather than comprehensive mental models*: The concept of fragmentary mental models (FMMs) was introduced in Year 1. In Year 2, we further speculated that the FMMs are called to mind to create a "just-in-time" mental model of a situation as the situation warrants.

8

- *Functional sensemaking, not just abstract understanding*: Sensemaking is about what an expert is able to do in a situation, not merely what is going on in the situation.

- *Expert/novice differences are based on larger repertoire of frames and FMMs*: With experience, people develop larger repertoires of FMMs. They have a better understanding of how to link these to their current goals and they develop scripts for action rather than route knowledge. Experts also appear to have more routines or ways of accomplishing things. This widens their functional sensemaking because they have more ways to frame the leverage points in a situation.

- *Initiation of sensemaking*: The general sensemaking process, involving patterns of use of the six activities described above, is usually triggered by some anomaly or uncertainty that contradicts our typical expectations or interpretations.

The application of these Year 2 findings to our Year 3 research will be discussed in the next section.

YEAR 3 EFFORT: CHARACTERISTICS OF EXPERT SENSEMAKING

In Year 2, substantial progress was made towards developing and refining the Data/ Frame Model of Sensemaking. Furthermore, several general differences between aspects of expert and novice sensemaking were uncovered, providing broad directions for potential training. A primary finding was that experts have better-developed mental models to aid their sensemaking, but that experts and novices both follow the same general processes as described in the Data/Frame Model. In Year 3, we attempted to further unpack these ideas and gain a deeper understanding regarding the nature and generality of differences between expert and novice sensemaking. Doing so was considered a critical step towards the development of specific training recommendations.

The team continued these investigations into the IO domain. This was done in order to leverage some of our previous studies, and to push closer towards providing tangible training suggestions for people performing a specific job in a specific context. Before proceeding, for clarity, we need to define what we mean by a mental model, and how mental models relate to other kinds of knowledge.

Knowledge Typology

We can discriminate experts from others by describing the facts and the skills experts possess that others do not—the declarative and procedural knowledge described by Anderson (1983). Klein and Militello (2004) suggested several additional categories of knowledge related to expertise, along with the two offered by Anderson: (a) perceptual skills, (b) mental models, (c) sense of typicality and associations, (d) routines, and (e) declarative knowledge. Mental models and routines are especially pertinent for present purposes, so we describe them further. First, experts understand the dynamics of events in their domain. They know how the natural and artificial systems in their domain are supposed to function. This mental representation of "how things work" is referred to as a mental model, an internal representation of a system in the external world. Mental models enable the decision maker to describe, explain, and predict

(Rouse & Morris, 1986). The "routines" category corresponds to procedural knowledge or "knowing how" discussed by Anderson (1983). Experts know a wider variety of tactics for getting things done. These routines can correspond to physical actions, or to mental routines. Sensemaking strategies can be seen as composing one type of mental routine.

In the Year 1 and Year 2 efforts, the team found little evidence of differences in sensemaking strategies between experts and novices, as described above. The predominant difference in sensemaking ability between experts and novices appeared to be due to differences in the comprehensiveness of the experts' mental models. However, given the utility of uncovering any expert sensemaking strategies for training purposes, we explored the issue further in Year 3. Specifically, the following questions were at issue in the Year 3 efforts that will shape the analyses and discussion of results. First, is there a set of sensemaking strategies or routines that experienced IO personnel use to develop their understanding in uncertain situations? And second, are there aspects of experts' mental models that pertain under sufficiently wide circumstances that training for them would be worthwhile?

A number of specific hypotheses regarding expert/novice differences in sensemaking processes were examined. First, even if there are few differences within sensemaking, it may be that there are differences in how experts coordinate their sensemaking and decision-making processes. For example, in an experimental design task, Schraagen (1993) found that experts exerted a stronger control structure between understanding the problem and pursuing specific designs, whereas novices had a greater tendency to switch back and forth between these two processes (see also Chi, Feltovich, & Glaser, 1981). Second, within the generic sensemaking process, experts may tend to traverse a different set of specific activity paths than do novices. For example, after encountering an anomalous event or other "framebreaker," experts may be more likely to question critical assumption in their frames, whereas novices are more likely to elaborate their frames in a shallow way to accommodate the data. Third, it is fairly widely believed that people should avoid early closure on a hypothesis derived from a frame (e.g., Cohen, Freeman, & Wolf, 1996). If such a strategy is indeed advantageous, we expect that experts will be less committed to their frames than novices in situations where uncertainty and information conflict is high, and thereby prevent premature closure on a frame. There are several observable phenomena implicated by this hypothesis: (a) Experts should have lower confidence in their understanding than novices in such situations, since they appreciate that any of several possible stories or frames may apply. (b) Experts would want to consider all of the information (or as much as possible) before coming to firm conclusions. Hence, in verbalizing their thoughts on a complicated scenario, they would consider and discuss more scenario "threads" or story lines than novices. (c) Experts should bring to mind, and so spontaneously think aloud about, a greater number of hypotheses or explanations than novices. (d) Fourth and finally, irrespective of whether there are general sensemaking processing differences, as described above, it could be that experts have more effective procedures for carrying out specific sub-processes. For example, regardless of whether experts engage in frame questioning any more than novices, they may utilize better questioning strategies when they are in the process of questioning their frames.

These hypotheses were tested in three efforts conducted during Year 3, as described presently.

Experiment 1: Expert/Novice Differences in the Sensemaking Process

In Experiment 1, data originally collected during Year 1 were recoded and reanalyzed, so as to permit evaluation of several of the hypotheses described above. In particular, we examined whether experts differed in their coordination of sensemaking and decision-making processes, whether experts spend relatively more time on initial sensemaking functions than do novices, and whether experts appear to be less fixated on a single interpretation than novices.

Method

Participants. Experiment participants were recruited from the LIWA, now known as 1st Information Operations Command. Eleven participants complete the experiment, five of whom were designated as "experts" by their commanding officer. The experts had an average of 20.3 years of overall Army service and 4.17 years of IO experience. The novices had an average of 18 years in the service and an average of 1.13 years of IO experience. The IO figures may be somewhat misleading, as IO is a relatively new field, and the designated IO experts also apparently had previously worked in areas that were more related to the new IO field than had the novices.

Materials. The experiment materials consisted of scenarios that took the form of sets of situation reports describing the progression of scenario events over the course of several months. The scenarios were constructed by the investigators such that each consisted of three primary themes or "stories within the story," with some other miscellaneous events added in as well. The scenarios and themes for each are presented in Table 1. Each situation report was constrained to fit on a single page. Maps of the scenario regions were used to supplement the situation reports.

Procedure. Participants were oriented to the task, and then asked to think out loud as they proceeded to read and digest the scenario situation reports. After several situation reports had been presented, participants were asked to describe their understanding of the situation at that point. These summary descriptions formed the basis for coding and analysis in the present experiment. The experiment sessions were completed within two hours. At the end of the session, participants were informed of the purpose of the experiment and thanked for their participation.

Coding. The participants' thinkaloud explanations of their understanding were transcribed and coded for additional analysis. To accomplish this, coders who were naïve as to the hypotheses initially read through the transcriptions and divided each into idea units. Idea units corresponded to a single proposition, and generally could be communicated in the form of a simple sentence. These units were then coded along three dimensions: story line, number of qualifiers, and unit content type. There were eight possible values for the story line, corresponding to all the possible permutations of each of the three story lines (i.e., line 1, line 2, line 3, lines 1 and 2, lines 1 and 3, lines 2 and 3, lines 1-3, none of the lines). The number of qualifiers contained in the idea unit was tallied by matching uttered words and phrases with a set list, including "maybe," "possibly," and "I'm guessing that..." There were eleven possible values for the unit content type, as described in Table 2.

Table 1

Scenarios and Major Story Lines Used in Experiment 1

Scenario 1: **Preparing for Winter**

- Story line #1:
  Refugees are moving to camps for winter. The camps are and should be preparing for this movement. Messages indicate where refugees are being sent and the status of the preparations.

- Story line #2:
  There is a growing sewage problem in Megen Camp. This poses health risks to those in the camp. There is some activism on the parts of the residents to take action and get this resolved.

