Protocols for Cognitive Task Analysis

Robert R. Hoffman

Florida Institute for Human and Machine Cognition

Beth Crandall and Gary Klein

Klein Associates Division of ARA

Debra G. Jones and Mica R. Endsley

SA Technologies

This document is intended to empower people in the conduct of Cognitive Task Analysis and Cognitive Field Research.

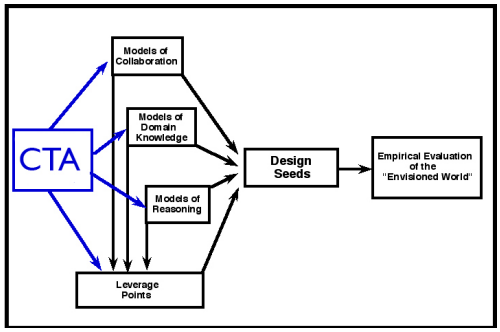
All rights reserved. Copyright 2008 by R. R. Hoffman, B. Crandall, G. Klein, D. G. Jones and M. R. Endsley. Individuals or organizations wishing to use or cite this material should request permission from:

Robert R. Hoffman, Ph.D.,
Institute for Human & Machine Cognition
40 South Alcaniz St.
Pensacola, FL 32502-6008
rhoffman@ihmc.us

Cite as:
Hoffman, R.R., Crandall, B., Klein, G., Jone, D., & Endsley, M. (2008). "Protocols for Cognitive Task Analysis." Technical Report, Florida Institute for Human and Machine Cognition, Pensacola, FL.

This document was prepared as a companion to *Working Minds: A Practitioner's Guide to Cognitive Task Analysis* by B. Crandall, G. Klein, G., & R. R. Hoffman. Cambridge MA: MIT Press, 2005. The CTA protocols are best conducted through an understanding of procedural discussions in *Working Minds*.

This document was prepared with the support of the Advanced Decision Architectures Collaborative Technology Alliance, sponsored by the US Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009.



1. Introduction
2. Bootstrapping Methods
    2.1. Documentation Analysis
    2.2. The Recent Case Walkthrough
    2.3. The Knowledge Audit
    2.4. Client Interviews
3. Proficiency Scaling
    3.1. General Introduction
    3.2. Career Interviews
    3.3. Sociometry
        3.3.1. General Protocol Notes
        3.3.2. General Forms
        3.3.3. Ratings Items
        3.3.4. Sorting Items
        3.3.5. Communication Patterns
            3.3.5.1. Within-groups
            3.3.5.2. Within-domains
    3.4. Cognitive Styles Analysis
4. Workplace Observations and Interviews
    4.1. General Introduction
    4.2. Workspace Analysis
    4.3. Activities Observations
    4.4. Locations Analysis
    4.5. Roles Analysis
    4.6. Decision Requirements Analysis
    4.7. Action Requirements Analysis
    4.8. SOP Analysis
5. Modeling Practitioner Reasoning
    5.1. Protocol Analysis
        5.1.1. General Introduction
        5.1.2. Materials
        5.1.3. Coding Schemes
        5.1.4. Effort
        5.1.5. Coding Schemes Examples
        5.1.6. Coding Verification
        5.1.7. Procedural Variations and Short-cuts
        5.1.8. A Final Word
    5.2. Cognitive Modeling Procedure
        5.2.1. Introduction and Background
        5.2.2 Protocol Notes
            5.2.2.1. Preparation
            5.2.2.2. Round 1
            5.2.2.3. Round 2
            5.2.2.4. Round 3
            5.2.2.5. Products
        5.2.3. Templates
            5.2.3.1. Round 1
            5.2.3.2. Round 2
            5.2.3.3. Round 3
6. The Critical Decision Method
    6.1. General Description of the CDM Procedure
        6.1.0. Preparation
        6.1.1. Step One: Incident Selection
        6.1.2 Step Two: Incident Recall
        6.1.3. Step Three: Incident Retelling
        6.1.4. Step Four: Timeline Verification and Decision Point Identification
        6.1.5. Step Five: Deepening
        6.1.6. Step Six: "What-if" Queries
    6.2. Protocol Notes for the Researcher

6.3. Boiler Plate Data Collection Forms
6.3.0. Cover Form
6.3.1. Step One: Incident Selection
6.3.2. Step Two: Incident Recall
6.3.3. Step Three: Incident Retelling
6.3.4. Step Four: Timeline Verification and Decision Point Identification
6.3.5. Step Five: Deepening
6.3.6. Step Six: "What-if" Queries
6.3.7. Final Integration
6.3.8. Decision Requirements Table
7. Goal-Directed Task Analysis
7.1. Introduction
7.2. Elements of the GDTA
7.2.1. Goals
7.2.2 Decisions
7.2.3. Situation Awareness Requirements
7.3. Documenting the GDTA
7.3.1. Goal Hierarchy
7.3.2. Relational Hierarchy
7.3.3. Identifying Goals
7.3.4. Defining Decisions
7.3.5. Delineating SA requirements
7.4. GDTA Methodology
7.4.1. Step 1: Review Domain
7.4.2. Step 2: Initial interviews
7.4.3. Step 3: Develop the goal hierarchy
7.4.4. Step 4: Identify decisions and SA requirements
7.5.5. Step 5: Additional interviews / SME review of GDTA
7.5.6. Step 6: Revise the GDTA
7.5.7. Step 7: Repeat steps 5 & 6
7.5.9. Step 8: Validate the GDTA
7.5. Simplifying the Process
7.6. Conclusion
8. References

1. Introduction

The purpose of this book to lay out guidance, including sample forms and instructions, for conducting a variety of Cognitive Task Analysis and Cognitive Field Research methods (including both knowledge elicitation and knowledge modeling procedures) that have a proven track record of utility in service of both expertise studies and the design of new technologies.

Our presentation of guidance includes protocol notes. These are ideas, lessons learned, and cautionary tales for the researcher. We also present Templates—prepared forms for various tasks.

We describe methods useful in boostrapping, which involves learning about the domain and the needs of the clients. We describe methods used in proficiency scaling and procedures for modeling knowledge and reasoning. We describe methods coming from the academic laboratory (e.g., Protocol Analysis), methods coming from applied research (e.g., Sociometrty) and methods coming from Cognitive Anthropology (e.g., Workpatterns Observations).

Along the way we present lessons learned and rules of thumb that arose in the course of conducting of a number of projects, both small-scale and large-scale. The guidance offered here spans the entire gamut—going all the way from general advice about records-keeping, to detailed advice about audio recording knowledge elicitation interviews; going all the way from general issues involved in protocol analysis, to detailed advice about the problems and obstacles involved in managing large numbers of multi-media resources for knowledge models.

Supporting material, including general guidance for the conduct of interviews and detailed guidance for the Critical Decision Method, can be found in: Crandall, B., Klein, G., & Hoffman R. R. (2006). *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. Cambridge, MA: MIT Press.

2. CTA Boostrapping Methods Protocols and Templates

In many projects that use Cognitive Work Analysis methodologies, one of the first steps is bootstrapping, in which the analysts familiarize themselves with the domain.

2.1. Documentation Analysis

Bootstrapping typically involves an analysis of documents (manuals, texts, technical reports, etc.). The literature of research and applications of knowledge elicitation converges on the notion that documentation analysis can be a critically important method, one that should be utilized as a matter of necessity (Hoffman, 1987; Hoffman, Shadbolt, Burton & Klein, 1995). Documentation analysis is relied upon heavily in the Work Domain Analysis phase of Cognitive Work Analysis (see Hoffman & Lintern, 2005; Vicente, 1999).

Documentation Analysis can be a time-consuming process, but can sometimes be indispensable in knowledge elicitation (Kolodner, 1983). In a study of aerial photo interpreters (Hoffman, 1987), interviews about the process of terrain analysis began only after an analysis of the readily-available basic knowledge of concepts and definitions. To take up the expert's time by asking questions such as "What is limestone?" would have made no sense.

Although it is usually considered to be a part of bootstrapping, documentation analysis invariably occurs throughout the entire research programme. For example, in the weather forecasting case study (Hoffman, Coffey, & Ford, 2000), the bootstrapping process focused on an analysis of published literatures and technical reports that made reference to the cognition of forecasters. That process resulted in the creation of the project guidance document. However, documentation analyses of other types occurred throughout the remainder of the project—analysis of records of weather forecasting case studies, analysis of Standard Operating Procedures documents, analysis of the Local forecasting Handbooks, etc.

Analysis of documents may range from underlining and note-taking to detailed propositional or content analysis according to functional categories Documentation Analysis can be conducted for a number of purposes or combinations of purposes. Rather than just having information flow from documents into the researcher's understanding, the analysis of the documents can involve specific procedures that generate records or analyses of the knowledge contained in the documents. Documentation Analysis can:

- Be suggestive of the reasoning of domain practitioners, and hence contribute to the forging of reasoning models,
- Be useful in the attempt to construct knowledge models, since the literature may include both useful categories for domain knowledge and important specific "tid-bits" of domain knowledge. In some domains, texts and manuals contain a great deal of specific information about domain knowledge.
- Be useful in the identification of leverage points—aspects of the work where even a modest infusion of or improvement in technology might result in a proportionately greater improvement in the work.

Often, "Standard Operating Procedures" documents are a source of information for bootstrapping. However, if one wants to use an analysis of SOP documents for other purposes, (e.g., as a way of verifying the Decision Requirements analysis), one may hold off on even looking at SOP documents during the bootstraping phase. Therefore, we include the protocol and template for SOP analysis later in this document, in the category of "Workplace Observations and Interviews.

The uses and strengths of documentation analysis notwithstanding, the limitations must also be pointed out. Practitioners possess knowledge and strategies that do not appear in the documents and task descriptions and, in some cases at least, could not possibly be documented. Even in Work Domain Analysis, which is heavily oriented towards Documentation Analysis, interactions with experts are used to confirm and refine the Abstraction-Decomposition Analyses. In the weather forecasting project (Hoffman, Coffey, & Carnot, 2000) an expert told how she predicted the lifting of fog. She would look out toward the downtown and see how many floors above ground level she could count before the floors got lost in the fog deck. Her reasoning relied on a heuristic of the form, "If I cannot see the 10th floor by 10AM, pilots will not be able to take off until after lunchtime." Such a heuristic had great value, but is hardly the sort of thing that would be allowed in a formal standard operating procedure document. Many observations have been made of how engineers in process control bend rules and deviate from mandated procedures so that they can more effectively get their jobs done (see Koopman & Hoffman, 2003). We would hasten to generalize by saying that all practitioners who work in complex sociotechnical contexts possess knowledge and reasoning strategies that are not captured in existing procedures or documents, many of which represent (naughty) departures from what those experts are supposed to do or to believe (Johnston, 2003; McDonald, Corrigan, & Ward, 2002).

2.2. The Recent Case Walkthrough

2.2.1. Protocol Notes

The Recent Case Walk-through is an abbreviated version of the Critical Decision Method, and is primarily aimed at:

- Bootstrapping the Analyst into the domain
- Establishing rapport with the Participant

The method can also lead to the identification of potential leverage points, a tentative notion of practitioner styles, or tentative ideas about other aspects of cognitive work.

2.2.2. Template

"The Recent Case Walkthrough"

Interviewer	
Participant	
Start time	
Finish time	

Instructions
Please walk me through your most recent problem. When was it? What were the goals you had to achieve? What did you do, step by step?

Narrative
(enlarge this cell as needed)

Build a timeline	
TIME	**EVENT**
	(Duplicate and expand these rows as needed)

Revise the Timeline

Instructions: Now let's go over this timeline. I'd like you to suggest corrections and additions.

Time	**Event**
	(Duplicate and expand these rows as needed)

Deepening
Instructions: Now I'd like to go through this time line a step at a time and ask some questions.

Information	What information did you need or use to make this judgment? Where did you get this information? What did you do with this information?

Mental Modeling	As you went through the process of understanding this situation, did you build a conceptual of the problem scenario? Did you try to imagine the important causal events and principles? Did you make a spatial picture in your head? Can you draw me an example of what it looks like?

Knowledge Base	In what ways did you rely on your knowledge of this sort of case? How did this case relate to typical cases you have encountered? How did you use your knowledge of typical patterns?

Guidance	At what point did do you look at any sort of guidance? Why did you look at the guidance? How did you know if you could trust the guidance? How do you know which guidance to trust? Did you need to modify the guidance? When you modify the guidance, what information do you need or use? What do you do with that information?

Legacy	What forms do you have to complete or specific products you have to produce?

Procedures	

Deepening Form

Timeline Entry	Probe	Response
	(Duplicate and expand these rows as needed)	

2.3. The Knowledge Audit

2.3.1. Template

"The Knowledge Audit"

Interviewer	
Participant	
Start time	
Finish time	

Your work involves a need to analyze cases in great detail. But we know that experts are also able to maintain a sense of the "big picture." Can you give me an example of what is important about the big picture for (your, this) task? What are the major elements you have to know and keep track of?
(Expand these rows, as needed)

When you are conducting (this) task, are there ways of working smart, or accomplishing more with less--ways that you have found especially useful?

What are some examples when you have improvised on this task, or noticed an opportunity to do something more quickly or better, and then followed up on it?"

Have there been times when the data or the equipment pointed in one direction, but your own judgment tells you to do something else? Or when you had to rely on experience to avoid being led astray by equipment or the data?

Can you think of a situation when you realized to that you had change what you were doing in order to get the job done?

Can you recall an instance where you spotted a deviation from the norm or recognized that something

was amiss?

Have you ever had experiences where part of a situation just "popped" out at you?--a time when you walked into a situation and knew exactly how things got there and where they were headed?

Experts are able to detect cues and see meaningful patterns that novices cannot see. Can you tell me about a time when you predicted something that others did not, or where you noticed things that others did not catch? Were you right? Why (or why not)?

Experts can predict the unusual. I know this happens all the time, but my guess is that the more experience you have, the more accurate your seemingly unusual judgments become. Can you remember a time when you were able to detect that something unusual happened?

2.4. Client Interviews

2.4.1. Protocol Notes

The "clients" are the individuals or organizations that use the products that come from the domain practitioners (or their organizations). For example, new technologies might be designed to assist weather forecasters (who would be the "end-users" of the systems) whose job it is to produce forecasts to assist aviators. The aviators would be the "clients."

One might make a great tool to help weather forecasters, but it if does not help the forecasters help *their* clients, it may end up an instance of how good intentions can lead to hobbled technologies.

For many projects in which CTA plays a role, it is important to identify Client's needs (for information, planning, decision-making, etc.), and understand the differing needs and requirements of different Clients.

Some "Tips"
Keep an eye out for inadequacies in the standard forms, formats, and procedures used in the User-Client exchanges. Keep an eye out for adaptation (including local kluges and work-arounds). It can be useful to use Knowledge Audit probes in the Client Interviews. One need not use all of the probe questions that are included in the Template, below.

2.4.2. Template

"Client Interviews"

Analyst:	
Participant	
Date	
Start time	
Finish time	
Participant Age	
Participant Job designation (or duty assignment), and title (or rank)	

How many years have you been _____ ? What kinds of projects have you been involved in?
(Expand the cells, as needed)

Have you had any training in (provider's domain)?

What is the range of experience you have had at (provider's domain or products)?

What is your current (job/project)?

Can you give me an example of a project where (provider information/services/products) had a

big impact?

Can you give me an example of a project where the provider's (products/information/services) made a big difference in success or failure?

What is it that you like about the provider's (products/information/services) that is provided to you?

Can you give me an example of a project where you were highly confident in the provider's (products/services/information)?

Can you give me an example of a project where you were highly skeptical of the provider's (products/services/information)?

How do you use the provider's (products/services/information) during your projects?

Have you ever been in a situation where you felt that you understood the (provider's domain) better than the provider?