- Story line #3:
  There are key distribution center and supply routes in the vicinity of the refugee movement and where the camps are located. Several events, including a tour of officials and an exercise, are scheduled in areas near these locations. Part of SFOR's job is to ensure that the routes are open so the entire mission can continue to function.

Scenario 2: **Rebuilding the Schools**

- Story line #1:
  There are growing ethnic tensions between teenagers near the school that plans to be integrated.

- Story line #2:
  The IO campaign is leading the charge for the school integration. IO plans how to get the message out and also plans ways to measure whether they are having success.

- Story line #3:
  There is a growing problem with the smuggling of contraband. There have been some arrests, and different families have been tied to or implicated in the smuggling.

Scenario 3: **Activity in the Republika Srpska**

- Story line #1:
  There is some suggestion that a person(s) indicted for war crimes (PIFWC) is moving around in the area and causing some trouble. Perhaps the PIFWC is spreading propaganda or perhaps even inciting anti-refugee activity.

- Story line #2:
  The refugee situation is the main story in the scenario. The fact that refugees are flowing across the border is something that needs to be managed. The number of refugees expected—10,000—should be significant and should trigger planning on behalf of SFOR. In the end, there was anti-refugee activity and people were hurt and killed. SFOR was implicated in propaganda on the airwaves.

- Story line #3:
  Rebuilding activities, thefts in the area, and the increase in patrols by the International Police Task Force (IPTF).

Coding was conducted by two raters who were naïve as to the hypotheses. The raters first coded participant responses independently, and then determined the final codes through consensus. The number of qualifiers was numerical and unbounded, rather than categorical. Nevertheless, the raters agreed on 85% of the units as to the exact number of qualifiers included. In one instance, the number differed by two, and in all other cases, the number of qualifiers differed by one. Cohen's Kappa (κ) was used to assess reliability on the other two dimensions. Kappa was moderate for the story line coding ($\kappa = .52$), and substantial ($\kappa = .70$) for the unit content type (Landis & Koch, 1977).

Results

    Qualifiers. The number of verbal qualifiers participants uttered per unit was used as an indirect indicator of their feelings of uncertainty (Browne & Curley, 1998). The novices ($M = .273$) tended to use more qualifiers than the experts on average ($M = .164$), $t(230) = 1.73$, $p = .085$. This finding suggests that the novices tended to be less confident than the experts, in contrast with the notion that the experts were more "open-minded" than the novices.

    Story Lines. The various scenario story lines constitute multiple threads of information that must be tracked. Many particular facts were presented within each line so as to produce a high data load situation. Under these conditions, do the experts track more of the threads than the novices? To determine this, the proportion of time spent discussing each story line was first determined for each participant. Then, the squared deviation between each of the resulting proportions and 1/3 (since there were three story lines) was computed and summed across the three lines. This deviance measure indicates the extent to which attention was focused on one particular line, as opposed to being evenly distributed across the major story lines. Lower scores indicate more even distribution, and higher scores indicate greater focus on one line. The experts ($M = .292$) did not attend to significantly more of the threads than the novices ($M = .210$), $t(10) < 1$.

    Unit Content. In order to clarify analyses, the eleven content codes were clustered into sets of classes to address specific research questions. First, however, all irrelevant statements were expunged from the data set. In order to address relationships between sensemaking and decision making, three general classes of cognitive function were used, as follows: 1) steady state, 2) decision making, and 3) sensemaking. Steady state included *internal fact*, decision making consisted of the codes *change situation* and *how to change*, and sensemaking comprised the remaining codes. The proportion of idea units that pertained to each class are presented in Table 3. As shown in Table 3, for this task, experts and novices were primarily engaged in sensemaking activities. However, the experts tended to devote a smaller proportion of their thoughts to sensemaking, as compared with decision-making activities, $\chi^2(2) = 5.54$, $p = .063$.

Table 2
Unit Content Codes Used in Experiment 1

1. *Internal fact*: Stating a fact that is **internal** to the story. This is a fairly direct, literal mentioning of message content as described in the scenario. It could also include a comment as to the importance of the fact.

2. *External fact*: Stating a fact based on prior, background knowledge (hence, **external** to the story), rather than information that was presented in the story.

3. *Inference*: Stating an **inference,** interpretation, or hypothesized reason (causal or other) about *why* things are happening. Use this code irrespective of whether the participant seems to be aware that he or she is going beyond the given data. Finally, each unit that is part of an explanation about what is going on in the story spanning two or more units should be given this code.

4. *Anomaly*: Describing a specific **anomaly** or inconsistency between facts or interpretations. Expressing surprise at some message that disconfirms inferences, etc.

5. *Inquiry*: Directing inquiry or seeking information, either by asking an information question directly, stating an action to take designed to improve understanding, or giving any indication of a specific information need to better understand the situation.

6. *Change situation*: Describing a goal or action to **change the situation** based on current understanding.

7. *How to change*: Asking a question or otherwise seeking information about how to change the situation.

8. *Don't know/don't understand:* A statement that indicates the participant lacks knowledge about the scenario generally, or does not understand what's going on. However, the statement does not go so far as to make a specific information request (as in *inquiry*).

9. *Multiple interpretations*: Acknowledging that **multiple** interpretations or possible explanations exist.

10. *Understanding change*: Acknowledging that their own understanding has changed, expanded, etc., in some way.

11. *Irrelevant*: A statement that cannot be taken as one of the above codes, and is otherwise **irrelevant**. Also use for non-informative utterances (e.g., "uh...," "duh...," "oh yeah," "okay") and comments made when "stepping outside" the game.

Table 3
Proportion of Idea Units by Cognitive Function and Level of Expertise

	Steady State	**Sensemaking**	**Decision Making**
Novices	.11	.78	.11
Experts	.10	.68	.22

In a separate analysis, the sensemaking class was further divided into one of three general sensemaking functions from the Data/Frame Model. The data did not support the finer distinctions needed to determine divisions into the six full sensemaking functions, but it was possible to divide the units into the three primary paths stemming directly from the steady state node (specifically, Elaborate, Question, and Compare) based on the content codes. Elaborate

consisted of *external facts* and *inferences*, Question was composed of *anomaly* (this code did not actually appear in the data set), *inquiry, don't know*, and *understanding change*, and Compare included *multiple interpretations*. The proportion of sensemaking idea units for each function are presented in Table 4. As can be seen, the majority of units pertained to frame elaboration, followed by questioning, and lastly by comparison between frames. Experts and novices did not differ significantly in these general kinds of sensemaking that they engaged in, $\chi^2(2) = 1.08, p = .584$.

Table 4

Proportion of Idea Units by Sensemaking Function and Level of Expertise, Experiment 1

	Elaborate	**Question**	**Compare**
Novices	.61	.37	.02
Experts	.64	.32	.05

Finally, a measure was developed to determine whether the experts tended to develop their understanding first and then propose actions to take to change the situation. Specifically, for each participant, the idea units were numbered from 1 to N (the last) according to their order of utterance. Then, again for each participant, the mean order was computed for both the decision making units and the sensemaking units. Finally, the mean ordered value for sensemaking was subtracted from the value for decision making. Note that this measure will be large if all of the sensemaking occurs very early in the process and all of the decision making occurs later. To the extent that sensemaking and decision making overlap, the measure will be small. On this measure, the experts ($M = 13.18$) tended to follow a much more strict control process of making sense of the situation first, and then making decisions, as compared with the novices on average ($M = 3.75$), $t(4) = 2.13, p = .101$.

Summary

As anticipated by the Data/Frame Model and consistent with evidence from Year 1, expert and novice sensemaking was similar in some respects. Specifically, experts did not spread their attention over a wider number of potentially important story theme "developments" than novices. Also, experts' and novices' mental activity was similarly distributed across sensemaking functions in the given task. For example, experts did not spend more time questioning aspects of their frames than novices. Both experts and novices were primarily engaged in elaborating their frames. The critical differences between experts and novices were more centered on the relationship between sensemaking and decision making. Most strikingly, experts maintained tighter control between sensemaking and decision making, engaging in sensemaking activities to develop their understanding first, and then moving on to decision processes (e.g., suggesting courses of action). Novices had a greater tendency to intermix their sensemaking and decision-making processes. Experts also spent proportionally less time on sensemaking, relative to decision making, as compared with novices. Finally, experts tended to be more confident than novices, as measured by the number of qualifiers they uttered.