Have you ever felt that you needed a kind of (product/service/information) that you are not given?

In your opinion, what characterizes a good provider?

What is the value of having providers who are highly skilled?

Can you think of any things you could do to help the providers get better?

3. Proficiency Scaling Methods Protocols and Templates

3.1. General Introduction

The psychological and sociological study of interpersonal judgments and attributions dates to the early days of both fields (e.g., Thorndike, 1918; see Mieg, 2001). In the past two decades, the areas of “Knowledge Elicitation” and “Expertise Studies” have been focused specifically on proficiency scaling, primarily for the purpose of identifying experts. The rich understanding of expert-level knowledge and reasoning provides a "gold standard" for domain knowledge and reasoning strategies. The rich understanding of journeymen, apprentices, and novices informs the researcher concerning training needs, the adaptation of software systems to individual user models, etc.

In parallel, the areas of the Sociology of Scientific Knowledge and Cognitive Anthropology, have been focusing some effort on studies of work in the modern technical workplace and thus have brought to bear methods of ethnography. This has also entailed the study of proficient skill. This paradigm has been referred to as the study of "everyday cognition" and "cognition in the wild" (Hutchins, 1995a,b; Lave, 1988; Scribner, 1984; Suchman, 1993). Researchers have studied a great many domains of proficiency, including traditional farming methods used in Peru, photocopier repair, the activities of military command teams, the social interactions of collaborative teams of designers, financiers, climatologists, physicists, and engineers, and so on (Collins, 1985; Cross, Christians, and Dorst, 1996; Fleck and Williams, 1996; Greenbaum and Kyng, 1991; Hutchins, 1995; Lynch, 1991; Orr, 1985). Classic studies in this paradigm include:

- Hutchins' studies of navigation (by both individuals and teams, by both culturally-traditional methods and modern technology-based methods) (1990) on the interplay of navigation tools and collaborative activity in maritime navigation,
- Hutchins' (1995b) study of flight crew collaboration and team work,
- Orr's study of how photocopier repair technicians acquire knowledge and skill by sharing their "war stories" with one another.
- Lave's (1988) studies of traditional methods used in crafts such as tailoring, and the nature of skills used in everyday life (e.g., math skills used during supermarket shopping).

Research in the area of the "sociology of scientific knowledge" (e.g., Collins, 1997; Fleck and Williams, 1996; Knorr-Cetina, 1981; Knorr-Cetina and Mulkay, 1983) also falls within this paradigm, and has involved detailed analyses of science as it is practiced, e.g., the history of the development of new lasers and of gravity wave detectors, the role of consensus in decisions about what makes for proper practice, breakdowns in the attempt to transmit skill from one laboratory to another, failures in science or in the attempt to apply new technology, etc.

In psychology and cognitive science, concern with the question of how to define expertise (Hoffman, 1998) led to an awareness that determination of who is an expert in a given domain can require its own empirical effort. In a *proficiency scaling* procedure, the researcher determines a domain-and organizationally appropriate scale of proficiency levels. *Proficiency scaling* is the attempt to forge a domain- and organizationally-appropriate scale for distinguishing levels of proficiency. Some alternative methods are described in Table 3.1.

Table 3.1. Some methods that can contribute data for the creation of a proficiency scale.

Method	Yield	Example
In-depth career interviews about education, training, etc.	Ideas about breadth and depth of experience; Estimate of hours of experience	Weather forecasting in the armed services, for instance, involves duty assignments having regular hours and regular job or task assignments that can be tracked across entire careers. Amount of time spent at actual forecasting or forecasting-related tasks can be estimated with some confidence (Hoffman, 1991).
Professional standards or licensing	Ideas about what it takes for individuals to reach the top of their field.	The study of weather forecasters involved senior meteorologists US National Atmospheric and Oceanographic Administration and the National Weather Service (Hoffman, Coffey, and Ford, 2000). One participant was one of the forecasters for Space Shuttle launches; another was one of the designers of the first meteorological satellites.
Measures of performance at the familiar tasks	Can be used for convergence on scales determined by other methods.	Weather forecasting is again a case in point since records can show for each forecaster the relation between their forecasts and the actual weather. In fact, this is routinely tracked in forecasting offices by the measurement of "forecast skill scores" (see Hoffman and Trafton, forthcoming).
Social Interaction Analysis	Proficiency levels in some group of practitioners or within some community of practice (Meig, 2000; Stein, 1997)	In a project on knowledge preservation for the electric power utilities (Hoffman and Hanes, 2003), experts at particular jobs (e.g., maintenance and repair of large turbines, monitoring and control of nuclear chemical reactions, etc.) were readily identified by plant managers, trainers, and engineers. The individuals identified as experts had been performing their jobs for years and were known among company personnel as "the" person in their specialization: *"If there was that kind of problem I'd go to Ted. He's the turbine guy."*

If research or domain practice is predicated on any notion of expertise (e.g., a need to build a new decision support system that will help experts but can also be used to teach apprentices), then it is necessary to have some sort of empirical anchor on what it really means for a person to be an "expert," say, or an "apprentice." A proficiency scale should be based on more than one method, and ideally should be based on at least three. We refer to this as the “three legs of a tripod.”

Social Interaction Analysis, the result of which is a sociogram, is perhaps the lesser known of the methods. A sociogram, that represents interaction patterns between people (e.g., frequent interactions), is used to study group clustering, communication patterns, and workflows and processes. For Social Interaction Analysis, multiple individuals within an organization or a community of practice are interviewed. Practitioners might be asked, for example, "If you have a problem of type x, who would you go to for advice?" or they might be asked to sort cards bearing the names of other domain practitioners into piles according to one or another skill dimension or knowledge category.

Hoffman, Ford and Coffey (2000) suggested that proficiency scaling for a given project should be based on at least three of the methods listed in Table A.1. It is important to create a scale that is both domain and organizationally-appropriate, and that considers the full range of proficiency. A rule of thumb in psychology is that it takes 10 years of intensive practice to become an expert (Ericsson, Charness, Feltovich, and Hoffman, 2005). Assuming six hours of concentrated practice per day for 40 weeks per year, this would mean 12,000 hours of practice at domain tasks. while convenient, such rules of thumb are just that, rules of thumb. in the project on weather forecasting (Hoffman, Coffey and Ford, 2000), the proficiency scale distinguished three levels: experts, journeymen, and apprentices, each of which was further distinguished by three levels of seniority, up to the senior experts who has as much as 50,000 hours of practice and experience at domain tasks.

Following Hoffman (1988; Hoffman, Shadbolt, Burton, and Klein, 1995; Hoffman, 1998) we rely on a modern variation on the classification scheme used in the craft guilds (Renard, 1968).

Table 3.2. Basic proficiency categories (adapted from Hoffman, 1998).

Naive	One who is totally ignorant of a domain.
Novice	Literally, someone who is new—a probationary member. There has been some ("minimal") exposure to the domain.
Initiate	Literally, someone who has been through an initiation ceremony—a novice who has begun introductory instruction.
Apprentice	Literally, one who is learning—a student undergoing a program of instruction beyond the introductory level. Traditionally, the apprentice is immersed in the domain by living with and assisting someone at a higher level. The length of an apprenticeship depends on the domain, ranging from about one to 12 years in the craft guilds.
Journeyman	Literally, a person who can perform a day's labor unsupervised, although working under orders. An experienced and reliable worker, or one who has achieved a level of competence. It is possible to remain at this level for life.
Expert	The distinguished or brilliant journeyman, highly regarded by peers, whose judgments are uncommonly accurate and reliable, whose performance shows consummate skill and economy of effort, and who can deal effectively with certain types of rare or "tough" cases. Also, an expert is one who has special skills or knowledge derived from extensive experience with subdomains.
Master	Traditionally, a master is any journeyman or expert who is also qualified to teach those at a lower level. Traditionally, a master is one of an elite group of experts whose judgments set the regulations, standards, or ideals. Also, a master can be that expert who is regarded by the other experts as being "the" expert, or the "real" expert, especially with regard to sub-domain knowledge.

This is useful first step, in that it adds some explanatory weight to the category labels, but needs refinement when actually applied in particular projects. In the weather forecasting project (Hoffman, Coffey and Ford, 2000) it was necessary to refer to Senior and Junior ranks within the expert and journeyman categories. (For details, see Hoffman, Trafton, and Roebber, 2009).

The Table entries fall a step short of being operational definitions. They are suggestive, however, and this is where the specific methods of proficiency scaling come in: Career evaluations (hours of experience, breadth and depth of experience), social evaluations and attributions (sociograms), and performance evaluations.

Mere years of experience does not guarantee expertise, however. Experience scaling can also involve an attempt to gauge the practitioners' breadth of experience—the different contexts or sub-domains they have worked in, the range of tasks they have conducted, etc. Deeper experience (i.e., more years at the job) affords more opportunities to acquire broader experience. Hence, depth and breadth of experience should be correlated. But they are not necessarily correlated, and so examination of both breadth and depth of experience is always wise.

Thus, one comes to a resolution that age and proficiency are generally related, that is, partially correlated. This is captured in Table 3.3.

Table 3.3. Age and proficiency are, generally speaking, partially correlated.

	LIFESPAN					
	10	20	30	40	50	60
Novice	Learning an commence at any age → → → (e.g., the school child who is avid about dinosaurs; the adult professional who undergoes job re-training)				Achievement of expertise in significant domains may not be possible	
Intern and Apprentice	Individuals less than18yrs of age are rarely considered appropriate as Apprentices	Can commence at any age → → →				Achievement of expertise in significant domains may not be possible
Journeyman		It is possible to achieve journeyman status (e.g., chess, computer hacking, sports)	Is typically achieved in mid- to late-20s, but development may go no further			
Expert Stage		It is possible to achieve expertise (e.g., chess, computer hacking, sports)		Most typical for 35 yrs of age and beyond		
Master Stage			Is rarely achieved early in a career	It is possible to achieve mid-career	Most typical of seniors	

To summarize, a rich understanding of expert-level knowledge and reasoning provides a "gold standard" for domain knowledge and reasoning strategies. The rich understanding of journeymen, apprentices, and novices informs the researcher concerning training needs, the adaptation of software systems to individual user models, etc.

A number of CTA methods can support the process of proficiency scaling. Proficiency scaling is the attempt to forge a domain- and organizationally-appropriate scale for distinguishing levels of proficiency. A proficiency scale can be based on:

1. Estimates of experience extent, breadth, and depth based on results from a Career Interview,
2. Evaluations of performance based on performance measures,
3. Ratings and skill category sortings from a sociogram.

If research is predicated on any notion of expertise (e.g., a need to build a new decision support system that will help experts but can also be used to teach apprentices), then it is necessary to have some sort of empirical anchor on what it really means for a person to be an "expert," say, or an "apprentice." Ideally, a proficiency scale will be based on more than one method.

3.2. Career Interviews

3.2.1. General Protocol Notes

Important things the Interviewer needs to accomplish:

- Convey respect for the Participant for their job and the contribution their job represents.
- Convey respect for the Participant for their skill level and accomplishments.
- Convey a desire to help the Participant by learning about their goals and concerns.
- Mitigate any concerns that the Participant might have that the Interviewer is there to finger-point or disrupt the work, or broker people's agendas.
- Prove that the Interviewer has done some homework, and knows something about the Participant's domain.

Important things the Interviewer needs to learn:

- What it takes for a practitioner in the domain to achieve expertise.
- The range of kinds of experience (breadth and depth) that characterize the most highly-experienced practitioners.

Acquire a copy of the participant's vita or resume and append it to the interview data.

Selection of Participants can depend on a host of factors, ranging from travel and availability issues to any need or requirement there might be to identify and interview experts. Whatever the criteria or constraints, it is generally useful to interview personnel who represent a range of job assignments, levels of experience, and levels of "local" knowledge. In most CTA procedures, it is necessary to have at least some Participants who conduct domain tasks on a daily basis. In most CTA procedures it is highly valuable to have Participants who can talk openly about their processes and procedures, seem to love their jobs, and have had sufficient experience to have achieved expertise at using the newest technologies that are in the operational context.

Lesson Learned

Practitioners who are verbally fluent and who seem to love their domain can sometimes be too talkative and can jump around story-telling and getting off on tangents, especially since in some cases they rarely have opportunities to talk about their domain and their experiences to interested listeners. They sometimes need to be kept on track by saying, e.g., "that's an interesting story—let's come back to that…"

Some "Tips"

- Until or unless each interviewer has achieves sufficient skill or has had sufficient practice with Participants in the domain at hand, interviews should be conducted by pairs of interviewers, who take separate notes.
- Expect that the interview guides will have to be revised, even several times, during_the course of the data collection phase.
- Think twice about audiotape recording the interviews—it can take a great deal of time to transcribe and code the data.
- For interviews that are audio recorded, do not conduct the interview in a room with ANY background noise (e.g., air conditioners, equipment, etc.) as this will significantly degrade the quality of the recordings. Consider using a y-connector allowing input from two lapel microphones, one clipped to the collar of the Participant, one to the collar of the Interviewer.
- After the recorder is started, preface it with a clear statement of the day, date, time, Participant name, and Researcher name, project name, and interview type.
- Be sure to ask the Participant to speak clearly and loudly.

It is often valuable, and sometimes necessary, to attempt to determine how much time a Participant has spent engaged in activities that arguably fall within the domain. One might focus on the amount of time the Participant has spent at some single activity. For instance, how much time do weather forecasters spend trying to interpret at infra-red satellite images? One might focus on something broader, such as how much time a forecaster has spent engage in all forecasting-related activities.

The rule of thumb from expertise studies is that to qualify as an expert one must have had 10 years full time experience. Assuming 40 hours per week and 40 weeks per year, this would amount to 16,000 hours. Another rule of thumb, based on studies of chess, is 10,000 hours of task experience (Chase and Simon, 1973). These rules of thumb will in all likelihood not apply in many applied research contexts. For instance, in the weather forecasting case study (Hoffman, Coffey, and Ford, 2000), the most senior expert forecasters had logged in the range of 40,000-50,000 hours of experience. The experience scale must always be appropriate for the domain, but also appropriate for and the particular organizational context.

When retrospecting about their education, training, and career experiences, the reflective Participant will recall details, but there will be some gaps and uncertainties. One strategy that has proven useful is to apply the "Gilbreth Correction Factor," which is simply the estimated total divided by two. In his classic "time and motion" studies of the details of various jobs, Gilbreth (1911, 1914) found that the amount of time people spend at tasks that are directly work-related is about half of the amount of time they spend "on the job." Gilbreth also found if the schedule of work and rest periods could be made, as we would say today, more human-centered, that work output could double. Gilbreth studied tasks that had a major manual component (e.g., the assembly of radios), the "factor of 2" kept cropping up again (e.g., estimated time savings if a bin holding a part could be moved closer to the worker). The GCF serves as a reasonable enough anchor, a conservative estimate of time-on-task, in comparison to the raw estimate from the Career Interview, which may be likely to be an over-estimate. Research in a number of domains

has shown that even with the GCF applied, estimates of time spent at domain tasks by the most senior practitioners often fall well above the 10,000 hour benchmark set for the achievement of expertise (see Hoffman and Lintern, 2005).

Mere years of experience does not guarantee expertise, however. Experience scaling also involves an attempt to gauge the practitioners' breadth of experience—the different contexts or sub-domains they have worked in, the range of tasks they have conducted, etc. Deeper experience (i.e., more years at the job) affords more opportunities to acquire broader experience. Hence, depth and breadth of experience should be correlated. But they are not necessarily correlated, and so examination of both breadth and depth of experience is always wise.

Thus, one comes to a resolution that age and proficiency are generally related. This is captured in Table 3.2.

Table 3.2. Age and proficiency are, generally speaking, partially correlated.

	LIFESPAN					
	10	20	30	40	50	60
Novice	Learning an commence at any age → → → (e.g., the school child who is avid about dinosaurs; the adult professional who undergoes job re-training)				Achievement of expertise in significant domains may not be possible	
Intern and Apprentice	Individuals less than18yrs of age are rarely considered appropriate as Apprentices	Can commence at any age → → →				Achievement of expertise in significant domains may not be possible
Journeyman		It is possible to achieve journeyman status (e.g., chess, computer hacking, sports)	Is typically achieved in mid- to late-20s, but development may go no further			
Expert Stage		It is possible to achieve expertise (e.g., chess, computer hacking, sports)		Most typical for 35 yrs of age and beyond		
Master Stage			Is rarely achieved early in a career	It is possible to achieve mid-career	Most typical of seniors	

3.2.2. Template

"Career Interviews"

Analyst:	
Participant	
Date	
Start time	
Finish time	

Name	
Age	
Title or rank or job designation or duty assignment	

High-school and college education
(Expand these cells, as needed)

Early interest in this domain?

Training/College/Graduate Studies

Post-Schoolhouse Experience

Attempt To Estimate Hours Spent At The Tasks

EXPERIENCE			
WHERE	WHEN	WHAT	TIME
		For each task/activity: Number of hours per day Number of days per week Number of weeks per year	
		TOTAL	

Notes
Determination: Qualification on the Determination:
Notes on Salient experiences and skills

3.3. Sociometry

3.3.1. General Protocol Notes

In the field of psychology there is a long history of studies attempting to evaluate interpersonal judgments of abilities and competencies (e.g., Thorndike, 1920). In recent times this, has extended to research on the extent to which expert judgments of the competencies of other experts conform to evidence from experts' retrospections about previously-encountered critical incidents, and the extent to which competency judgments, and competencies shown in incident records, can be used to predict practitioner performance (McClelland, 1998).

In the field of advertising there has been a similar thread, focusing on perceptions of the expertise of individuals who endorse products, and the relations of those perceptions to judgments of product quality and intention to purchase (e.g., Ohanian, 1990).

There is a similar thread in the field of sociology. Sociometry is a branch of sociology in which attempts are made to measure social structures and evaluate them in terms of the measures. Psychiatrist Jacob Moreno first put the idea of a "sociogram" forward in the 1930s (see Fox, 1987) as a proposal of a means for measuring and visualizing interpersonal relations within groups.

In the context of cognitive task analysis and its applications, Sociometry involves having domain practitioners make evaluative ratings about other domain practitioners. (see Stein, 1992, 1997). The method is simple to use, especially in the work setting. In its simplest form (see Stein, 1992, Appendix 2), the sociogram is a one-page form on which the participant simply checks off the names of other individuals from who the participant obtains information or guidance.

Stein (1992, Appendix 2) provides guidance on using techniques from social network analysis in the analysis of sociometric data (e.g., how to calculate who is "central" in a communication network).

Results from a sociogram can include information that can be useful in proficiency scaling, leverage point identification, and even cognitive modeling. They can be useful in exploring hypotheses about expertise attributions and communication patterns within organizations. For instance, one study (Stein, 1992) in a particular organization found that the number of years individuals had been in an organization correlated with the frequency with which they were approached as an information resource. However, the individual who was most often approached for information was not the individual with the most years in the organization.

A sociogram can be designed for use with practitioners who know one another and work together (e.g., as a team, within an organization, etc.) or it can be designed for use in the study of a "community of practice" in which each practitioner knows something about many others, including others with whom the practitioner has not worked.

Most sociograms, like social network analyses, focus on a single relation: Whom do you trust? From whom to you get information? With whom do you meet? And so on. There is no principled reason why a sociogrammetric interview, in the context of cognitive task analysis, might ask about a variety of social factors that contribute to proficiency attributions. The listings

of possible sociogram probe questions we present here is intended to illustrate a range of possibilities.

Questions in sociograms can ask for self-other comparisons in terms of knowledge and experience level, in terms of reasoning styles, in terms of ability, in terms of subspecialization, etc. On those same topics, questions in a sociogram can ask about patterns of consultation and advice-seeking.

Questions in a sociogram can involve a numerical ratings task, or a task involving the sorting of cards with practitioner names written upon them.

The template we present is intended as suggestive. A sociogram need not precisely follow the guidance that we present here. Questions can be combined in many ways, as appropriate to the goals of the research and other pertinent project constraints.

Asking practitioners who work together in an organization to make comparative ratings of one another can be socially awkward and can have negative repercussions. The Participant needs to feel that they can present their most honest and candid judgments while at the same time preserving confidentiality and anonymity. A solution is to have the Researcher who facilitates the interview procedure remain "blind" to the name and identity of the co-worker(s) that the Participant is rating.

Ideally, before the sociogram is conducted, the Researchers already have some idea of who the experts in an organization are. Ideally, the sociogram should involve all of the experts as well as journeyman and apprentices.

If the number of participants to be involved in a sociogram is greater than about seven, the conduct of a full Sociogram can become tedious for the Participant. (A glance ahead will show that each of the sociograms involves having each Participant answer a number of questions about each of the other Participants).

One solution to this problem is to have each Participant conduct the sociogram for only a subset of the other Participants, with the constraint that each participant is assessed by at least two others. Another solution is to divide the pool of Participants into two sets, conduct the sociograms within each of the two sets, and then randomly select n Participants from set 1 and n Participants from set 2, and conduct another sociogram in which they rate individuals whom they had ***not*** rated in the first sociogram. Ideally, the same proficiency scale assignments should arise out of the converging data sets.

As for all of the Templates we present, we suggest that the first page of the sociogram data collection form include the following:

3.3.2. General Forms Templates

Interviewer			
Participant			
Date			
Start time			
Finish time			

<u>For Within-Group Sociograms</u>

We would like to ask you some questions regarding:
(Name or ID number)
How long (months, years) have you known/worked with one another?

3.3.3. Ratings Scale Items

We would now like you to answer a set of specific questions that use a numerical rating scale.

Please utilize the full range of the scale.

Do not hesitate to say things that reflect either positively or negatively on you or your colleagues.

We are not here to point fingers, or challenge anyone's skill or authority.

All the data we collect will remain anonymous and confidential. That is, no one's name will be associated with particular results.

Please be open, honest, and candid.

Compared to you, what is their level or extent of experience? Check one of the boxes below.

They have had much more experience	They have had more experience	They have had somewhat more experience	Our levels of experience are about the same	They have had somewhat less	They have had less experience	They have had much less experience
7	6	5	4	3	2	1

Comments

(Expand the cells, as needed)

Compared to your own knowledge of your domain, what is their level of knowledge? Check one of the boxes below.

They know much more	They know more	They know somewhat more	Our levels of knowledge are about the same	They know somewhat less	They know less	They know much less
7	6	5	4	3	2	1

Comments

Compared to your level of ability at conducting your domain tasks, what is their level of

ability? Check one of the boxes below.

Their work is consistently better	Their work is usually better	Their work is often better	Their work is about as good as mine	Their work is often not as good	Their work is usually not as good	Their work is consistently not as good
7	6	5	4	3	2	1

Comments

Is s/he good at particular tasks within your domain than others? What types?

_____

_____

Compared to your level of ability at conducting *those kinds* of tasks, what is their ability level? Circle one of the numbers below

Their work is consistently better	Their work is usually better	Their work is often better	Their work is about as good as mine	Their work is often not as good	Their work is usually not as good	Their work is consistently not as good
7	6	5	4	3	2	1

Comments

Compared to your level of ability at rapidly integrating data and quickly sizing up a situation, what is their level of ability? Check one of the boxes below.

Much greater skill	Greater skill	Moderately greater skill	About the same skill level as me	Somewhat lower skill	Lower skill	Much lower skill
7	6	5	4	3	2	1

Comments

Compared to the general accuracy of your own products or results, what is the general accuracy of their products or results? Check one of the boxes below.

Much greater	Greater	Moderately greater	About the same as mine	Somewhat lower	Lower	Much lower
7	6	5	4	3	2	1
Comments						

Compared to your own degree of ability at communicating about your domain with others, what is their level of ability? Check one of the boxes below.

Much greater skill	Greater skill	Moderately greater skill	About the same skill level as me	Somewhat lower skill level	Lower skill	Much lower skill
7	6	5	4	3	2	1
Comments						

Using traditional craft guild terminology, what would you say is their degree of expertise? Circle one of the numbers below.

Master Sets the standards for performance in the profession.	**Expert** Highly competent, accurate, and skillful. Has specialized skills.	**Journeyman** Can work competently without supervision or guidance.	**Apprentice** Needs guidance or supervision; shows a need for experience and on-the-job training, but also some need for additional formal training.	**Initiate** Still shows a need for needs formal training.
1	**2**	**3**	**4**	**5**
Comments				

Working Style

What seems to be their “style" of reasoning or working? Can you place them somewhere on the continuum below? Use a check mark

Understand the domain events as a dynamic causal system; form conceptual models of what's going on.		They conduct their work on the basis of standard procedures, checklists, etc. rather than a deep understanding of the domain
←----------→		
Comments		

Reflective Practitioner or Journeyman? What seems to be their "style" of reasoning or working? Can you place them somewhere on the continuum below? Use a check mark.		
Are reflective, and question their own reasoning and assumptions, and are even critical of their own ideas.		Do not seem to be deep a thinker. They do not think about their own reasoning or question their strategies or assumptions.
←----------→		
Comments		

Leader or Follower? What seems to be their "style" of reasoning or working? Can you place them somewhere on the continuum below? Use a check mark		
Are they one of the visionaries or innovators?		They seem to refine and extend the pathfinding ideas of others.
←----------→		
Comments		

3.3.4. Sorting Task Items

Experimenter's Protocol
Participant is invited to write the names of 12 domain practitioners on 3x5 cards. Names can be selected by instruction or by Participant choice. "People whose names come to mind, people you know or know of." "People you have worked with." "People who represent a range of experience in the field." A card bearing the Participant's own name can be added into the pile of cards.

Using traditional craft guild terminology, what would you say is their degree of expertise? Sort into piles.				
Master Sets the standards for performance in the profession.	**Expert** Highly competent, accurate, and skillful. Has specialized skills.	**Journeyman** Can work competently without supervision or guidance.	**Apprentice** Needs guidance or supervision; shows a need for experience and on-the-job training, but also some need for additional formal training.	**Novice** Still shows a need for formal training.
Comments				

Place the cards into two piles, according to these two general styles.	
"Reflective Practitioner"	**"Proceduralist"**
They understand the domain at a deep conceptual level. They understand the important issues.	They conduct their work on the basis of standard procedures and have a superficial rather than a deep understanding of the domain. They do not seem to care much about underlying or outstanding issues.
Comments	

For each name, note any specialization or specialized areas of subdomain knowledge or skill.
Comments

Who are the real pathfinders or founders of the field? Sort into piles.	
Founders and pathfinders	**Those who refine and extend the seminal ideas of others**
Comments	

Who would you say are the real visionaries in the field? Sort into piles.	
Visionaries	**Those who refine and extend the innovations and visions of others**
Comments	

Compared to you, what is their level of skill at conducting domain procedures? Sort into piles.				
Much higher skill	**Higher skill**	**About the same as me**	**Less skill**	**Much less skill**
Comments				

Compared to you, what is their level or extent of experience? Sort into piles.				
Much more experience	**More experience**	**About the same level of experience as me**	**Less experience**	**Much less experience**
Comments				

Compared to your own knowledge of your domain or field, what is their level of knowledge? Sort into piles.				
They know much more	**They know more**	**Our levels of knowledge are about the same**	**They know less**	**They know much less**
Comments				

Compared to the general quality or accuracy of your own products or results, what is the general quality or accuracy of their products or results?				
Greater	**Somewhat greater**	**About the same as mine**	**Somewhat less**	**Less**
Comments				

Compared to your own degree of ability at communicating about your field both to people

within the field and those outside of the field, what is their level of ability?				
Much greater	**Greater skill**	**About the same skill level as me**	**Lower skill**	**Much lower skill**
Comments				

3.3.5. Communications Patterns Items

3.3.5.1. Within-Groups, Teams, or Organizations

We would like to ask you some questions regarding:
(Name or ID number)
How long (months, years) have you known/worked with one another?

How often do you collaborate in the conduct of tasks?
How often do you conduct separate tasks, but at the same time?
How many times in the last month have you gone to them for guidance or an opinion concerning the interpretation of data? (If "rarely" or "never," is this because of rank?)
Does s/he ever come to YOU for guidance or an opinion?
How many times in the last month have you gone to them for guidance or an opinion concerning the understanding of causes or events?
How many times in the last month have you gone to them for guidance or an opinion concerning the equipment you use or the operations you perform using the technology?
If "rarely" or "never," is this because of rank or status?
Does s/he ever come to YOU for guidance or an opinion?
How many times in the last month have you gone to them for guidance or an opinion concerning organizational operations and procedures? (If "rarely" or "never," is this because of rank?)
Does s/he ever come to YOU for guidance or an opinion?

3.3.5.2. Within-Domains

Who in the field of _____ has come to you for guidance or an opinion?
On what topics or issues has your opinion been solicited?
Do you know of_____?
How do you know of _____?
Do you personally know _____?
If so, how long have you known _____?
Have you worked or collaborated in any way with _____?
Have ***you*** gone to ***them*** for guidance or an opinion?
Does s/he ever come to you for guidance or an opinion?

3.4. Cognitive Styles Analysis

3.4.1. Introduction

In any given domain, one should expect to find individual differences in skill level, but also such things as motivation and style. Different practitioners prefer work problems in different ways at different points, for any of a host of reasons.

Proficiency and style can be related in many ways. Proficiency can be achieved via differing styles. Proficiency can be exercised via differing styles. On the other hand, proficiency takes time to achieve and thus the styles that characterize experts should be at least partly coincident with skill level.

The analysis of cognitive styles, as a part of Proficiency Scaling, can be conducted with reliance on more than one perspective. Ideas concerning cognitive style come from:

- The proficiency scale of the craft guilds. Cognitive style can be evaluated from the perspective that style is partly correlated with skill level. Some designations and descriptions are presented in Table 3.3.

- Experiences of applied researchers who have conducted CTA procedures, and research in the paradigm of Naturalistic Decision Making (Reference). Some styles designations and descriptions are presented in Table 3.4

The ideas presented in these Tables are offered as suggestions. The categorizations are not intended to be inclusive or exhaustive—not all practitioners will fall neatly into one or another of the categories. Any analysis of reasoning styles will have to be crafted so as to be appropriate to the domain at hand. Any given project may find itself in need of these, some of these, or other appropriate designations.

Table 3.3. Likely reasoning style as a function of skill level. (For the definitions of the levels, see Table 3.1.)

Low Skill Levels (Naïve, Initiate, Novice)
• Their performance standards focus on satisfying literal job requirements. • They rely too much on the guidance (e.g., out puts of computer models), and take the guidance at • They use a single strategy or a fixed set of procedures--they conduct tasks in a highly proceduralized manner using the specific methods taught in the schoolhouse. • Their data-gathering efforts are limited but they tend to say that they look at everything. • They do not attempt to link everything together in a meaningful way—they do not form visual, conceptual models or does not spend enough time forming and refining models. • They are reactive, and "chase the observations." • They are insensitive to contextual factors or local effects. • They cannot do multi-tasking without suffering from overload, becomes less efficient and more prone to error. • They are uncomfortable when there is any need to improvise. • They are unaware of situations in which data can be misleading, and do not adjust their judgments appropriately. • They are less adroit. • Things that are difficult for them to do are easy for the expert. • They are less aware of actual needs of the end-users. • They are not strategic opportunists.