Experiment 2a: Expert/Novice Differences – Governor Scenario

Experiments 2a and 2b were conducted at 1st IO Command during the Year 3 effort. Although the same participants completed each, the efforts are discussed separately as they consisted of different scenarios and had different specific objectives. Experiment 2a was focused on the composition and strategies for managing the hypothesis components of frames across a wide range of experience levels. With respect to the overall list of hypotheses for the Year 3 effort, Experiment 2a is primarily devoted to the issue of whether experts exhibit tendencies that suggest they are less likely to succumb to early closure on a hypothesis. As in Experiment 1, confidence was examined, though here it was measured directly. Furthermore, the specific scenario used facilitated the detection of differences in hypothesis formation and evidence handling.

Method

Participants. Sixty participants completed the experiment. Twenty were laypeople ("lay"), with no military background and who were recruited from the general populace outside the Dayton area. The remainder was recruited from 1st Information Operations (IO) Command. Of those, 23 were IO novices ("novice") who had nearly completed IO training but had no IO field experience, and 17 had field experience in IO ("field") in addition to IO training. Of the latter group, 4 were further identified as IO "experts" ("expert") via peer nomination, and they are treated separately from the rest of the field group (i.e., the experts are not counted in the field sample, except where mentioned explicitly). It is worth noting that several participants (including those nominated by others to be experts) noted that IO experts do not exist, as the field is too new and complicated. However, they were still able to indicate those who were currently the best among themselves.

The experts had an average of 12.3 years of overall Army service and 4.75 years of IO experience. The field experienced participants had 15.4 years of overall Army experience and 2.8 years of IO experience. The novices had an average of 12.9 years in the service and an average of 1.84 years of IO experience. The overall Army service figures may be somewhat misleading, as many IO personnel serve as private contractors in essentially the same function as their Army counterparts.

Materials. The materials consisted of a written description of a scenario describing a real situation that had occurred in Afghanistan. The incident had been elicited in an earlier experiment utilizing CTA interview techniques. Specifically, incidents were elicited in which IO personnel interviewees had interpreted a situation incorrectly. The full incident is provided in Appendix A. The scenario was constructed with very little modification to the original incident notes, with the exception of leaving off the eventual resolution of the problem that had occurred in the original incident. Participants in the current experiment thus received the following scenario:

*Assume that you have just been assigned as the IO officer for a specified
geographic area of focus in Afghanistan. The basic mission of U.S. forces at the
time of your assignment is to eliminate a permissive environment for the
adversary and deny them sanctuary. Since there are too few U.S. forces, the
Afghans are assisting in the effort.*

*In an outbrief, the previous officer provided information on the governor of the
province that contains your sector. The information included that the governor
has been actively supporting the current government, not the Taliban. The
governor does not have a military background and he was not a mujahideen
during the Soviet occupation. He is from a tribe that is in the minority here. He
has no record of participating in building militias or in drug trafficking. Also, he
has not had any known allegiances with U.S. adversaries in the past 10 years. He
is secure, and subsistence is not an issue for him. Recently, official reports
established that the Afghan governor was seen associating with U.S. adversaries
on more than one occasion. He did not appear to be providing any financial
support to those elements. Also, at one point the governor left town for two or
three weeks. The former IO officer found out that the governor was out of town
for good reasons, and that he had been out of town for extended periods before.
Finally, a junior officer in a lateral organization confronted the governor about
his association with the adversary using a constructive engagement approach.
The governor apparently gave a roundabout answer, saying that he is trying to
change society's attitude of "if you do wrong once, you'll do it again."*

Participants were asked to provide two kinds of responses based on the scenario. First,
they were asked to report their confidence that they understood the situation on a 0-100% scale
(0% = completely certain that I do not understand the situation; 100% = completely certain that
I do understand the situation). Second, they were instructed to describe their understanding of the
situation and to specify in detail the kinds of things they would want to know more about and
any other actions they would take. The participants were also given standard instructions to
thinkaloud (Ericsson & Simon, 1993). The problem scenario and response instructions were
intentionally kept vague in order to minimize imposed constraints and to allow differences in the
way participants structured the problem. As in other challenging fields, IO personnel are often
confronted with issues and events that are ambiguous and ill-structured including definition of
the central problem.

Procedure. Participants took part in the experiment individually in a quiet room at either
1st IO Command or the researchers' offices. The researchers explained that they were interested
in how people of varying levels of expertise developed an understanding or made sense of
situations. Participants were then told that they would be given a couple of scenarios, one at a
time, and for each, that they would be asked to: (1) read the scenario out loud, (2) report their
confidence, (3) describe their understanding of the situation in a thinkaloud procedure, and (4)
report their confidence again. Participants were also provided with a written copy of the
confidence and thinkaloud instructions for reference. A cassette recorder was used to record the
participants' verbalizations. The participants indicated when they felt they had provided
sufficient response to the scenario.

Coding. The participants' thinkaloud explanations of their understanding and questioning were transcribed and coded for analysis. Coding was conducted by two raters who were naïve as to the hypotheses. The raters first coded participant responses independently, then determined the final codes through consensus. Coding was conducted at the participant response level, as opposed to the idea unit level coding performed in Experiment 1. For example, the participants' responses were coded as to whether they contained particular hypotheses as well as to how particular evidence from the scenario was treated. The coding dimensions were (rater agreement given for each code in parentheses):

- *Hypothesized* that the governor is: (1) Anti-U.S. (71%), (2) Pro-U.S. (77%), (3) Pragmatic (96%), or (4) Other (90%). The first and third of these hypotheses correspond to Camps C and A in the original incident (see Appendix A). Note that a participant could have any combination of these hypotheses, since, for example, a participant could entertain and explore the hypothesis that the governor is Anti-U.S. at one point, and then shift to explore the idea that the governor is being Pragmatic at another point.

- *Evidence* was: (1) Questioned (90%; 92%), (2) Interpreted (73%; 81%), or (3) Discounted or Reversed (90%; 92%). This was coded for specific evidence that was taken to indicate that the Governor was either Anti-U.S or Pro-U.S. in the original incident. The specific evidence considered is shown in Table 5.

- *Suspended Judgment*: The participant states that judgment of the governor should be suspended until further evidence can be garnered (56%).

- *Culture*: The participant states that his or her understanding is based, at least in part, on cultural considerations (90%).

Table 5
Evidence Implying that the Afghan Governor was Either Anti-U.S. (Negative) or Pro-U.S. (Positive), Experiment 2

Positive Evidence	**Negative Evidence**
Governor… ■ has been actively supporting the current government, not the Taliban ■ does not have a military background ■ was not a mujahideen during the Soviet occupation ■ has no record of participating in building militias or in drug trafficking ■ has not had any known allegiances with US adversaries in the past 10 years ■ is secure and subsistence is not an issue for him	Governor… ■ was seen associating with US adversaries on more than one occasion ■ left town for two to three weeks ■ apparently gave a roundabout answer

Confidence. Initial confidence depended on experience level. Participants in the lay group were less confident ($M = 53.2$) than were those in the novice ($M = 68.5$) and field ($M = 70.6$) groups, $F(2,55) = 5.76$, $p = .005$. However, the effect was reduced to non-significance for post-report confidence, $F(2,54) = 1.72$, $p = .190$, apparently because confidence increased in terms of both mean and standard deviation for the novice group.

Hypotheses. The hypotheses considered by participants were analyzed in several ways. First, the total number of hypotheses considered was computed. The number of hypotheses considered by participants in the field group ($M = 1.41$) was higher than the number considered by the novice group ($M = .78$), $t(38) = 2.21$, $p = .033$. However, the number considered by the lay group was not different from either of the others ($M = 1.15$).

The various combinations of considered hypotheses were examined, as well. The hypothesis consideration sets for most participants included: Anti-U.S. alone (17%), Anti-U.S. and Pro-U.S. (22%), and no stated hypotheses (32%). The remainder (30%) were scattered over many other possible combinations. Also, the distribution across consideration sets did not depend on experience level.

In terms of individual hypotheses in the consideration sets, 47% of participants considered the Anti-U.S. hypothesis and that did not differ by experience level. Inclusion of the Pro-U.S. hypothesis did depend on experience level, $\chi^2(2) = 6.00$, $p = .050$, such that the novices (17%) were less likely to consider the Pro-U.S. hypothesis than either participants in the lay group (50%) or the field group (47%). Inclusion of the Pragmatic hypothesis appeared to be much greater in the field group (12%) than in the lay (5%) or novice groups (0%). However, the result was not statistically significant, probably because the hypothesis was considered so infrequently. Miscellaneous other hypotheses were considered 20% of the time, and that figure did not depend on experience level.

Evidence Handling. Positive evidence was questioned by participants 25% of the time, interpreted by 28% of participants, and discounted or reversed by 13% of participants. None of these figures depended on experience level. Questioning of negative evidence did depend on experience level, with field experienced participants (77%) questioning significantly more than laypeople (45%), $\chi^2(1) = 3.78$, $p = .052$, and with novices falling in the middle (65%). Negative evidence was interpreted by 28% of participants, and discounted or reversed by 12% of participants. Neither of these depended on level of experience.

Suspend Judgment/Culture. Participants in the field group (59%) and novice group (57%) were more likely to mention that they should suspend judgment and obtain more data than were the laypeople (30%), $\chi^2(1) = 4.04$, $p = .044$. Also, noting the importance of taking the Afghan culture into account in trying to make sense of the governor's actions increased considerably with experience level, $\chi^2(2) = 11.48$, $p = .003$. Specifically, culture was noted by 53% of the field group, 22% of the novice group, and 5% of the laypeople.

Summary

The general sensemaking strategies were similar across experience levels, as in the prior studies. The overall distribution of hypotheses was similar across experience levels, and participants in the higher experience levels were not more likely to consider that the governor was pragmatic. The novices, however, were least likely to consider that the governor was pro-U.S. Further, participants across experience levels handled evidence in generally similar ways. Across the board, participants questioned, interpreted, and discounted evidence; rarely were all the facts simply taken at face value.

Particular aspects of sensemaking were found to differ in specific ways in Experiment 2a. Although evidence handling was generally similar across experience levels, the likelihood of questioning negative evidence (i.e., evidence suggesting that the governor might be anti-U.S.) tended to increase with experience level, and differed statistically between the field and lay groups. Also, the field group considered more hypotheses than the novices, and the field and novice groups were more likely to deliberately suspend their judgment than the lay group. In seeming contradiction with this finding, the lay group was less confident than the novices and field groups. Finally, there was a sharp increase as a function of experience in the extent to which participants considered culture as a critical factor to consider.

Experiment 2b: Expert/Novice Differences—Bus Scenario

The primary aim of Experiment 2b was to move past examination of sensemaking differences at a general level, and instead focus specifically on the questioning strategies participants used when confronted with anomalous information. The basic idea was based on earlier work by Dunbar on scientific thinking. In an experiment of simulated scientific discoveries, Dunbar (1993) asked participants to attempt to discover how genes control other genes. The simulated environment was based on actual biological models developed by Monod and Jacob, and participants were placed in the same original (and incorrect) hypothesis state as those biologists at the start of their research. Participants were able to conduct experiments that provided evidence inconsistent with their original hypothesis. At this point, participants dealt with the inconsistent evidence depending on how they set their goals. One group maintained the goal of seeking evidence to support their original hypothesis. None of these participants ever discovered the actual mechanism governing gene regulation. Others instead set a new goal of attempting to explain the cause of the inconsistent findings. These participants tended to generate the correct hypothesis. These, and other findings, led Dunbar (1993) to conclude that people's goals determine when and how inconsistent evidence is used. In particular, in the face of inconsistent evidence, when people change their goal to one of determining the cause of unexpected findings, they make discoveries.

Dunbar (1995) extended these findings in a naturalistic study of actual biological scientists working in their laboratories. As in other naturalistic studies, important aspects of the natural situation were the influences of experience and group interactions on individual scientists' handling of inconsistent evidence. In terms of experience, the more experienced scientists were much more willing to let go of their original hypothesis, and set a new goal of understanding the inconsistent findings. The experienced scientists were especially sensitive to

inconsistent findings arising in their control conditions and explicitly thought of control conditions as serving two purposes: checking experimental conditions and exposing hidden mechanisms. However, the very experienced scientists appeared to display what Dunbar termed a "falsification bias," that is, they would discard good data that actually confirmed their hypothesis.

In Experiment 2b, we sought to determine whether experienced IO personnel would be more likely than novices and laypeople to engage in questioning strategies matching those of experienced scientists. That is, does the strategy of directing inquiry towards explaining the cause of inconsistent findings generalize beyond the scientific domain?

Method

Participants. Participants were the same as in Experiment 2a.

Materials. The materials consisted of a written description of a scenario describing a real situation that had occurred in Kosovo. As in Experiment 2a, the incident had been elicited in an earlier study utilizing CTA interview techniques. The full incident is provided in Appendix A. The scenario was constructed with very little modification to the original incident notes, with the exception of leaving off the eventual resolution of the problem that had occurred in the original incident. Participants in the current experiment thus received the following scenario:

> *Assume that you have just been assigned as the IO officer for a specified zone in Kosovo. The overall mission is to keep things safe and secure in this geographic area.*
>
> *In an outbrief the previous officer provided information on a Serb bus situation in your area. The Serbs are a protected minority in Kosovo, live in enclaves, and are afraid to venture forth from these enclaves. The U.S. has been escorting Serb college students out of their enclave to a college in north-central Kosovo. The students are being transported in a bus donated by Denmark, with an armored vehicle in the lead and in trail. It is very expensive to provide the full escorts, but about two years before your arrival, a bomb blew up a bus, and 17 were killed. So there is a fair amount of nervousness on the part of the Serbs. A U.S. decision was made fairly recently to increase the efficiency of the escorts while maintaining safe and secure travel. Specifically, they first reduced the two armored vehicles to one for a period, and then down to zero armored vehicles but with a guard on the bus.*
>
> *The decision was announced about two weeks after being made. The Commander decided to stop the escorts before the last day of school, so that the mindset would be to not expect buses the next year. This provided a whole summer to adjust to the idea. The thinking was that a demonstration of additional air surveillance and security, along with the motivation of the students to finish the term, would carry them through. Some drop in ridership was expected, though it was assumed that over time the levels would go back up. At the time of the decision, the Commander began to have the number of students riding the bus to and from school per day presented to him in daily briefs. Prior to the announcement, it was found that the buses were driving 60-80 students per day.*
>
> *When the planned change was announced, the scheme of going from two to one to zero was described. The students/riders were shown all kinds of "gee whiz" stuff, including a UAV video, a CD show presenting aerial views, etc. They were told to wave at the UAVs,*

and that they could see themselves on the monitors (demonstrating the quality of the surveillance). The U.S. also issued posters saying, "We can see you. KFOR anytime, anywhere—we can provide security." The U.S. started to hear complaints from the college students themselves, from the bus drivers, and from some local leaders. The students said they were worried about access to the buses (the new routes would be different), possible cost increases and convenience. The bus drivers said they were worried about personal security. Of the leaders, the moderate Serbs gave grudging approval and cooperation, though the hardliners did not. At this point, ridership was still averaging 60-80 students/day. Rumors were heard that the students would not ride when less security was provided.

About two weeks after the announcement, the change was implemented to go from two vehicles to one vehicle. At this point, the bus drivers seemed to be mollified. The ridership dropped to 50-60 students/day. Then, in another two weeks, the switch from one vehicle to zero vehicles and a lightly armed guard was implemented. The number of riders dropped down to 10 students/day. Also, the bus drivers are planning a strike because of the lack of security.

Participants were asked to provide the same two kinds of responses based on the scenario as in Experiment 1a.

Procedure. The procedure was the same as in Experiment 1a.