Medium Skill Levels (Apprentice, Journeyman)
• Despite motivation, it is possible to remain at this level for life and not progress to the expert level. • They begin by formulating the problem but then like individuals at the Low Skill Levels, those at the Medium Skill Levels tend to follow routine procedures—they tend to seek information in a proceduralized or routine way, following the same steps every time. • Some time is spent forming a mental model. • Their reasoning is mostly at the level of individual cues, some ability to recognize cue configurations within and across data types. • They tend to place great reliance on the available guidance. • They tend to treat guidance as a whole, and not look to sub-elements—they focus on determining whether the guidance is "right" or "wrong." • Journeymen can issue a competent product without supervision but both Apprentices and Journeymen have difficulty with tough or unfamiliar scenarios.

High Skill Levels (Expert, Master)
• They have high personal standards for performance. • Their process begins by achieving a clear understanding of the problem situation. • They are always skeptical of guidance and know the biases and weakness of each of the various guidance products. • They look at agreement of the various types of guidance--features or developments on which the guidances agree. • They are flexible and inventive in the use of tools and procedures, i.e., the conduct action sequences based on a reasoned selection among action sequence alternatives. • They are able to take context and local effects into account. • They are able to adjust data in accordance with data age. • They are able to take into account data source location. • They do not make wholesale judgments that guidance is either "right" or "wrong." • They reason ahead of the data. • They can do multi-tasking and still conduct tasks efficiently without wasting time or resources. • They are comfortable improvising. • They can tell when equipment or data or data type is misleading; know when to be skeptical. • They create and use strategies not taught in school. • They form visual, conceptual models that depict causal forces. • They can use their mental model to quickly provide information they are asked for. • Things that are difficult for them to do may be either very difficult for the novice to do, or may be cases in which the difficulty or subtleties go totally unrecognized by the novice. • They can issue high-quality products for tough as well as easy scenarios. • They provide information of a type and format that is useful to the client. • They can shift direction or attention to take advantage of opportunities.

Table 3.4. Some Cognitive Styles Designations (based on Pliske, Crandall, and Klein, 2000).

SCIENTIST The "reflective practitioner" Thinks about their own strategies and reasoning, engages in "what if?" reasoning, questions their own assumptions, and is freely critical of their own ideas.		
STYLE	**AFFECT**	They are often "lovers" of the domain. They like to experience domain events and watch patterns develop. They are motivated to improve their understanding of the domain. They are likely to act like a Mechanic when stressed or when problems are easy.
	ACTIVITIES	They actively seek out a wide range of experience in the domain, including experience at a variety of scenarios. They show a high level of flexibility. They spend proportionately more time trying to understand problems, and building and refining a mental model. They are most likely to be able to engage in recognition-primed decision-making. They spend relatively little time generating products since this is done so efficiently.
CORRELATED SKILL OR PROFICIENCY LEVEL		They possess a high level of pattern-recognition skill. They possess a high level of skill at mental simulation. They possess skill at using a wide variety of tools. They understand domain events as a dynamic system. Their reasoning is analytical and critical. They possess an extensive knowledge base of domain concepts, principles and reasoning rules.

PROCEDURALIST Seem to lack the intrinsic motivation to improve their skills and expand their knowledge, so they do not move up to the next level. They do not think about their own reasoning or question their strategies or assumptions. However, in terms of performance and/or experience, they are journeyman on the proficiency scale.		
STYLE	**AFFECT**	They sometimes are lovers of the domain. They like to experience domain events and see patterns develop. They are motivated to improve their understanding of the domain.
	ACTIVITIES	They spend proportionately less time building a mental model. They engage in recognition-primed decision making less often. They are proficient with the tools they have been taught to use. They are less likely to understand domain events as a complex dynamic system. They see their job as having the goal of completing a fixed set of procedures, although these are often reliant on a knowledge base. They sometimes rely on a knowledge base of principles of rules, but this tends to be limited to types of events they have worked on in the past.
CORRELATED SKILL OR PROFICIENCY LEVEL		Typically, they are younger and less experienced.

MECHANIC They conduct their work on the basis of standard procedures and have a superficial rather than a deep

understanding of the domain. They do not seem to care much or wonder about underlying concepts or outstanding issues.		
STYLE	**AFFECT**	They are not interested in knowing more than what it takes to do the job; not highly motivated to improve. They are likely to be unaware of factors that make problems difficult.
	ACTIVITIES	They see their job as having the goal of completing a fixed set of procedures, and these are often not knowledge-based. They spend proportionately less time building a mental model. They rarely are able to engage in recognition-primed decision making.
CORRELATED SKILL OR PROFICIENCY LEVEL		They sometimes have years of experience. They possess a limited ability to describe their reasoning. They are skilled at using tools with which they are familiar, but changes in the tools can be disruptive.

DISENGAGED The most salient feature is emotional detachment and disengagement with the domain or organization.		
STYLE	**AFFECT**	They do not like their job. They do not like to think about the domain. They are very likely to be unaware of factors that make problems difficult.
	ACTIVITIES	They spend most of the time generating routine products or filling out routine forms. They spend almost no time building a mental model and proportionately much more time examining the guidance. They cannot engage in recognition-primed decision making.
CORRELATED SKILL OR PROFICIENCY LEVEL		They possess a limited knowledge base of domain concepts, principles, and reasoning rules. Skill is limited to scenarios they have worked in the past. Their products are of minimally acceptable quality.

In the Cognitive Styles analysis, the Researchers conduct a card-sorting task that categorizes the domain Practitioners according to perceived cognitive style. The procedure presupposes that all of the Researchers who will engage in the sorting task have had sufficient opportunity to interact with each of the Participants (e.g., opportunistic, unstructured interviews, initial workplace observations, etc.). The Researchers will have engaged in some interviews, perhaps unstructured and opportunistic ones, with the domain Practitioners. That is, the Researchers will have bootstrapped to the extent that they feel they have some idea of each Practitioner's style.

The result of the procedure is a consensus sorting of Practitioners according to a set of cognitive styles. Individually, the styles should seem both appropriate and necessary. As a set, the styles should seem appropriate and complete, at least complete with regard to the set's functionality in helping the Researchers achieve the project purposes or goals.

3.4.2. Protocol For Cognitive Styles Analysis

1. The Analysis is conducted after the Researchers have become familiar with each of the domain practitioners who are serving as Participants in the project. That is, the Researchers have already had opportunities to discuss the domain with the Participants whose styles will be categorized.
2. The Researchers formulate a set of what seem to be reasonable and potentially useful styles categories, having short descriptive labels. These may be developed by the Researchers, or may be adapted from the listing provided in this doccument.
3. Each Practitioner's name is written on a file card. Working independently, all of the Researchers sort cards according to style.
4. The Researchers review and discuss each other's sorts. They converge on an agreed-upon set of categories and descriptive labels.
5. The Researchers re-sort. The Results are examined for consensus or rate of agreement.
6. The Researchers come to agreement that those Practitioners attributed the same style are similar with regard to the style definition.
7. The Researchers come to agreement that there are no Practitioners who seem to manifest more than one of the styles. That is, the final set of styles categories allows for a satisfactory partitioning of the set of Participants.
8. Steps 3-6 may be repeated.

3.4.3. Protocol for Emergent Cognitive Styles Analysis

A variation on this task involves having the Researchers conduct the sorting task without worrying about creating any short descriptive labels for the perceived styles, and without worrying about devising a clear description of the styles. The procedure lets the categories "emerge" from the individual Researcher's perceptions and judgments without influence from a set of pre-determined styles descriptors.

1. Each Practitioner's name is written on a file card. Working independently, each of the Researchers deliberates concerning the style of each of the Practitioners,
2. The Researchers sorts cards according to those judgments. An individual Practitioner may be in a category alone, or may fall in a category with one or more others.
3. The Researchers review and discuss each other's sorts. They converge on an agreed-upon set of categories. These are expressed in terms of the theme or themes of each category.
4. Working independently, the Researchers re-sort. The results are examined for consensus or rate of agreement.
5. Only then are the categories are given descriptive labels.
6. The Researchers confirm that those Practitioners attributed the same style are similar with regard to the style definition.
7. The Researchers confirm that there are no Practitioners who seem to manifest more than one of the styles. That is, the final set of styles categories allows for a satisfactory partitioning of the set of Participants.

3.4.4. Protocol for Cognitive Styles-Proficiency/Skill Levels Comparison

Another use of Cognitive Styles Analysis is to aid in proficiency scaling. In this use, Cognitive Styles are compared to proficiency levels, since there should be some relations. For example, it would be more likely to find that individuals who are potentially designated as expert in a sociogram or a career interview or a performance evaluation would be more likely to manifest the style of the "Reflective Practitioner" than that of the "Proceduralist."

Proficiency data for comparison to those from a Cognitive Styles Analysis can themselves come from a sorting procedure.

1. Working individually, the Researchers sort cards containing the names of the practitioners according to expertise level (see Table 4.1) or along a continuum.
2. The Researchers compare each individual an individual in the next lowest level and an individual in the next highest level, to confirm the levels assignments.
3. The Researchers compare the results of the Levels sort and the Styles sort to determine whether there is any consensus on a possible relation of style to skill level.

3.4.5. Validation

Researcher(s) who did not participate in the sorting task can be provided with the Styles categories' labels and definitions and can engage in the styles sorting (and the skill levels comparison). Their styles sorting should match the sorting arrived at in the final group consensus. Their skill levels sorting should affirm the consensus on the relation of styles to skill level.

4. Workplace Observations and Interviews

4.1. General Introduction

Workspace analysis can focus on:

- The workspace in which groups work,
- The spaces in which individuals work,
- The activities in which workers engage,
- The roles or jobs,
- The requirements for decisions,
- The requirements for activities,
- The "standard operating procedures."

These alternative ways of looking at workplaces will overlap. For example, at a particular work group, the activities might be identified with particular locations where particular individuals work. Roles might be identified with particular work locations. And so on. The listing above is not meant to be a complete, universal, or even consistent "slicing of the pie." Indeed, there is considerable overlap in the Templates presented below. Our purpose in providing these alternative ways of empirically investigating cognitive work is intended to suggest options. Where appropriate, ideas from the Templates might be used, or combined in ways other than those we present here.

It must be kept in mind that ***no*** observation is unobtrusive. In arguing for the advantages of so-called unobtrusive observations, people sometimes assert that in the ideal case, the observer should be like a "fly on the wall." The analogy is telling, because generally speaking when someone is concentrating at work, a fly in the room, on the wall or otherwise, can be very distracting.

The goal is not to capture behavior that has not been influenced in any way by the presence of the Researcher/observer, but to capture *authentic* behavior on the part of the Worker. The Researcher/observer is far more likely to observe authentic behavior if he or she has to some extent already been accepted into the culture of the work than if he or she tries to be a "fly on the fall." Acceptance means that the Workers regard the Researcher/observer as informed, sincere, and oriented toward helping, and the Workers know that the Researcher/observer respects them for their experience and skill.

This understanding of observational methods means that the possibilities for observation include merging interviewing and observing. The analysis can be conducted with the assistance of an informant, who accompanies the Researcher/observer and answers questions.

It should not be assumed that the Observations or Interviews should be limited to the categories set out here. Furthermore, these generic categories may need to be modified to better fit the given work domain. Nor should it be assumed that the Template probe questions presented here are necessarily appropriate to the domain, organization, or workplace.

Finally, it should not be assumed that the probe questions presented in the Templates exhaust all of the questions that might be appropriate to the domain, organization, or workplace.

When probing concerning the deficiencies of a tool or workstation, the practitioner's initial response to the probe may be meditative silence. Only with some patience and re-probing will deficiencies be mentioned. Often, the practitioner will have discovered deficiencies but will not have explicitly thought about them. Re-probing can focus on make-work, work-arounds, action sequence alternatives, alternative displays, changes to the tool, etc. It is critical to ask about each piece of equipment and technology—whether it is legacy, how it is used, what is good about it, whether its use is mandated, etc.

As can be seen in the Templates, below, the Templates combine various probes in differing ways. The two following tables provide an overview of the sorts of questions that are addressed, looking across the alternative procedures and their Templates.

Possible Probes
What sub-tasks or action sequences are involved in this activity? What information does the practitioner need? Where does the practitioner get this information? What does the practitioner do with this information? What forms have to be completed? How does the practitioner recover when glitches cause problems? How does the practitioner do work-arounds? Is each piece of technology a legacy system or a mandated legacy system?

Phenomena to Look For
• Multi-tasking and multiple distractions? • Do procedures require preparation of reports having categories and descriptions with little meaning? • Are the categories and descriptions really relevant to job effectiveness? • Is the environment one in which it is hard to develop skill? • Do task demands match with the equipment? • Are apprentices or trainees overwhelmed?—have to work through tons of data? • What circumstances induce conservativism? • Do workers test their own processes so as to learn the extent of their capabilities? • Do workers have leeway to make risky decisions of other forms of opportunities to benefit from learning? • Does any mentoring occur? What are the opportunities, or the obstacles/impediments? • Is their routine work (such as reporting functions) that have to be performed and detract from learning or understanding? • Do tasks or duties force the worker to adopt one or another style or strategy?

The Roles Analysis and Locations analysis interviews are conducted for the purpose of specifying work patterns at the social level (information-sharing) and at the individual level (cognitive activities). To focus on Roles, the Participant is asked questions about each of the jobs

(duty assignments) that focus on the flow of information, such as "What information does this person need?" and, "What does s/he do with the information?" To focus on Locations, the Participant is asked questions about each of the tools (workstations), such as "What tasks are conducted here?," What is the action sequence involved?" and, "What makes the task difficult?"

Workplace, Locations and Roles Analyses, Activities Observations, and SOP Analyses involve variations on similar probe questions, but the purposes of the analyses can differ, even though all can culminate in Decision Requirements Tables (DRTs) and/or Action Requirements Tables (ARTs).

It is important that the Analyses be conducted in the work place, since artifacts in the workplace can serve as contextual cues to both the Participant and the Researcher.

The Importance of Opportunism

Once the Work Place Map has been prepared, every time the researchers visit the workplace they should bring with them a folder containing copies of the Work Place Map and also copies of the Template for the Activities Observations. At any moment there may be an opportunity to take notes about work patterns (e.g., information-sharing) or an observation that suggests leverage point. For example, during the weather forecasting case study (Hoffman, Coffey, and Ford, 2000), there were a number of occasions where pilots and forecasters made comments that were of use. On one occasion, a group of pilots was effectively stranded at the airfield due to inclement weather, and were gathered chatting in a hallway. This was taken as an opportunity to conduct an informal and unplanned interview about weather impact on air operations.

4.2. Workspace Analysis

4.2.1. Protocol Notes

The main goal of this analysis is to create a detailed representation of the workspace, to inform subsequent analyses of work patterns and work activities. The analysis consists "simply" of creating a detailed map of the workplace. We use scare quotes because a detailed map, drawn to scale, takes considerable care and attention to detail.

The effort includes:

- Photographing the workplace to any level of detail, including views of individual workspaces, photographs of physical spaces alone, photographs of individuals at work, and so on. The workplace should be photographed from a variety of perspectives. e.g., from the entranceways, from each desk or workstation looking toward the other desks, and so on. It is important to note all of the desks, individual worker's workplaces, the locations of cabinets, records, operating manuals, etc.

- Creating preliminary sketches map of the workplaces that are subsequently used to create a refined map of the workplace, noting any special features that are pertinent to the research goals (e.g., locations of information resources, obstacles to communication, etc.).