Coding. The participants' thinkaloud explanations of their understanding and questioning were transcribed and coded for analysis. Coding was conducted by two raters who were naïve as to the hypotheses. The raters first coded participant responses independently, and then determined the final codes through consensus. Coding was conducted at the participant response level, as in Experiment 2a. The coding dimensions were (rater agreement given for each code in parentheses):

- *Detect Drop*: The individual indicated detecting that the massive drop in ridership was the central problem in the scenario. That is, the ridership drop was recognized as a critical cue that triggered sensemaking (81%).

- *"Why Stop Riding?" Question*: The individual explicitly wanted to try to find out or just asked why the students had stopped riding the bus (92%).

- *Challenge Assumption ("Who Decides?" Question)*: The individual explicitly challenged the assumption that riding the bus was the students' decision. That is, the participant considered the hypothesis that someone else might be influencing the students to stop riding the bus. Note that the challenging hypothesis might still be wrong. For example, a participant might hypothesize that some gang leaders within the students are influencing the decision, but still challenge the basic assumption (92%).

- *Correct Hypothesis.* The correct hypothesis was among those discussed by the participant. For the bus scenario, the actual state of affairs was that the mothers were

22

not letting their students ride the bus. If participants said anything along the lines of it being about the family, mothers, etc., they were scored as having considered the correct hypothesis (100%).

Results

Confidence: As in Experiment 1 and Experiment 2a, confidence appeared to depend on experience level, such that participants in the field group were more confident ($M_1 = 76.6$; $M_2 = 78.2$) than were those in the lay ($M_1 = 70.0$; $M_2 = 74.1$) and novice ($M_1 = 70.8$; $M_2 = 75.5$) groups. The effect did not reach statistical significance in the current experiment, though the trend is consistent across the three efforts. As can be seen, confidence increased from the initial, pre-thinkaloud report to the post-thinkaloud report, $t(54) = 2.54$, $p = .014$. This effect was expected in the current experiment, as the scenario contained more information and less conflicting information as compared with the scenario in Experiment 2a.

Content Codes: Most participants recognized the drop in ridership as the primary problem in the scenario (73%) and that figure did not depend on level of experience. As noted above, the protocols were coded for key kinds of inquiries participants made, in particular inquiries that would lead directly to developing an accurate understanding of the scenario. The two key inquiry types were: "Why are the students not riding?" and "Is someone else influencing the student's decision?" The proportions of participants who asked each of these key decisions by experience level are presented in Table 6. As shown, participants with field experience were three to four times more likely to ask one of these critical questions than were those with no field experience (trained or not), $\chi^2(1) = 5.31$, $p = .02$ for the "why" question, and $\chi^2(1) = 5.31$, $p = .02$ for the "who" question. The results were not due to the experienced participants simply "knowing" the answer. Only three participants hypothesized the correct answer (coded liberally as "family decides" is the reason). Also, accuracy did not depend on field experience, $\chi^2(1) = 0.04$.

Table 6
Proportion of Participants Who Asked Each Question by Level of Expertise, Experiment 3

Key Inquiries	Experience Level			
	Lay	Novice	Field	Expert
"Why not ride?"	.10	.04	.23	.50
"Who decides?"	.05	.09	.23	.50

Summary

Experienced IO personnel were much more likely to question important aspects of their frames than laypeople and trained novices. At one level, the kinds of questions they asked were quite basic, lacking the obvious technical sophistication that might often be assumed to be associated with experience and expertise. Nevertheless, these simple questions were exactly the kind needed to develop a useful frame on which to base decisions and actions, and are quite similar to questioning strategies of experienced scientists.

The three research efforts conducted during Year 3 provide additional insight into the nature of sensemaking expertise. In terms of the hypotheses specified at the fore, we found that IO experts exhibit a stronger delineation between sensemaking and decision making processes. Experts attempted to first clearly understand the situation they faced, and then proceeded to formulate courses of action and other decision-making activities. Furthermore, IO experts were much more likely than others to formulate high-quality questions geared towards understanding anomalous evidence. The kinds of questions IO experts asked varied in terms of generality. For example, a very general kind of question would be, "Why did this anomalous event occur?" whereas a question specific and highly relevant to the IO domain was, "Who are the key decision makers in this situation?" Although IO expertise differences were found at both the broadest (e.g., coordination between sensemaking and decision making) and most specific (e.g., how frames were questioned) levels, we did not detect differences in the relative time spent on distinct sensemaking functions. Finally, regarding the issue of reduced fixation for experts, the evidence appears to be quite mixed. In support of the idea, we found a greater number of considered hypotheses with more experience, as well as more suggestion that judgment ought to be suspended. We also found more experienced IO personnel were more likely to question anomalous evidence. However, we also found consistent evidence that confidence in these uncertain and anomalous situations increases with experience. In addition, experts did not attempt to track more information than novices. We attempt to reconcile these last findings as a part of the theoretical discussion, to follow.

Theoretical Implications

The Data/Frame Model of Sensemaking, as represented at the completion of this project, is illustrated in Figure 1. As shown, the model consists of six sensemaking functions, along with temporal path relations that link the functions. See Table 7 for a description of the functions.
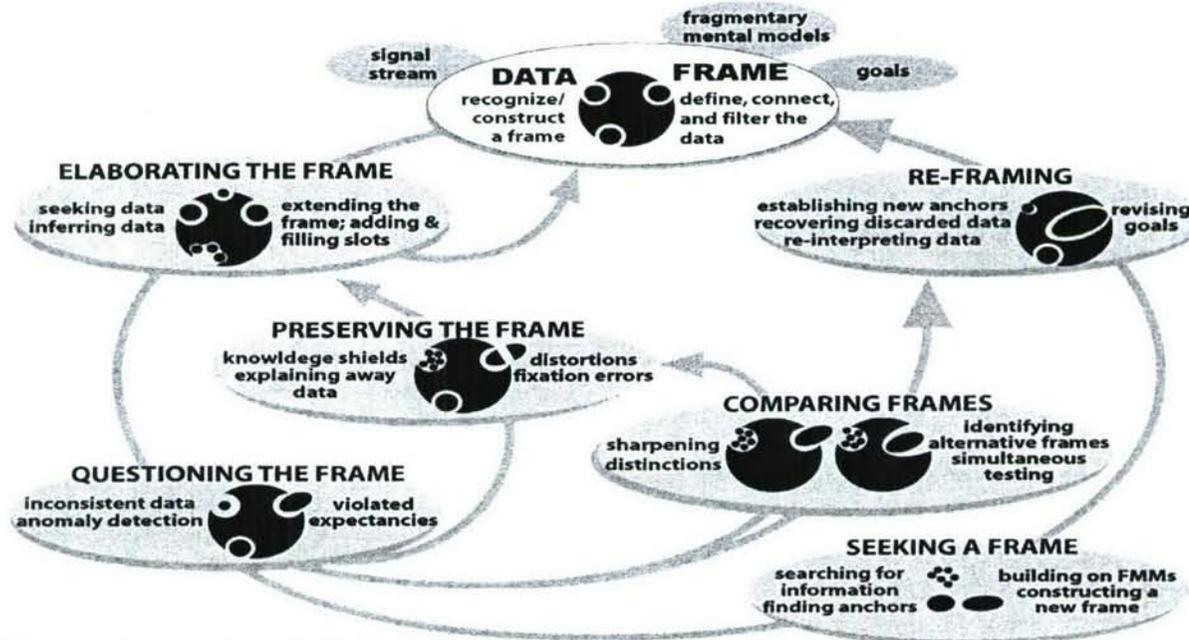

Figure 1. Data/Frame Model of Sensemaking.

Table 7

Description of Sensemaking Functions

I. ELABORATING THE FRAME	

Seeking Data: Facts gathered directly from a situation which must be compared and fitted with the data to determine their adequacy.

Inferring Data: Data are perceived and become the basis for generating inferences.

Extending the Frame: As more is learned about the situation, the frame becomes more elaborate from drawn inferences and observation.

Adding and Filling Slots: The general structure of the frame may be known, but as more information is sought out, inferred, and constructed, one may begin to piece together a more comprehensive picture.

Internal Knowledge: This accounts for previous experience and individually constructed information.

II. QUESTIONING THE FRAME

Inconsistent Data: One realizes that the data do not match the frame.

Anomaly Detection: Detection of a unique circumstance in a situation one normally encounters instigates the assessment of use of an alternative strategy.

Violated Expectancies: Frames provide people with expectancies; when they are violated people begin to question the accuracy of the frame.

III. PRESERVING THE FRAME

Knowledge Shields: Preserving a frame in the face of countervailing evidence. In other words, minimizing the importance of the contradictory data.

Explaining the Data Away: Justifying why inconsistent data do not match the frame.

Distortions: This can be thought of as the byproduct of elaborating on a frame. One has the opportunity to notice the flaw in the way he or she understands the events and to alter or replace the flawed interpretation, but instead, discards the anomalous data and persists in the flawed situation account. Novices tend to get trapped in the "garden path" (relying on data/anchors that are inaccurate or are not diagnostic) while experts are able to break free, using their stronger knowledge base along with a "logical competitor set" strategy.

Fixation Errors: Reducing consideration of alternative frames or preserving a frame that should be discarded. Often, the process of preserving an inaccurate frame leads from one error to another, as waypoints and landmarks are continually misidentified.

IV. COMPARING THE FRAME

Sharpening Distinctions: Gathering evidence in support of one (the primary) frame.

Seeking Distinguishing Evidence: Elaborating on an opposing frame.

Identifying Alternative Frames: Once broken free of fixation, one selects various frames to sharpen the distinctions that were relevant, pinpointing critical details.

Simultaneous Testing: Testing members of the set, or the most likely member, as a way for decision makers to engage in the activity of comparing frames and achieving a differential diagnosis. Information-seeking becomes more directed and efficient when the decision maker can work from a set of related and competing frames.

Table 7 (Cont.)
Description of Sensemaking Functions

V. SEEKING A FRAME

Searching for Information: This process entails selecting the relevant frame to generate an explanation. Frames dictate which data from the environment will be recognized as pertinent, how data will be interpreted, and what role they will play when incorporated into the evolving frame. Sensemaking suffers when the frame is inadequate; available information or data modifies one's schema/frame of the present environment, which directs one's explanation and/or sampling of the environment.

Finding Anchors: One to two key data elements serve as anchors to elicit the initial frame. People consider, at most, three or four anchors in deriving a frame. The quality of the frame enables one to select the right data from the environment, interpret it accurately, and see pertinent patters and connections in the data stream.

Building on FMMs: FMMs are local cause/effect connections. They are evoked in order to create a "just-in-time" mental model of a situation or phenomenon. Instead of positing that mental models are part of long-term memory, they are seen as constructions based on the requirements of the situation. The various types of FMMs (i.e., causal relationships, rules, principles, etc.) help guide the selection and interpretation of data in the sensemaking process. A set of FMMs contribute to the frame constructed by the sensemaker in order to help guide the selection and interpretation of the data.

Constructing a New Frame: Evolving frames dictate what data from the environment seem relevant and how they should be interpreted. Data may be used to confirm an explanation or frame when the data are unclear and alternate explanations are possible.

Schema: An expert understanding by evoking principles that can be used to solve a problem. Rules generated by novices do not contain actions that are explicitly tied to solution procedures. Novices know what problem cues are relevant, but do not know what to do with the knowledge.

VI. REFRAMING

Establishing New Anchors: This defines associations vs. noise, and allows one to discard unnecessary, irrelevant data and to recognize the importance of new data.

Recovering Discarded Data: Using data that now has relevance to the frame.

Re-Interpreting Data: This step stems from use of data that was originally thought to have no relevance. The frame may need to be re-interpreted because of importance of the new data.

Revising Goals: Due to the recovery of the discarded data, goals may need to be revised from the original frame.

Sensemaking Recovery: This occurs when an individual pursing a certain goal finds that the current situation is not what was expected, and recognizes what the situation actually is; thus, the individual is able to form an intelligible world view.

The collection of six functions appears to be necessary and sufficient for covering the sensemaking activity that we have witnessed in the three years of research. Furthermore, the model in its current form has much to offer in the way of practical applications, as described in the section below. However, in order to further develop useful applications, the workings of these functions need to be specified more precisely and in greater detail. In addition, there is a need to determine the characteristics of quality products from each function as a guide for

training and technological support. Part of the Year 3 efforts provided an initial step by developing some ideas regarding expertise effects on frame questioning.

Practical Implications

There are several practical implications of this effort for both training and tool development. One training concept follows directly from the Year 3 effort, and so would apply to the IO domain in particular. Recall that in the Year 3 efforts, participants of varying experience levels responded to scenarios that were constructed from previously conducted CTA interviews. Our Year 3 results describe how the most and least experienced participants differed with respect to their questioning strategies, among other things. Hence, these differences highlight learning principles or objectives for training. In addition, the responses from the most experienced respondents can be used explicitly as feedback for trainees. For example, IO trainees could be given a scenario based on the original CTA and asked to respond on their own, say in written form. Then, after they have completed their response, the trainees could be provided with the learning principles and one or two of the expert responses to provide specific examples of how the principles manifest in the context of the scenario. In this way, the trainees are provided with the opportunity to practice in a short exercise and then receive meaningful process feedback, but without the cost of a facilitator. Thus, the general idea is to provide high impact training exercises that are short and inexpensive for sensemaking skills. We plan to provide samples of these training exercises to 1st IO Command for use in their onsite training programs. In addition, we are in the process of establishing follow-on links with applied funding programs for the purposes of further developing and evaluating sensemaking training for IO.

The model and empirical results are also potentially useful in intelligence analysis and related domains. Many efforts to improve intelligence analysis are largely focused on the development of technologies such as social network analysis to understand communication traffic and collaboration paths. Note that such tools are geared towards condensing and synthesizing large amounts of raw data, so that analysts need only process a summary record. Some of these kinds of analysis tools will undoubtedly be useful in supporting analysts, perhaps by illuminating previously unseen data and patterns of activity. However, it would be a mistake to believe that they can replace sensemaking functions or directly transmit understanding (frames) to the analyst. From the point of view of the Data/Frame Model, the sensemaker always draws on his or her own frames in order to interpret and question presented data, even if that data is highly processed and cast in graphic or pictorial form. Hence, analysis tools can change the nature of the data that the analyst is attempting to make sense of, but it will not change the core cognitive functions involved. This implies that information technology (IT) development should be centered on supporting the core sensemaking functions of analysts (or other decision makers), rather than attempting to circumvent them. This ought to include tools that directly support analysts' attempts to think through and clarify their understanding of situations, given the facts they already have at hand. Klein Associates is currently pursuing applied efforts in the intelligence analysis domain, and the Data/Frame Model is figuring prominently in support of those efforts (Klein, 2004).

Finally, the Data/Frame Model is proving invaluable for system design. Klein Associates has recently completed Phase I of a Small Business Innovative Research project dedicated to the development of IT that aids sensemaking. The focus of the project is to provide Navy and other

system designers with a methodology and support tools for capturing and leveraging sensemaking requirements. Currently, these support tools do not exist. Our Phase I effort has successfully developed and we piloted a sensemaking methodology that is rooted in the Data/Frame Model (Klein, Long, Hutton, & Shafer, 2004). Our proposed Phase II effort will apply and validate this methodology and then formalize designer support tools.

References

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Browne, G. J., & Curley, S. P. (1998). Reasoning with category knowledge in probability forecasting: Typicality and perceived variability effects. In G. Wright & P. Goodwin (Eds.), *Forecasting with judgment* (pp. 169-200). Chichester, England: John Wiley & Sons.

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.

Cohen, M. S., Freeman, J. T., & Wolf, S. (1996). Meta-recognition in time-stressed decision making: Recognizing, critiquing, and correcting. *Human Factors, 38*, 206-219.

Crandall, B., & Gamblian, V. (1991). Guide to early sepsis assessment in the NICU. Instruction manual prepared for the Ohio Department of Development under the Ohio SBIR Bridge Grant program. Fairborn, OH: Klein Associates Inc.