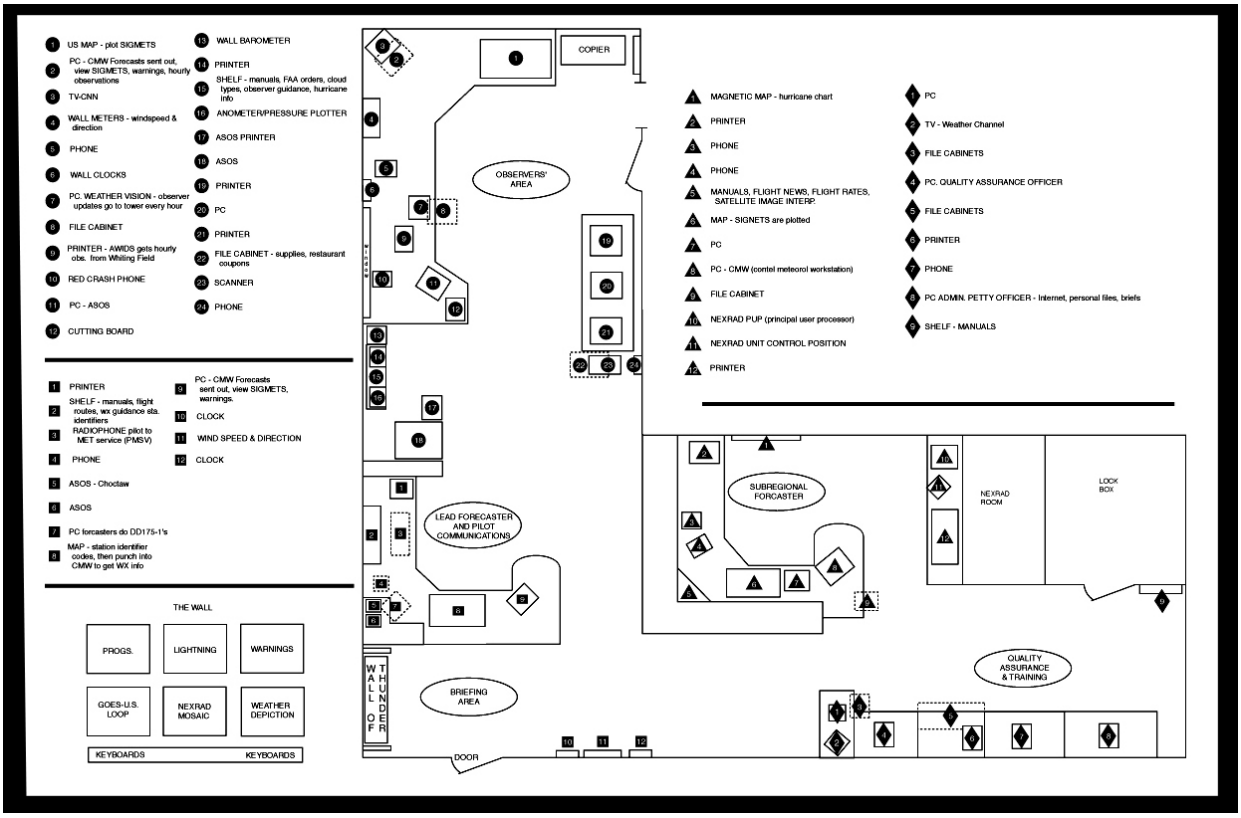
One never knows beforehand what things in the work place may seem unimportant at first but that may, in a later analysis, turn out to be very important. A detailed map and set of photographs can be repeatedly referred to throughout the research program, and mined for details that are likely to be missed at the time that the photographs and map are made. A simple device such as a paper cutter or a jumbled pile of reference manuals could turn out in a later analysis to be an important clue to aspects of the work.

Here are examples from the weather forecasting project (Hoffman, Coffey, and Ford, 2000):

- A particular workstation was adorned with numerous Post-It Notes© , serving as an indicator of user-hostility.
- One of the Participant interviews revealed information about problems forecasters were having with a particular software system. Subsequently, the photographs were examined and revealed that the forecasters had applied a relatively large number of Post-It Notes at the workstation, confirming the interview statements.
- Interviews with Participating expert forecasters revealed that some of them had kept files of hardcopies of images and other data regarding tough cases of forecasting that they had encountered. Using the workplace map, the researchers were able to create a second map showing the locations of the information resources, which in turn could be used in the study of work patterns.
- Photographs of the workplace showing the Participants at work showed how forecasters provided information to pilots, and revealed obstacles to communication and ways in which the workspace layout was actually a hindrance. The workplace was subsequently reconfigured.

The workplace map from the weather project is presented in the Figure 4.1 below. We present this figure to illustrate the level of detail that can be involved.

Figure 4.1. An example workplace map.



Related to the importance of subtle indicators, and the importance of conducting a photographic survey, it may be advisable to take photographs of the workplace be taken throughout the course of the investigation, if not at planned periodic intervals. The complex sociotechnical workplace is always a moving target, and changes made in the workplace can be important clues to the nature of the work. For instance, in the course of conducting the weather forecasting case study, the main forecasting workstation and its software were changed out completely.

Activities Observations, Roles analysis, Locations analysis, Decision Requirements analysis, Action Requirements analysis, and SOP analysis can all rely on the workplace map, and in many circumstances will rely heavily on it. See the comment above concerning opportunism.

4.3. Activities Observations

4.3.1. Protocol Notes

This simple and deceptively straight-forward protocol is for observing people at work. What is not so simple or straight-forward is deciding on what to observe and how to record the observations. This necessarily involves some sort of reasoned, specific, and functional means or categorizing or labeling work activities.

Like the Workspace Analysis, this procedure does not involve interviewing. This distinguishes Workspace Analysis and Activities Observations from the other Workplace Observations and Interviews procedures, which combine an element of observing with an element of interviewing.

4.3.2. Template

"Activities Observations"

Observer
Informant(s) (if any)
Date
Start time:
Finish time:

Observation Record

Time	**Actor**	**Activities**
		(Duplicate and expand the rows as needed.)

4.4. Locations Analysis

This analysis looks at the work from the perspective of individual locations (desks, cubicles, etc.) within the total workplace.

4.4.1. Template

"Work Locations Analysis"

Researcher/Observer
Participant/Informant
Date
Start time:
Finish time:

Instructions for the Participant
We would like to go through the workplace one work location at a time, and ask you some questions about it. This may take some time and we need not rush through it in a single session. Some of our questions may require a bit of thought. Feel free to take your time, and of course ask any questions of us that come to mind.

(Expand the cells and duplicate this table, as needed.)

Work Space (desk, workstation)
Location (see Workspace Map)
Work Space Layout (Researcher sketches or photographs)

Who works here?

Name	**Notes**
	(Duplicate and expand the rows, as needed.)

Tools and Technologies

Tool/Technologies	**Description, Uses, Notes**

Role or activities that are enacted at the work space (Expand the cells and duplicate this table, as needed.)

Activity
What are the goals of this activity?
What skills, knowledge, and experience are needed for successful accomplishment of this activity?
What information is needed for successful accomplishment of this activity?
What about this work space makes the goal easy to achieve?
What about the work space makes the goal difficult to achieve?
Kluges and work-arounds

4.5. Roles Analysis

4.5.1. Protocol Notes

The analysis of Roles can be aided by keeping on hand the finished workplace map, but even more important is to conduct the analysis in the workplace rather than outside of it.

For the Roles Analysis, a Participant/informant is asked about goals, tasks, needed knowledge, and technology.

Finally, it should not be assumed that the probe questions presented in the Template exhaust all of the questions that might be appropriate to the domain, organization, or workplace.

4.5.2. Template

"Roles (jobs, duty assignments)"

Researcher	
Participant/ Informant (if any)	
Date	
Start time	
Finish time	

Instructions for the Participant
We would like to go through the workplace one job assignment at a time and ask you some questions about it. This may take some time and we need not rush through it in a single session. Some of our questions may require a bit of thought. Feel free to take your time, and of course ask any questions of us that come to mind.

Role (job, post, or duty assignment)
Goals
Tasks
Needed skills, knowledge, and experience.
What makes the job easy?
What makes the job difficult?

4.6. Decision Requirements Analysis

4.6.1. Protocol Notes

This procedure is an interview for workpatterns analysis, and is conducted using Template Decision Requirements Tables (DRT) to provide interview structure.

For each DRT, the Participant is asked to confirm/flesh out the action sequence, confirm/flesh out the specification of support and tools, confirm/flesh out the notes on information needs, confirm/flesh out the notes on usability and usefulness, and to confirm/flesh out the notes on deficiencies.

The Interviewer makes notes right on the draft forms.

Following the Interview, the DRTs are re-drafted into their final form.

A DRT is an identification and codification of the important judgments and decisions that are involved in performing a particular task. In addition, the table captures the dynamic flow of decision-making.

In the analysis, one DRT is completed for each task or worker goal. The DRT specifies:

- The action sequence involved in the task,
- The equipment, tools, forms, etc. that are used to conduct the task,
- A specification of the information the person needs in order to conduct the task,
- Notes about what is good and useful about the support,
- Notes about ways in which the support makes the task unnecessarily difficult or awkward, or that might be improved.

As a consequence of these depictions, the DRT is intended to suggest new approaches to display design, workspace design, and work patterns design.

4.6.2. Template

"Decision Requirements Analysis"

Interviewer	Decision Designation or Identifying goal
Participant	
Date	
Start time	
Finish time	

What is the decision? What are the goals, purposes, or functions? Why does this decision have to be made?
How is the decision made? What information is needed?
What are the informational cues?
In what ways can the decision be difficult?
How do you assess the situation or broader context for this decision?
How much time or effort is involved in making this decision?
What is the technology or aid, and how does it help?
Are any work-arounds?
Are there any local "kludges" to compensate for workplace or technology deficiencies?
What kinds of errors can be made?
What kinds of additional aids might be useful?

When this decision has to be made, are there any hypotheticals or *"what ifs?"*
What are your options when making this decision?

4.7. Action Requirements Analysis

4.7.1. Protocol Notes

This procedure is an interview for workpatterns analysis, and is conducted using Template Action Requirements Tables (ART) to provide interview structure.

For each ART, the Participant is asked to confirm/flesh out the action sequence, confirm/flesh out the specification of support and tools, confirm/flesh out the notes on information needs, confirm/flesh out the notes on usability and usefulness, and to confirm/flesh out the notes on deficiencies.

The Interviewer makes notes right on the Template forms.

Following the Interview, the ARTs are re-drafted into their final form.

An ART is an identification and codification of the important activities that are involved in performing a particular task. In addition, the table captures the dynamic flow of activity

In the analysis, one ART is completed for each task or worker goal. The ART specifies:

- The action sequence involved in the task,
- The equipment, tools, forms, etc. that are used to conduct the task,
- A specification of the information the person needs in order to conduct the task,
- Notes about what is good and useful about the support,
- Notes about ways in which the support makes the task unnecessarily difficult or awkward, or that might be improved.

As a consequence of these depictions, the ART is intended to suggest new approaches to display design, workspace design, and work patterns design.

4.7.2. Template

"Action Requirements Analysis"

Interviewer	Task Designation or Identifying goal
Participant	
Date	
Start time	
Finish time	

What is the action sequence?
What cognitive activities are involved in this task/activity?
In what ways can the activity be difficult? What about the support or information depiction makes the action sequence difficult?
What are the informational cues? How are they depicted?
What is the technology or aid, and how does it help? What is good or useful about it?
Are any work-arounds?
Are there any local "kludges" to compensate for workplace or technology deficiencies?
What kinds of errors can be made?
What kinds of additional aids might be useful?

4.8. SOP Analysis

4.8.1. Protocol Notes

It is not unheard of for "Standard Operating Procedures" (SOP) documents to be a poor reflection of actual practice, for a variety of reasons. Workers may rely on short-cuts and work-arounds that they have discovered or created. SOPs may specify (sub)procedures that are not regarded as necessary. SOPs may simply be ignored, and the reasons for this might be informative. An SOP might be outdated (given that task requirements evolve) or ill-specified.

For these and other reasons, the analysis of SOP documents can be a window to the "true work."

The analysis of SOP documents depends on the cooperation of a highly-experienced domain practitioner who is willing to talk openly and candidly.

For each SOP, the Participant is asked the following questions:

1. Can you briefly describe that this procedure is really about? What are its goals and purposes; what is the basic action sequence?
2. Who does this procedure?
3. How often is the procedure conducted?
4. What is good or easy about the action sequence?
5. When you conduct the sequence, do you really do it in a way that departs from the specifications in the SOP? Are there short-cuts? Do they use a "cheater sheet?"
6. How often have you actually referred to the SOP document for guidance?

When probing concerning the deficiencies of a tool or workstation, the practitioner's initial response to the probe may be meditative silence. Only with some patience and re-probing will deficiencies be mentioned. Often, the practitioner will have discovered deficiencies but will not have explicitly thought about them. Re-probing can focus on make-work, work-arounds, action sequence alternatives, alternative displays, changes to the tool, etc. It is critical to ask about each piece of equipment and technology—whether it is legacy, how it is used, what is good about it, whether its use is mandated, etc.

In many projects that use Cognitive Work Analysis methodologies, one of the first steps is bootstrapping, in which the analysts familiarize themselves with the domain. Ordinarily, SOP documents would be a prime source of information for bootstrapping. However, if one wants to use the SOP Interview and analysis as a way of verifying the DRT analysis, one may hold off on even looking at SOP documents during the bootstrapping phase.

In the weather forecasting project (Hoffman, Coffeyand Ford, 2000), it took about 5 minutes to flesh out the answers for each SOP.

From the Interviewer's notes, formal SOP tables can be created. An example appears below in Table 4.2.

Table 4.2. An example completed SOP Table, for Terminal Air Forecast Verification

Number and Name	2110 TAF Verification
Description	Takes about 2 minutes. You input your own TAF and verify the previous one.
Who does this procedure?	Forecaster
How often?	Every 6 hours (four times per day) unless the field is closed.
Frequency of reference to the SOP	P has referred to this SOP a couple of times. Once you've done it a few times, you know it.
Comments	Is done on time about 80 percent of the time. Sometimes you get busy, you get tired. You forget to check the home page and when you do you see that GOES isn't downloading or something and so you get involved in fixing that, etc. The OPs Assistant looks at the TAF Verifications to prepare the skill scores.

Invariably, there will be some missing information in the SOP tables that are created after a single interview. Therefore, SOP Interviews generally have to be conducted in two waves, one to draft the SOP tables and the second to flesh them out and fine-tune them.

Another common circumstance is when the domain includes workers with special areas of proficiency or responsibility. Groups of tasks (and the corresponding SOPs) may be largely unknown to any one individual. In such cases, SOP Interviews will need to be conducted with more than one individual.

4.8.2. Template

Analysis of "Standard Operating Procedures" Documents

Interviewer
Participant
Date
Start time
Finish time

SOP Number and Name
What are its goals and purposes; what is the basic action sequence?
Can you briefly describe that this procedure is really about?
What are the purposes and goals of the procedure?
Who conducts this procedure?
How often?
What are the circumstances that trigger the procedure?
Is the procedure optional, or mandatory, and why?
Frequency of reference to the SOP. How often have you actually referred to the SOP document for guidance?
What information is needed to conduct the task?
What knowledge does a person have to possess in order to be able to conduct this procedure?

Can the procedure be conducted easily or well by individuals who have not practiced it?
What is easy about the procedure?
What is difficult about the procedure?
When you conduct the sequence, do you really do it in a way that departs from the specifications in the SOP? Are there short-cuts? Do you use a "cheater sheet?"
Comments

5. Modeling Practitioner Reasoning

5.1. Protocol Analysis

5.1.1. General Introduction

In this document we have been using the word "protocol" to refer to descriptions of procedures and guidance to the researcher for the conduct of procedures. For the suite of methods that is referred to as "Protocol Analysis," we have a different use and intended meaning of "protocol." In this context, a protocol is a record of a process in which a domain practitioner has performed some sort of task. The record might include data concerning actions and action sequences, cognitions, and sometimes, also often includes data concerning communication and collaboration. This protocol must somehow be analyzed or "mined" for data that can be used in building or verifying a cognitive model (e.g., of reasoning, heuristics, strategies, cognitive styles, knowledge, etc.) or testing hypotheses about cognition. Exploration of the protocol for such information is protocol analysis.