Dunbar, K. (1993). Scientific reasoning strategies for concept discovery in a complex domain. *Cognitive Science, 17*, 397-434.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 365-398). Cambridge: MIT Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Cambridge, MA: MIT Press.

Kaempf, G. L., Wolf, S., Thordsen, M. L., & Klein, G. (1992). *Decision making in the AEGIS combat information center* (Contract N66001-90-C-6023 for the Naval Command, Control and Ocean Surveillance Center). Fairborn, OH: Klein Associates Inc.

Klein, G. (2004). *A Data/Frame Model of Sensemaking. Presented at the Friends of the Intelligence Community (FOIC) Meeting held January 21-23 (Washington, D.C.)*. Fairborn, OH: Klein Associates.

Klein, G., Long, W. G., Hutton, R. J. B., & Shafer, J. (2004). *Battlesense: An innovative sensemaking-centered design approach for combat systems* (Final report prepared under contract # N00178-04-C-3017 for Naval Surface Warfare Center, Dahlgren, VA). Fairborn, Ohio: Klein Associates Inc.

Klein, G., & Militello, L. (2004). The knowledge audit as a method for cognitive task analysis. In H. Montgomery & R. Lipshitz & B. Brehmer (Eds.), *How professionals make decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, G., Phillips, J. K., Battaglia, D. A., Wiggins, S. L., & Ross, K. G. (2002). *Focus: A model of sensemaking* (Interim Report - Year 1 prepared under Contract 1435-01-01-CT-31161 [Dept. of the Interior] for the U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA). Fairborn, OH: Klein Associates.

Klein, G., Phillips, J. K., Rall, E. L., Thunholm, P., Battaglia, D. A., & Ross, K. G. (2003). *FOCUS: A model of sensemaking* (Year 2 Interim Report prepared under Contract #1435-01-01-CT-3116 [Dept. of the Interior] for the U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA). Fairborn, OH: Klein Associates.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1988). *Rapid decision making on the fireground* (Technical Report DTIC No. AD-A199492, http://www.dtic.mil). Alexander, VA: U.S. Army Research Institute.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, 19*(3), 462-472.

Landis, J. R., & Koch, G. G. (1977). A one way components of variance model for categorical data. *Biometrics, 33*, 671-679.

Phillips, J. K., Baxter, H. C., & Harris, D. S. (2003). *Evaluating a scenario-based training approach for enhancing situation awareness skills* (Technical Report prepared for ISX Corporation under contract DASW01-C-0036). Fairborn, OH: Klein Associates Inc.

Ross, K. G., Battaglia, D. A., Hutton, R. J. B., & Crandall, B. (2003). *Development of an instructional model for tutoring tactical thinking* (Final Technical Report for Subcontract No. SHAI-COMM-01; Prime Contract No. DASW01-01-C-0039 submitted to Stottler Henke Associates Inc., San Mateo, CA). Fairborn, OH: Klein Associates.

Ross, K. G., Thunholm, P., Uehara, M. A., McHugh, A., Crandall, B., Battaglia, D. A., Klein, G., & Harder, R. (2003). *Unit of Action Battle Command: Decision-making process, staff organizations, and collaborations* (Report prepared through collaborative participation in the Advanced Decision Architectures Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0009). Fairborn, OH: Klein Associates.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits on the search for mental models. *Psychological Bulletin, 100*(3), 349-363.

Schraagen, J. M. (1993). How experts solve a novel problem in experimental design. *Cognitive Science, 17*, 285-309.

Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage Publications.

APPENDIX A:
CRITICAL INCIDENTS USED IN SCENARIO DEVELOPMENT

Afghan Governor Incident

Background:

The incident is from an Information Operations (IO) officer ("the Major") who was deployed to Afghanistan from April to October 2002. He commented that the decision making in Afghanistan was so foreign and different that he has thought a lot about it.

Incident Overview:

The basic mission in Afghanistan was to deny the adversary sanctuary and eliminate a permissive environment for the U.S. adversaries. There were too few U.S. forces. So the Afghans were going to have to do this for us. The Major was the IO officer for this area. The event happened in late June. They received official reports that a governor was associating with U.S. adversaries. The official was the governor of a province in a larger area. Other people are also important in the province. "Governor" does not imply the same power and authority that it does in the U.S. The official is from a tribe that is in the minority here. The real power is with the other groups. Next, in July, the official disappears for a few weeks. This is very suspicious. In this part of the world, you don't just disappear.

What was going on, and why did he have these "bad contacts"? One possibility was that he was trying to build bridges with the opposing elements. Another possibility was that he wasn't very smart, but was an opportunist who didn't realize the implications of having these contacts. The final possibility was that he himself was anti-U.S. Most U.S. people jumped to the third possibility, because they didn't appreciate his decision making, didn't appreciate the role of elders. The Major disagreed, thought it was the second. In reality, it was the first.

NOTE:

The three beliefs most U.S. military subscribed to for why the governor was associating with known U.S. adversaries were that the governor was associating with U.S. adversaries because he is either:

   a) Trying to build bridges with other important groups/factions within his geographic area
   b) Unwittingly going against stated policy not to meet with U.S. adversaries
   c) Anti-U.S. himself

We described people who subscribed to these beliefs as being in Camp A, Camp B, or Camp C.

Time	Incident Events
Before he arrives in Afghanistan learns of mission	The Major learns that Afghans are assisting in the ongoing effort to deny sanctuary and dry up permissive environment. The prevailing belief among U.S. comrades is that all Afghan government officials are corrupt. Government officials cannot be trusted, whether they say they support the U.S. effort or not.

Time	Incident Events
~20 JUN 02	Reports establish that the Afghan "governor" of their geographic area of focus was seen associating with U.S. adversaries. The Governor's association with adversaries seemed to be an isolated activity—he didn't do other things to suggest corruption or adversarial intent. Most people in the leadership and in Intel were in Camp C at this point, saying, "How could the governor not be anti-U.S. when he lets U.S. adversaries roam around the area for which he's responsible?" The Major wants to figure out the governor's motivation for associating with U.S. adversaries. He also wants to have enough evidence to be able to confidently present his assessment of the governor's association with U.S. adversaries to the leadership. At this point, he doesn't present his initial assessment of situation to leadership because he isn't confident of his assessment. The Major has to distinguish between things he can believe versus things he can use in the military decision making process. The Major is in Camp B at this point. He believes this because the governor didn't fit an anti-U.S. profile: • He didn't have a military background (he had not been a mujahideen, and that was a bit unusual) • He had no known allegiances with U.S. adversaries in previous 10 years • He was actively supporting current government (not Taliban) • He didn't appear to financially support U.S. adversaries Further, the Major couldn't find obvious motivation for the association: • He was secure • Subsistence wasn't an issue for him • He hadn't participated in building militias or in drug trafficking. The Major was also aware that the U.S. IO didn't have a way to think about how the Afghanis ought to be acting. For example, actions that Afghanis believe are completely innocuous are completely beyond the pale to the U.S.
10-20 JUL	A second report is made of the Governor associating with U.S. adversaries. This second meeting reinforces the prevailing view that "Afghan government officials can't be trusted." The Major finds that he is starting to move away from Camp B, towards Camp C. The Major is responsible for engagements with the Afghan leadership and he wanted to have someone engage the governor in a constructive way to find out more about what he was doing. The people in Camp C want to pursue a confrontational approach with the governor. The Major delegates the meeting with the governor to a lateral unit.

Time	Incident Events
late JUL 25-30 JUL	The Major now receives a report that the governor has been out of town for almost two or three weeks. People don't just disappear in this area of the world for no good reason. And when people do disappear, it's usually not good news. The Major knows Intel will have more information and that this would help them to make sense of the report. The Major gives the report to Intel and asks them to make sense of it. Intel gave a range of reasons for why he could be away from town for weeks. But, because of pace of operations, the information doesn't figure into the Major's or most others' assessment of the situation. The Major now has one foot in Camp B and one foot in Camp C. He feels that something more is going on than meets the eye. A few other people take on a Camp B perspective.
~10 AUG	The U.S. finds out that the governor was out of town for good reasons (The Major didn't say what they were). Also, the Major learns more about the background of the governor in the late July/early Aug timeframe. He learns that the governor had been out of town for extended periods before.