"Protocol Analysis" if often referred to as a research method, when in fact it is a data analysis method. The reason is that protocol analysis is usually employed in conjunction with a single knowledge elicitation task, the "Think-Aloud Problem Solving" (TAPS) task (see Hoffman, Militello, and Eccles, 2005). In the TAPS task, the participant is presented with some sort of puzzle or problem case and is asked to work on the problem while thinking out loud. In their verbalization the participant is asked to attempt to explicate the problem. An audio recording is made of the deliberations. That recording is transcribed and the transcription is then analyzed for content. It is this last activity that is protocol analysis. Thus, many references to Protocol Analysis as a task or as a CTA method are in fact references to TAPS +PA, that is, TAPS plus Protocol Analysis as the data analysis method.

In this section of the document we some guidance concerning protocol analysis.

5.1.2. Materials

To conduct a study in which expert's performance at their familiar tasks is examined, one must select particular problems or cases to present to the expert. Materials can come from a number of sources.

- Since experts often reason in terms of their memories of past experiences or cases (their "war stories") (Kolodner, 1991; Slade, 1991; Wood and Ford, 1993), experts can be asked to retrospect about cases that they themselves encountered in the past.
- Hypothetical test cases (e.g., Kidd and Cooper, 1985, Prerau, 1989) can be generated from archives data or can be generated by other experts. A set of test cases can be intended to sample the domain, or it can be intended to focus on prototypical cases, sometimes to sample along a range of difficulty. For example, Senjen (1988) used archived data to generate test cases of plans for sampling the insects in orchards. The test cases were presented to expert entomologists, who then conducted their familiar task--the generation of advice about the design of sampling plans.

- Occasionally, experts come across a particularly difficult or challenging case. Reliance on tough cases for TAPS knowledge elicitation can be more revealing than observing experts solving common or routine problems (Klein and Hoffman, 1993). Mullin (1989) emphasized the need to select so-called well structured test case problems to reveal ordinary "top-down" reasoning, and using so-called ill structured or novel test case problems in order to reveal flexible or "bottom-up" reasoning. Hoffman (1987) tape recorded the deliberations of two expert aerial photo interpreters who had encountered a difficult case of radar image interpretation. The case evoked deliberate, pensive problem solving and quite a bit of "detective work." In this way, the transcripts were informative of the experts' refined or specialized reasoning.

5.1.3. Coding Schemes

In the traditional analysis, each and every statement in the protocol is coded according to some sort of a priori scheme that reflects the goal of the research (i.e., the creation of models of reasoning). Hence, the coding categories include, for example, expressions of goals, observations, and hypotheses.

A number of alternative coding schemes for protocol analysis have been discussed in the literature. (See for instance, Cross, Christians, and Dorst, 1996; Newell, 1997; Pressley and Afflerbach, 1995; Simon, 1979). Without exception, the coding scheme the researcher uses is a function of the task domain and the purposes of the analysis. If the study involves the reasoning involved in the control of an industrial process, categories would include, for instance, statements about processes (e.g., catalysis), statements about quantitative relations among process variables, and so on (see Weilinga and Breuker, 1985). If the study is about puzzle-solving (e.g., cryptarithmetic), categories might include, for example, statements about goals, about the states of operators, about elementary functions, and so on. In the study concerns expert decision making, categories might include: noticing informational cues or patterns, hypothesis formation, hypothesis testing, seeking information in service of hypothesis testing, sensemaking, reference to knowledge, procedural rules, inference, metacognition, and so on.

Statements need not just be categorized with reference to a list of coding categories. the scheme might be more complex and might involve sub-categories, for instance, statements about operators might be broken down into statements about assigning values to variables, generating values of a variable, testing an equation for variable y value x, and so on to a fine level of analysis (see Newell, 1997).

Also, working backwards from a detailed assignment of each and every statement in a protocol, one can cluster sequences of statements into functional categories (e.g., a sequence of utterances that all involved a forward search or a means-end analysis, etc. (see Hayes, 1989).

5.1.4. Effort

In addition to these method-shaping questions is an important methodological consideration for protocol analysis: No matter what task is used to generate the protocol, transcription and analysis is very time and labor-intensive (see Burton, et al., 1987, 1988; Hoffman, 1987). It takes on the order of seven to ten hours for even an experienced transcriber to transcribe each hour of audio

taped TAPS, largely because even the best typist will have to pause and rewind very frequently, and cope with many transcription snags (e.g., how do I transcribe hesitations, "um's, "ah's," etc.?). Coding takes a considerable amount of time, and the validity check (multiple coders, comparison of the codings, resolution of disagreements, etc.) takes even more time.

Both Burton, et al. and Hoffman compared TAPS+PA with other methods of eliciting practitioner knowledge (interviews, sorting tasks, etc.), and both found that TAPS+PA is indeed time consuming and effortful, and yields less information about domain concepts than did contrived techniques. The upshot is that in terms of its total yield of information, TAPS with Protocol Analysis is one of the *least* efficient method of eliciting domain knowledge from practitioners. On the other hand, if the analysis focuses just on identifying leverage points or culling information about practitioner reasoning, and not on the coding of each and every statement in the protocol, then a review of a protocol can be useful.

5.1.5. Coding Schemes Examples

To illustrate protocol analysis coding schemes, we present four examples. The first three all come from the weather forecasting case study (Hoffman, Coffey, and Ford, 2000).

Example 1: Abstraction-Decomposition

The first coding scheme we use to illustrate protocol analysis shows how protocol analysis is shaped by project goals. The Abstraction-Decomposition Analysis scheme evolved out of research on nuclear safety conducted by engineer Jens Rasmussen at the RIS∅ National Laboratory in Denmark (Rasmussen, Pjtersen, and Schmidt, 1990). That research was initially focused on engineering solutions to the avoidance of accidents, but the researchers discovered that human error remained a major problem. Hence, their investigation spilled over into the ergonomic domain and also the analysis of the broader organizational context. So what started as an engineering project expanded into a fuller Work Analysis. The researchers developed a scheme that can be used for coding interviews and TAPS sessions. The coding representation shows in a single stroke both the category of each proposition and how each proposition fits into the participant's strategic reasoning process and goal-orientation. This scheme is illustrated in Figure 6.1. The rows refer to the levels of abstraction for analyzing the aspect of the work domain that is under investigation. The columns refer to the decomposition of the important functional components.

Levels of Abstraction ↓

Levels of Decomposition →

	Whole System	**Subsystem**	**Component**
Goals			
Measures of the goals			
General functions and activities			
Specific functions and activities			
Workspace configuration			

Figure 5.1. The coding scheme of Abstraction-Decomposition Analysis.

Table 5.1. shows how this scheme would apply to the domain of weather forecasting. In particular, it refers to a workstation system for the Forecast Duty Officer, called the Satellite, Alphanumerics, NEXRAD and Difax Display System (SAND).

Table 5.1. An instantiation of the Abstraction-Decomposition Analysis.

	Forecasting Facility	**Forecast Duty Officer's Desk**	**FDO - SAND Workstation**
Goals	Understand the weather, Carry out standing orders.	Forecast the weather	Supports access and analysis of weather data products from various sources.
Measures of the goals	Duty section forecast verification statistics.	24-hour manning by a senior expert	Forecast verification, System downtime, Ease of use.
General functions and activities	Prepare forecast products and services to clients.	Understanding and analysis of weather data.	Supports access to satellite images, computer models, etc.
Specific functions and activities	Carry out all Standard Operating Procedures.	Prepare Terminal Area forecasts, request NEXRAD products, etc.	Supports comparison of imagery to lightning network data to locate severe storms.
Physical form	The Operations floor layout.	Workspace layout.	CRT location and size, keyboard configuration, desk space, etc.

Once a template for use in a work domain has been diagrammed in such a manner, each of the statements in a specific interview or problem solving protocol can be assigned to the appropriate cell, resulting in a process tracing that codes the protocol in terms of the work domain. A hypothetical example appears in Figure 5.2. The numbered path depicts the sequence of utterances in an interview in which a forecaster was asked to describe the Standard Operating Procedure involved in the use of the SAND system, supported by probe questions about what makes the system useful and what makes it difficult.

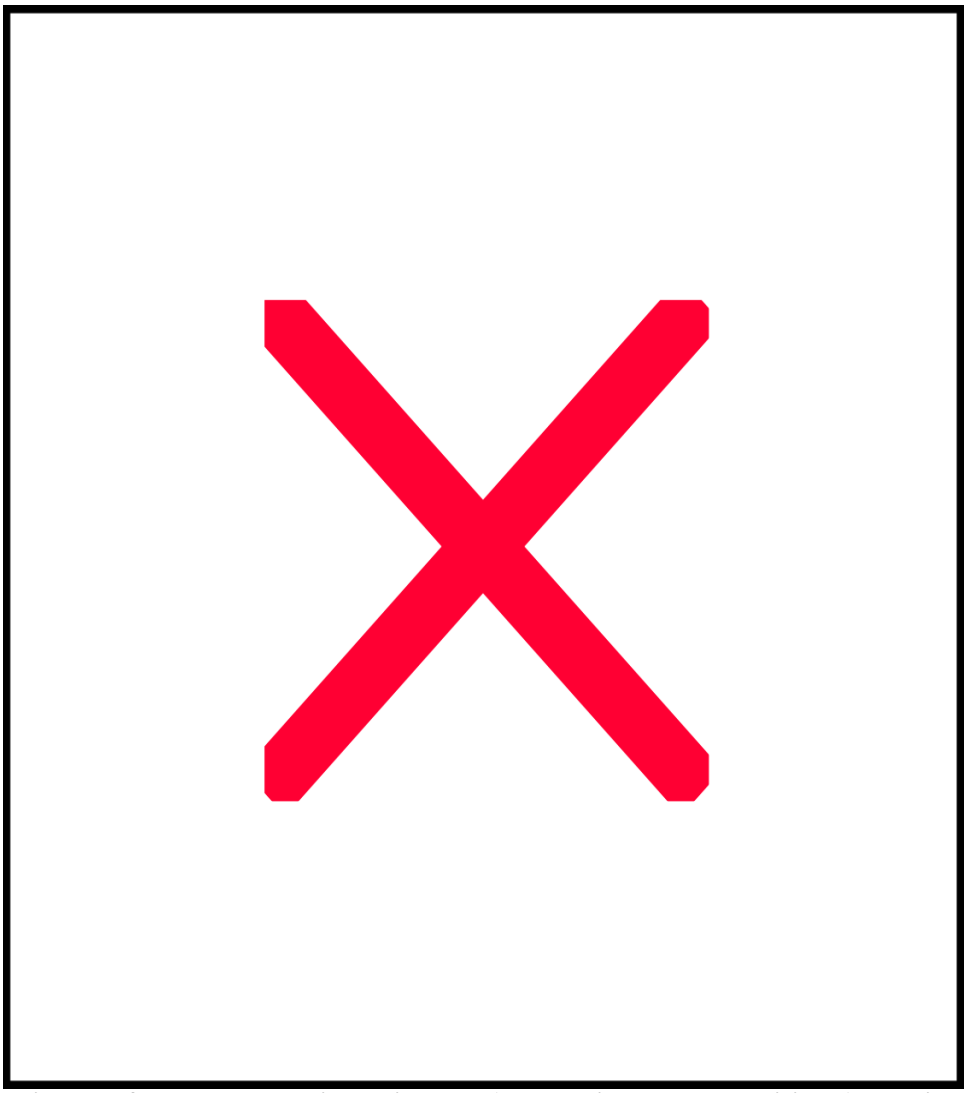


Figure 5.2. Protocol tracing using the Abstraction-Decomposition Analysis. (Reprinted from Hoffman and Lintern, 2005, with permission.)

Example 2: Coding of Propositions for a Model of Knowledge

In the analysis of an interview transcript, statements were highlighted that could be incorporated in Concept-Maps of domain knowledge. From the transcript:

> *'Cause usually in summer here we're far enough south that ah...**high pressure will dominate**, and you'll keep the **fronts and that cold polar continental air north of us**. Even if the front works its way down, **usually by the time it gets here it's a dry front and doesn't have a lot of impact**... I also think of ah... I think of **tornados** too. **Severe thunderstorms** and tornadoes are all part of the summer regime. And again, tornadoes and severe thunderstorms here are **not quite as severe as they are inland** like we talked about last time, because they need to **get far enough in to get the mixing, the shearing and the lift.** And that takes a while to develop and unfold. You really don't see that kind of play **until this maritime air starts to cross the land-sea interface.***

Starting at the beginning of this excerpt, one sees the following propositions:

(In the Gulf Coast region) **high pressure will dominate** (in the summer)
(High pressure) **keeps fronts north of us** (the Gulf Coast region)
(High pressure) **keeps cold polar continental air north of us** (the Gulf Coast region).

Example 3: Coding for Leverage Points

In an analysis of SOP documents, the Participant went through each of the Standard Operating Procedures and was probed about each one, as illustrated in Table 5.2. The purpose of the coding was to identify leverage points, indicated by the boldings.

Table 5.2. Notes from an interview concerning a SOP. SIGMET stands for SIGnificant METeorologcical event.

Task/Job: Observer updates SIGMETs	
Action sequence: Conducted every hour (valid for 2 hours) SIGMET is defined and identified Change is saved as a jpeg file Change is sent to the METOC home page and to thereby to the Wall of Thunder	
What supports the action sequence?	What is the needed information?
METOC PC Powerpoint with an old kludge. There are a series of 4 maps - SE Texas - East Coast, - FL, AL, MS, - VA through GA, - TX, OK, Arkansas.	How to find the SIGMETS (they come from Kansas City). **Have to plot them - sometimes look up stations in the station identifier book. Have to know more geography than many observers know.**
What is good or useful about the support and the depiction of needed information?	
The SIGMETS themselves are really good - give customers good information at a glance. The PowerPoint maps are designed for METOC - they have all geographical information and 3 letter identifiers (e.g., PNS for Pensacola) that is needed. Other forecasting offices have blank maps of the US.	
What about the support or information depiction makes the action sequence difficult?	
Too labor intensive - the whole system is archaic. There is a commercial Web site that has a Java program with map and red box for SIGMET. you can highlight the box and get text describing the SIGMET. This always seems to be updated. **Limited capability to customize the shapes of the SIGMET areas.** **The map cannot move as you are in the act of drawing a SIGMET--you have to change functions and scroll the map with the mouse.** **The alphanumerics are hard to see even if you zoom.** **The map shown on the CRT is not large enough, details are hard to read.** It is a sectored map--cuts off at Texas. **You sometimes have to hunt for station identifiers--end up searching via "Yahoo." Some stations have several IDs.** Map cannot scroll outside a limited area. **Nothing in the work environment reminds the Observer to conduct the task.** **NOTE: The final display of SIGMETs does not support a zoom function. They aren't easy to see on Data Wall.** The work is often done on the hardcopy map lying on the table--where you can see all the regions and station identifiers in a glance, and read all the details. After figuring it out on the hardcopy map the Observer inputs it into the computer.	

The entries in this table are the researcher's notes from the interview—including some paraphrases and synopses of what the participant said—and are not an utterance-for utterance protocol. This underscores the idea that protocol analysis, as a data analysis method, can be divorced from the TAPS task and can be used for a variety of purposes.

Example 4: Coding an Unstructured Interview to Identify Rules for an Expert System.