A junior officer in a lateral organization approaches the governor about his association with U.S. adversaries using a constructive approach. The governor gives a roundabout answer. He says that he is trying to change society's attitude of "If you do wrong once, you'll do it again." The governor also said indirectly that he would continue to meet with all "stakeholders" in the region. The inferences and interpretations made by the lateral organization about the results of this meeting don't carry much weight with the Major or his organization because a junior officer in the lateral organization took the meeting.

The Major considers the Camp A hypothesis for the first time because that's what the governor described in essence, but he is still in Camp B. More people move from Camp C into Camp B, and a few even move into Camp A. The Major said to himself, "I just can't believe he would meet with U.S. adversaries if he knew how it was perceived by the U.S. and those that support us. If he did understand this, he wouldn't do this." |

Time	Incident Events
Late Aug	The Major and a senior leader in his vertical hierarchy meet with the governor. This is the second time U.S. military meets with governor. The goal of the meeting is twofold: to understand why the governor has been meeting with U.S. adversaries and to convince him to discontinue his association with the adversaries. In the meeting, the governor lists everybody he's been meeting with except the adversaries. The major and senior leader interprets this as the governor's way of telling them he knew what they were saying and why. It was also his way of letting them know he was having these meetings with the U.S. adversaries for the reason he had stated earlier—bridge building. The Major's senior leader tells him he believes what the governor is saying: that he's trying to build bridges with ALL stakeholders in the geographic region. The Major finally understands that for most civilians the person who they consider their leader tells them what decisions to make. Leaders (official or unofficial) are expected to make decisions for the community. People follow the decisions of the leader because it's a cultural practice, not out of fear. The inferences drawn by the senior leader carry a lot of weight with the Major because he knows and trusts the leader's expertise. The Major is firmly in Camp A now and many others have moved into this camp as well.

Kosovo Bus Incident

This incident is from a civilian contractor who was deployed to Kosovo from December 2001 to August 2002. The overall mission was peacekeeping as a part of KFOR, 4 Alpha. They were supposed to keep things safe and secure in this zone in Kosovo.

The specific task was to streamline the escorts of Kosovo Serbian college students to school. The incident lasted from April to May of 2002. The Serbs are a protected minority in Kosovo, live in enclaves, and are afraid to venture forth from these enclaves. The U.S. was escorting Serb students out of their enclave using a bus donated by Denmark. They transported the college students in that bus, with an armored vehicle in the lead and in trail. It was very expensive to provide the full escorts, but in 1999 a bomb blew up a bus, and 17 were killed. So there was a fair amount of nervousness on the part of the Serbs. Also, the Serbs want the U.S. in Kosovo forever, because once we are convinced it is safe enough to leave, there will be a push for an independent Kosovo.

Given the expense, the Commander decided to increase the efficiency of the escorts while maintaining safe and secure travel. They were going to downshift from the two armored vehicles, down to one armored vehicle, and down to zero armored vehicles and a guard on the bus. At the same time, they were going to provide air surveillance and take other steps such as staged ground forces on the route. They were issuing posters saying "We can see you." "KFOR anytime, anywhere—we can provide security." The Commander decided to stop the escorts before the last day of school, June 12, so that the mindset would be set to not expect buses the next year. The

idea was to provide a whole summer to adjust to the idea, and the following year the escorts wouldn't be used.

The Commander felt they had a good, gradual plan. Then there was a bus driver problem, and then a Danish problem. Finally, there was an apparent student security perception problem. However, it eventually turned out that it was not the student's perceptions that were the problem. Instead, the parents were the real target audience here, as they were the real decision makers. It took a long time for the U.S. to figure this out, because their mindset was that college students would be making their own decisions about going to school. Also, they had access to a cultural advisor, a U.S. citizen who was Serb by birth, born in Montenegro. She did not pick up the potential problem with the mothers.

Time	Incident Events
April 1, 2002	The decision is made to change the escorts. In order to track the issue, the Commander has the following information added to his daily briefs: • The number of Army vehicles per day that were tied up as escorts. • The total number in the pool riding the bus to school and from school —and how many were students (other Serbian non-students were also slipping onto these buses for safe transportation). • Number of students and number of free riders per day. The buses are driving 60-80 students per day. The riders are mostly college students going to a college in north-central Kosovo.
April 15	The U.S. announces a change in the future. The scheme of going from two to one to zero is described. The U.S. shows the students all kinds of gee-whiz stuff, UAV video, a CD show presenting aerial views, etc., and demonstrates the quality of the surveillance by having the riders wave at the UAVs, so they could see themselves on the monitors. Some local leaders complain about the plan, and some complained throughout. Moderate Serbs, though, give grudging approval and cooperation, although the hardliners don't. They also start to hear complaints from the college students, themselves. The students say they are worried about access to the buses (the new routes will be different) and about higher costs. Plus they will have to start earlier (5:30 a.m. instead of 6:00 a.m.). The new route will take them to a train, and they will have a 10 minute walk from the train to the university. This walk has been checked and the safety is OK. The U.S. hope was that the demonstration of security, and the motivation of the students to finish the term, would carry through.

Time	Incident Events
April 15 (Cont.)	The bus drivers complain about security, and they develop a plan for a strike. Finally, there is a complication with the Danish. The Danish do not want anyone on their buses who is armed. That rules out an armed guard. This also sends the U.S. military over the edge—"They provide the buses, we provide the security." At this point, the buses are still averaging 80 students per day. Some rumors are heard that the students will not ride when less security is provided. The U.S. assessment is that the bus drivers were the critical element here.
May 1	The change is implemented to go from two vehicles to one vehicle. An agreement is made with the Danish. There will be no rifles on the buses, but guns (small arms) were allowed. At this point, the bus drivers seem to be mollified. They experience a drop to 50-60 students per day. This is not a big drop, and it is not unexpected. The U.S. assumes over time that the ridership will go back up.
May 15	They now go from one vehicle to zero vehicles and a lightly armed guard. Suddenly, there is a dramatic drop in riders, down to 10 students per day. This is unexpected, and they initiate some data gathering. The PSYOP and Civil Affairs (CA) Teams talk more in depth with the students. The students' answer is simply that they are not riding because of the lack of security. Also, the bus drivers are planning a strike because of the lack of security. For now, the focus is on the bus driver problem, and how to keep the buses running. They initiate negotiations with the drivers over security.
May 20	The ridership is still way down, and the U.S. is still surprised over the dramatic drop in riders. The Commander asks PSYOP and CA teams what happened, and why. There is not a problem with weather or exams. So, "What are we missing?" The Commander tells the interviewers he doesn't just want to know that they don't feel secure, but why? Why don't they feel secure? Now they start getting useful answers. The students say, "I'd like to go to school. But my mom won't let me because of the insufficient security." So, the parents of the students are the real target audience, or at least they are a critical target audience that was not previously considered. The earlier answer from the students, about a lack of security, was not the real answer, just an excuse. The real answer was that their mothers wouldn't let them ride. The Americans did not realize how much parental authority there was in this region; the Americans assumed that college students were their own agents, and never considered that most of them still live at home and that their parents maintain authority for key decisions. In America, who would ever think to check with the parents of college students about going to school? They had initiated the PR to the students, but never thought to include the mothers.

Time	Incident Events
May 20 (Cont.)	Now, the commander tells the PSYOP and CA to actually go to the parents, to the homes, to talk about security. This was done: they went to about 80 homes, plus homes of the leaders. Some more students started riding the buses: number of riders went up to 20-30 per day, for the rest of the year.
June 12	Last day of school.

Postscript.

The following year, regular school escorts were still stopped, but a random escort program was added. These ranged from armored vehicles front and back plus a guard on the bus, down to no armored vehicles and no guard. These random escorts were combined with random additional patrols (vehicle and foot mobile patrols scattered all along the convoy route) so the locals would see KFOR both on the street or in nearby vehicles. All together, this has provided some reassurance, and ridership is back to normal. However, complaints are now being directed to the UN that the U.S. is unresponsive and that the UN needs to either conduct the escorts or direct the U.S. to beef them up.