A project for the U. S. Air Force Military Airlift Command was aimed at developing an expert system to assist in the creation of airlift plans (Hoffman, 1986). The software used in planning was complex, and only one individual had achieved expertise in its use. The researchers conducted an unstructured interview with that expert. An example excerpt is presented in the left-hand column in Table 5.3. The focus of the interview at this point was on the kinds and nature of the information provided in a file about airfields. The excerpt is cropped in the center column, with boldings in that column indicating the coded pieces. The right-hand column shows the encoded propositions in an intermediate representation, a step closer to implementability as concepts for a knowledge base and procedural rules for an inference engine.

Table 5.3. Coding of a transcript from an unstructured interview. The purpose was to identify concepts for a knowledge base and potential inference rules for an inference engine.

Transcript	Encoded Transcript	Concepts and Rules
I: What is the difference between MOG and airport capability?		
E: Ah… MOG maximum on the ground is parking spots… on the ramp. Airport capability is how many passengers and tons of cargo per day it can handle at the facilities.	E: Ah… **MOG maximum on the ground is parking spots**… on the ramp. **Airport capability is how many passengers and tons of cargo per day it can handle at the facilities.**	**Concept:** **Maximum number of aircraft allowed on the ground = MOG** **Concept:** **Airport capability is the number of passengers and number of tons of cargo per day the airport can throughput.**
I: Throughput… ah… throughput as a function of… E: It all sorta goes together as throughput. If you've only got… if you can only have ah… if you've got only one parking ramp with the ability to handle 10,000 tons a day, then your… your throughput is gonna be limited by your parking ramp. Of the problem could be vice versa.	… your **throughput is gonna be limited by your parking ramp.** Of the problem could be vice versa.	**Rule:** **Parking ramp capacity can limit throughput.)**
I: Yeah?…		
E: So it's a [unintelligible phrase in the audio recording]		
I: So what if you had only one loader, so that you could only unload one wide-body airplane at a time? You wouldn't want to schedule five planes in the ground simultaneously. How would you restrict that?		
E: We know we're not gonna get all the error out of it. We're gonna try and minimize the error. And then we'll say that… ah… we'll say an arrival-departure interval of one hour, so that means that the probability of having two wide-bodies on the ground tryin' to get support from the loader is cut…	E: We know we're not gonna get all the error out of it. We're gonna try and minimize the error. And then we'll say that… ah… we'll say an **arrival-departure interval of one hour, so that means that the probability of having two wide-bodies on the ground tryin' to get support from the loader is cut…**	**Rule:** **If you need to restrict the number of aircraft on the ground then manipulate the arrival-departure interval.**

Note in this example that there are many propositions in the Expert's statements (e.g., "The airport designation file only includes the field's common name, latitude, longitude, MOG, and capability) that were not coded because they were not useful in composing implementable

statements. In addition there were statements made by the Interviewer (e.g., "The airport capability can be restricted by the number of loaders.") and that also could have been coded, and even implemented as rules, but were not because they were only tacitly affirmed by the Expert. (This sort of slippage is one of the disadvantages to unstructured interviews).

Like the first three examples, this fourth one also makes the point that the coding scheme depends heavily on the purposes of the analysis.

5.1.6. Coding Verification

In some cases it is valuable to have more than one coder conduct the protocol coding task. In some cases it is necessary for demonstrating soundness of the research method and the conclusions drawn from the research. For research in which data from the TAPS task are used to make strong claims about reasoning processes, especially reasoning models that assert cause-effect relations among mental operations, the assessment of inter-coder reliability of protocol codings is regarded as a critical aspect of the research (see Ericsson and Simon, 1993).

There are a number of ways of using multiple coders and conducting a validity check. In the simplest procedure, two researchers independently code the statements in the protocol and a percentage of agreement is calculated. Researchers typically look to find a high rate of 85 percent or greater agreement among multiple coders (see Hoffman, Crandall, and Shadbolt, 1998).

Analysis of multiple codings can show that the coding scheme is well defined, consistent, and coherent. On the other hand, it is likely that there will be disagreements, even among coders who are practiced and are familiar with the task domain. Disagreements can be useful pointers to ways in which the coding scheme, and the functional categories on which it is based, might be in need of refinement.

5.1.7. Procedural Variations and Short-cuts.

In another procedure:

1. Two or more researchers, working independently, go over the typed transcript and highlight every statement that can be taken as an instance of one of the coding categories.
2. Each researcher codes each highlighted statement in terms of the coding categories.
3. Each researcher codes the highlighted statements from ***OTHER*** researcher's highlightings.
4. Both the highlightings and the codings from the researchers are compared.

A short-cut on this general approach to coding verification is to have multiple coders and a reliability check, and once an agreement rate of 85% or more is achieved, all remaining transcripts can be coded by individual researchers.

Because the process of transcription alone, let alone coding in addition, is very time-consuming and effortful, researchers may want to consider another short-cut. In this procedural variation, there are two researchers, one serving as Interviewer to facilitate the interview (or other

empirical procedure) and the other serving as a Recorder. about what everyone says, especially things that he participant says that are salient in terms of the project goals. After the notes are transcribed, each researcher independently highlights every statement that can be taken as a reference to a coding category.

For some research, the assessment (elaborate or otherwise) of inter-coder reliability may not be necessary. For instance, the identification of leverage points in the analysis of the Standard Operating Procedures in the weather forecasting case study (Hoffman, Coffey, and Ford, 2000) did not require an assessment of inter-coder reliability because the leverage points were explicitly elicited from the expert using interview probe questions. Furthermore, the coding scheme was simple—a statement either was or was not an expression of a possible leverage point. Likewise, in the protocol analysis of the Concept Mapping interview, the identification of statements that could be used in Concept Maps did not mandate verification that *all possibl*e statements that could be used in Concept Maps were in fact identified and used. Such an analysis would have actually detracted from the main the goal of the analysis, which was to identify propositions that could be used in Concept-Maps on particular topics, and that were not already in the Concept-Maps that had been created.

5.1.8. A Final Word

In planning to conduct a protocol analysis procedure in the context of systems design, there are a number of questions the researcher might consider at the outset. These are presented in Table 5.4.

Table 5.4. Some questions for consideration when planning for a protocol analysis procedure.

What are the purposes? • If the purpose is to develop reasoning models, then the categorization scheme needs to include such (slippery) categories as "observation," "goal," and "hypothesis." • If the purpose is to identify leverage points, the categorization scheme can be simple (i.e., highlighting statements that refer any sort of obstacle to problem solving). •
What is the level of analysis? • Do I cut a coarse grain—"notice," "choose," "act"? • Or do I cut a fine grain based on a functional analysis of the particular domain? (e.g., "look at data type x," "notice pattern q," "choose operation y," "perform procedure z," and so on). •
Do I need to code each and every statement? • What constitutes a statement? Do I separate them by the hesitations on the recording? • How do I cope with synonyms, anaphor, etc.? • Statements are often obviously dependent on previous statements. Context dependence is *always* a feature of protocol analysis because context dependence is always a feature of discourse. •
How intensive must the validity check be? • Indeed, must there be a validity check at all given the purposes of the research? • Do I need to have independent coders code the protocol, and then compare the codings for inter-coder reliability? How many coders—two? Three? What rate of agreement is acceptable? How do I cope with the inevitable disagreements?

Answers to these questions will determine what, precisely, is done during the protocol analysis.

5.2. The Cognitive Modeling Procedure

5.2.1. Introduction and Background

The Documentation Analysis, Work Space Analysis, Participant Interviews, and other bootstrapping methods can allow the researcher to forge a preliminary model of domain practitioner reasoning or reasoning style, and even variations that capture the reasoning of less versus more proficient Practitioners, or variations representing the ways that reasoning strategies are shaped by problem types.

As is traditional in cognitive science, such models take the form of information processing flow diagrams or decision trees, with boxes indicating mental capacities such as memory, boxes indicating mental operations such as inference, and arrows indicating transformations or information flows among the hypothetical components.

The question always arises as to how to validate such models, that is, insure that the models are veridical or valid. The main approach taken in traditional experimental psychology is to conduct programmes of experimental research in which models are developed, refined, and then tested under controlled circumstances that allow convergence on the "truth" concerning the causal relations among mental processes.

For the purposes of most applied research--which is invariably conducted under significant constraints of time, resources, and particular goals--it is out of the question to step out of the Cognitive Work Analysis and Cognitive Field Research venue to design and conduct series of formal experiments. Might there not be some other way to take a "fast track into the black box?"

One way would be to simply present to the domain Practitioners the model that the researcher forged in the bootstrapping process, and elicit their commentary, i.e., "Does this model capture your reasoning?" There are two main difficulties with this approach:

- Knowledge elicitation using abstract or conceptual-level probes can be ineffective. Research using the CDM procedure, for example, has shown repeatedly that the more effective probes are those that are couched in terms of specific experiences (e.g., "How did you reason in this particular case?") rather than in terms of typical patterns (i.e., "Describe how your typical pattern of reasoning.).

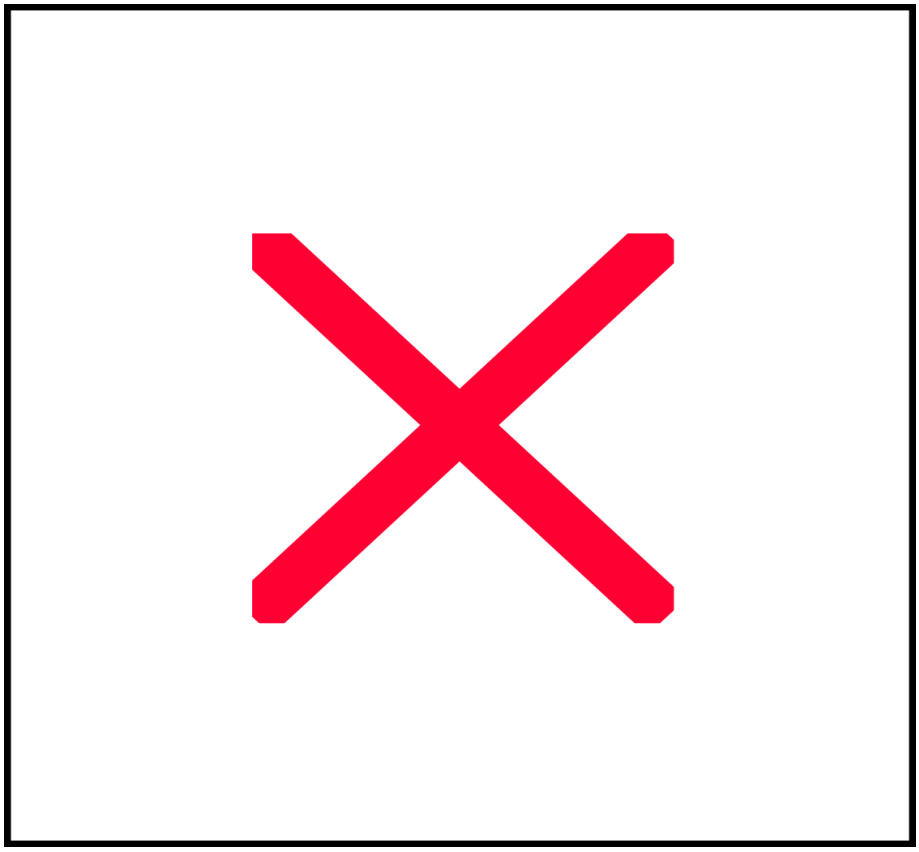
- Any responses to the Base Model may be "just-so stories," biased by the Base Model itself. That is, the responses may reflect the Practitioner's attempt to please the interviewer.

A second way to obtain validation would be via observations, to see if the patterns of behavior suggested by the model conform to actual task behavior. The main difficulty here is that cognitive activities (i.e., data inspection in service of mental model formation versus data inspection in service of hypothesis testing) may not be discernable on the basis of overt behaviors.

The issue of validating cognitive models is a significant one, and in theory requires formal controlled, and costly experimentation.

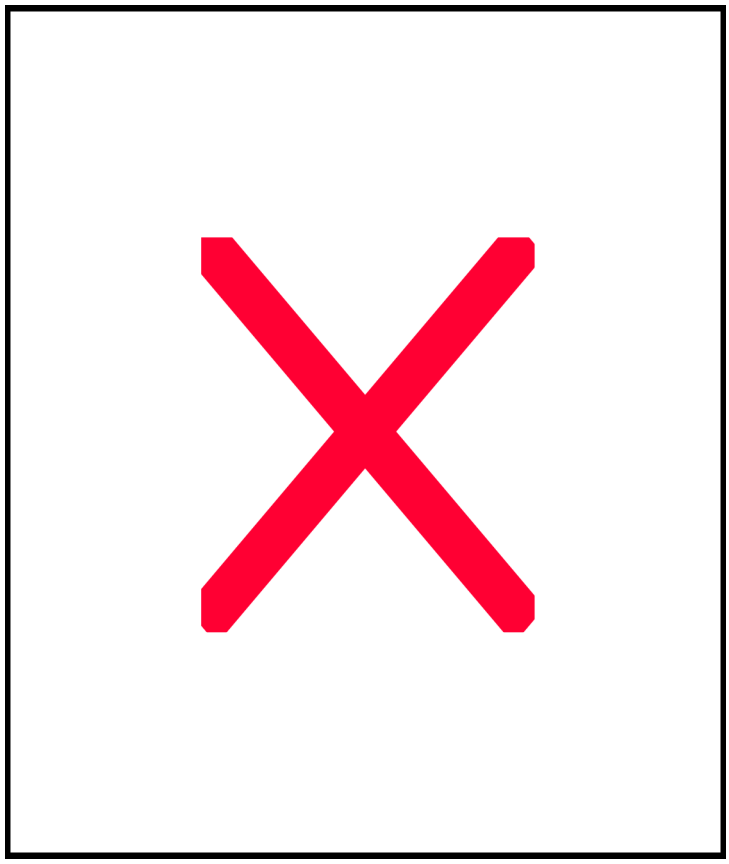
5.2.2 Protocol Notes

The "Cognitive Modeling Procedure" (CMP) is one way of obtaining validation in direct, and more efficient manner. In this interview, one begins with the "Base Model of Expertise." (See Hoffman and Shadbolt, 1996). The Base Model is shown in the following Figure.



5.2.2.1. Preparation

Armed with this Base Model, the researcher takes what was learned from the bootstrapping activities (i.e., documentation analysis, unstructured interviews, etc.) and tweaks the Base Model to make it accord with the domain. In the case of the weather forecasting domain, the Base Model would be modified as in the following Figure.



Next, the researcher takes the elements in the Base Model and recombines them so as to create two "bogus models." The bogus models will include boxes containing concept terms and action paths (arrows) labeled by actions. The following Table lists the sorts of terms and labels that reasoning models can include. This listing of terms was compiled from the models presented in a great number of recent studies of various domains of real-world decision-making (Hoffman and Shadbolt, 1995).

Some Useful Concept-terms	**Some Useful Labels for Action Paths**
UNDERSTANDING OF THE CURRENT SITUATION	COMPARE
MENTAL SIMULATION	AGREE?
KNOWLEDGE OF CONCEPTS	DISAGREE?
KNOWLEDGE OF PRINCIPLES	TEST HYPOTHESIS
KNOWLEDGE OF "RULES OF THUMB"	REVISE/CHANGE
MEMORY/PAST EXPERIENCES	SEARCH
PATTERN	RECOGNIZE
CUE	VERIFY/CONFIRM
PREDICTION	REFUTE
HYPOTHESIS	RECOGNIZE PROBLEM
GUIDANCE	MODIFY
PLAN	IMPLEMENT
INPUT/OBSERVATIONS	REMEMBER
JUDGMENT	EVALUATE
GOAL/SUB-GOAL	DIAGNOSE
EXPECTATION	INFER
PROTOTYPE/SCHEMA	DETECT
	IDENTIFY
	ASSESS CONFIDENCE QUALITY ALTERNATIVES/COURSES OF ACTION
	ANALOGIZE
	RECOMMEND
	DECIDE/CHOOSE/SELECT
	PRIORITIZE
	SCHEDULE

The concept-terms used for boxes and for action paths are not limited, except in that as they need to make reference to cognitive activities and task-relevant behaviors, and in that they have to make functional sense in terms of the data that have already been collected up to the point when the Cognitive Modeling Procedure is conducted.

To the greatest extent possible, the concept terms and action path labels should not rely on technical or psychological jargon. (See the Table above.) For instance, if one were to use the concept "mental model," the concept would have to be explained to the Participant. Instead, one might use lay terminology such as "Understanding of the Current Situation."

The two bogus models need to be reasonable. For example, one would probably not have a "disagree" link lead directly to a box labeled "output judgment."

One or both of the bogus models should include some sort of feedback loop, to let the participant know that such loops are acceptable. This is especially important since it is a good bet that one or more such loops will be involved in proficient reasoning in any domain. If one were to present two artificial models that did not contain one or more feedback loops, this might give the Participant the impression that the Interviewer was expecting purely linear models.

On the other hand, one should avoid having the feedback loop in the bogus models be an actual Duncker refinement cycle. One of the main things one seeks in the Cognitive Modeling Procedure is to have the Participant re-create this cycle and thereby validate the Base Model.

One or both of the bogus models should include some notion that hearkens of hypothesis testing, and some notion that hearkens of situational awareness, again since it is a good bet that such notions will be involved in proficient reasoning in any domain

On the other hand, the two bogus models should not conform precisely to those models that actually do, at this stage of inquiry, seem to describe Practitioner reasoning.

Try to make the two bogus models fairly simple (see the Base Model that is presented here). The artificial models should have only about 5 boxes and 3 linking arrows. the bogus models should serve as an invitation for the Practitioner to adapt the terms to their domain and flesh the models out.

Example "bogus models" are presented below in the Boiler Plate sub-section, and these can be used in composing the CMP.

5.2.2.2. Round 1

In Round 1 of the CMP, the Practitioner/Participant is presented with the two bogus models and is asked two probe questions:

- Which of these seems to best describe how you approach your domain task(s)?
- Are there any ways in which this (selected) model seems to be incorrect or in need of refinement?

(Don't be surprised if the participant shrugs, or even chuckles.) The Participant is told that they can feel free to concoct a diagram of their own, and that there is no one correct way of diagramming their reasoning of strategies. Indeed, they should be told that they may need more than one diagram to capture their different strategies or the ways in which their reasoning might depend on the nature of the case or situation at hand.

It is recommended that the participant be explicitly informed that they are allowed some quiet time alone, and perhaps even a span of a few days, to ponder the models and consider their answers to the probe questions. Otherwise, the pressure of having to create a more immediate answer might have a negative impact on the reflective thought that this task requires.

The caution here being that if the researcher leaves material with the Practitioner, there is a high likelihood that on the researcher's return, nothing will have been done. Domain Practitioners, even those who may be enthused about the project, will often, if not usually "drop the ball" if left to their own devices. It is highly recommended that if the material is to be left with the Practitioner, at the time when the material is first presented an appointment should be made for a specific date and time no more than a week hence when the interview and discussion will be conducted.

The notion in this interview is to have the domain Practitioner correct the bogus model that was selected as being the one that came closest to matching their own reasoning. model. The Practitioner is expected to and add richness to the over-simplification of the bogus model. What should come out as a consequence should be in greater conformity with the Base Model, i.e., validation of the Base Model or a Base Model of a particular reasoning style.

Although this interview method engages the Practitioner at an abstract or conceptual level of analyzing their own thinking, it does not seem to suffer from the abstractness. When a model does not fit, the Practitioner can see this, and can use the bogus models as support in clarifying their description of their own reasoning style.

After the completion of Round 1, the Researcher prepares a polished version of the Participant's model. Inevitably, the polished models will differ from the models crafted by the participants in Round 1, in ways that can be subtle and possibly profound. This is because of the degrees of freedom one has in concocting these sorts of models, and also because of the deliberate attempt on the part of the researcher to "clean up" the Participant's descriptions and create an elegant, if not totally logical, depiction. There are many degrees of freedom involved in the construction of reasoning models, and there is no single correct way to create a model. Placement of concepts and identification of paths can vary considerably without fundamentally altering the important flows in reasoning sequences. Thus, it is prudent to present the polished model to the Participant and invite comment. Fine-tuning of the model may be called for.

5.2.2.3. Round 2

After the Round 1 data have been collected and the resulting models from a number of domain Practitioners have been placed into a polished graphic, it is possible to run Round 2. In Round 2, each Participant Expert is presented with copies of all of the models that had been derived in Round 1. Added to this set are some foils derived from the analysis of Participants who are not expert, and some (new) bogus models. The task presented to each Participant Expert is to:

- Judge which model goes with which of the Practitioners (including their own)
  or
  Rate the degree of expertise manifested in each of the models
- Judge which model(s) describe the reasoning of individuals who are not experts
- Judge which model(s) are bogus, and
- Rate their degree of confidence in each of their judgments.

It is best to limit the number of models to less than 10 (say, one bogus model, two apprentice models, and 4 expert models). For greater numbers of models, the task can become overwhelming.

The Interviewer keeps track of task time and takes notes.

Round 2 can be used to explore a number of hypotheses about the work domain. For instance, if the Practitioner-participants are all experts, one might expect that their models would tend to be similar, and hence in Round 2 they would show considerable confusion and even fail to recognize their own model. If this hypothesis is of interest, it is desirable for the researcher to wait some weeks, or even months, between the conclusion of Round 1 and the beginning of Round 2.

The researcher may be interested in testing hypotheses about the workplace, such as, "Do these Practitioners discuss their reasoning amongst themselves?" "Do the all have the opportunity to learn of each others' reasoning styles and strategies by seeing each other in action?" "How closely do they communicate and cooperate in their organizational context?" Results from Round 2 can address such questions.

Hypotheses for Round 2

HYPOTHESIS	ALTERNATIVE HYPOTHESIS
Experts will recognize their own models, with confidence, if only because they will have already seen their polished model during the Round 1 discussion in which the Analyst's cleaned-up version is presented for confirmation.	If there is a time delay between CMP Round 1 and Round 2, experts will confuse the models and will fail to recognize their own. As skill approaches expertise, one would expect the reasoning models to converge, setting the stage for error or confusion. Experts may falsely recognize models of other Experts, insofar as those other models conform to some degree to their own reasoning, or to the normative model of proficient reasoning.
Experts will recognize the models of individuals who are not expert, with confidence, especially if the non-expert models tend to be simplifications.	Expert models might be relatively simple as well as complex, depending on the level of detail the Expert went into during CMP Round 1. So, just because a model is simple, that does not mean that it is necessarily a bogus or an apprentice model. Indeed, a participant may regard even a bogus model as being an expert's model.
Experts will correctly identify the bogus models, assuming that the bogus models include some sort of clear violation of domain practice or logic.	If the bogus models are oversimplifications rather than clear violations of domain practice, then they may be confused with the expert models.
Experts are likely to use a "divide-and-conquer" strategy in which they first attempt to identify the models of the senior experts, and their own model, and then partial out the remaining alternatives into the apprentice and bogus categories.	Other strategies are possible. A less frequent one is to peruse all of the models carefully before making any decisions.
Advanced apprentices and journeymen should be able to say which model goes with each expert, assuming that they have had the opportunity to work with each of the experts. Presumably, the apprentices and journeymen observe the experts' reasoning patterns and recognize stylistic differences and individual preferences.	None of the participants will show confidence in assigning the models, since they may rarely, if ever, discuss or even have the opportunity to witness, one another's reasoning style or strategy.

5.2.2.4. Round 3

In Round 3, the Analyst observes the Participant when he or she is conducting an actual task in the operational context. The researcher notes as much as possible about everything the Practitioner does and says, seeking behavioral evidence that might confirm or disconfirm aspects of the Participant's model. For instance, if a participant asserts in Round 1 that they always begin their familiar task by looking at a particular data type, then there should be corresponding behavior that should be observable-- if the researcher is located in the workplace just at the moment the participant begins to conduct their familiar task.

In addition to taking detailed notes on the Practitioner's behavior, the researcher can use the refined participant's model as a "checklist."

It is possible during Round 3 to ask occasional probe questions (assuming that such intrusion in the operational setting is appropriate, acceptable, and agreed to in advance). For instance, suppose again that a Practitioner asserted in Round 1 that they begin their familiar task by inspecting a particular data type (that should be observable in their behavior) and also said in Round 1 that they then attempt to come to an understanding of the current situation. This is mental modeling, which in all likelihood will not be readily observable. The researcher might simply ask, "What are you thinking?" just at the point where the Practitioner has conducted the initial examination of the data for an appropriate span of time, or has turned away from the data.

Round 3 has to be conducted opportunistically. For instance, in the weather forecasting project it was decided that the best time to make the Round 3 observations would be at the very beginning of the watch period of a forecaster who had been off of the watchbill for a span of days. In this case, the forecaster would almost certainly have to begin their watch by examining data to develop their initial mental model of the overall weather situation--something most of the forecasters asserted (in Round 1) that they did. If they were observed on any other day, they would have already had a mental model of the current weather situation when they walked in the door. Furthermore, the observations had to be conducted at a time when the local weather was not in a persistence regime--when the weather is dominated by an air mass or a stationary front, making the weather situation basically the same from day to day.

Opportunism notwithstanding, the researcher needs to keep focused on the main goal of Round 3--observe whatever can be observed, and even ask probe questions, to see what elements of the Practitioner's model are mirrored in actual behavior.

The Round 3 results can also be examined with regard to the Base Model of domain expertise that was developed in the Preparation phase--which elements were affirmed, which need refinement.

5.2.2.4. Round 3

Following Round 3, the data collected in all three Rounds can be pulled together into the final products:

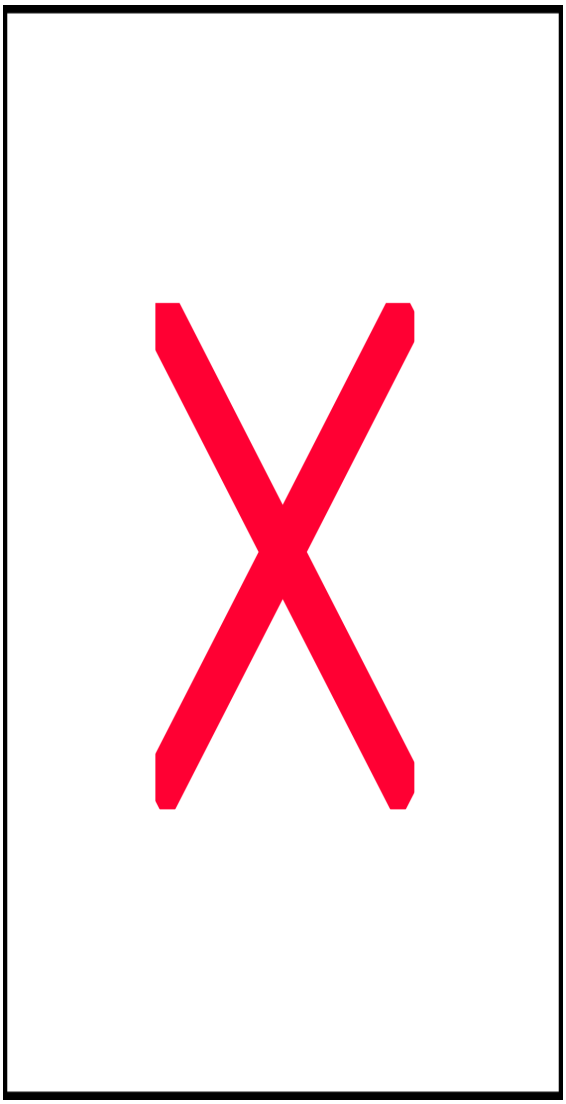
- Refined Base Model of domain expertise
- Refined models of the reasoning of the individual participants

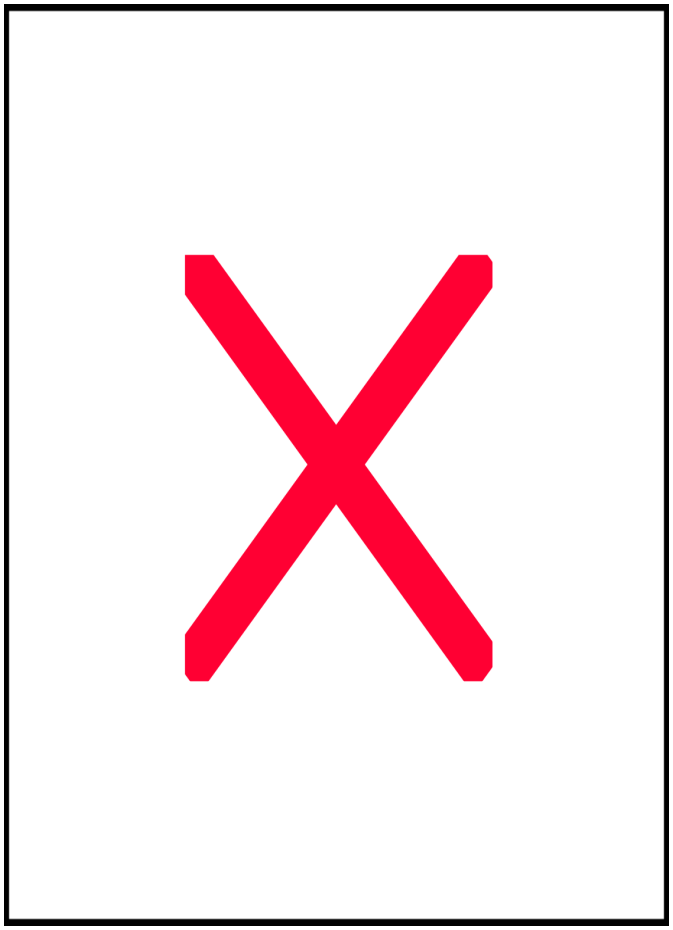
- Conclusions regarding any hypotheses that were testable in Round 2.

5.2.3. Templates

5.2.3.1. Round 1

Example Bogus Models.





Analyst	
Participant	
Date	
Start Time	
Finish Time	

Instructions

Based on the material we have collected so far in this project, we have attempted to create something like a logical flow diagram that attempts to capture the sorts of strategies that domain Practitioners use.

On the next page you will see diagrams of two possible reasoning strategies. Another way of saying what this is that these are two alternative ways of describing what you do when you conduct your usual or typical task.

We would like you to look at these and think about them. You may have some immediate reactions, but you may want to take this booklet with you and think about the strategies for a day or so.

Which of these seems to best describe how you approach your domain task(s)?

Are there any ways in which this (selected) model seems to be incorrect or in need of refinement?

Please feel free to concoct diagrams of your own. You might use a strategy other than those shown here. You might use more than one strategy depending on the given task situation.

Please keep in mind that there is no one correct way of diagramming your reasoning of strategies. Indeed, you may need more than one diagram to capture your different strategies, or those of other Practitioners.

After you have thought about the two strategies that are depicted in the diagram, please complete pages 4 and 5.

Insert your "Bogus Models" page here

Which of these diagrams seems to best describe how you approach your domain task(s)? In what ways does it seem to capture your strategies?
Are there any ways in which this (selected) diagram seems to be incorrect or in need of refinement? What changes would you make? Feel free to make changes right on the diagram page itself.
Feel free to concoct diagrams of your own, ones that better represent your strategy or strategies.
DISCUSSION

5.2.3.2. Round 2

Analyst	
Participant	
Date	
Start Time	
Finish Time	

Instructions

Based on the material we have collected so far we have been able to recreate graphical representations of the reasoning styles of each of our Participants. Those models appear on these pages that I will lay out on the table.

For each of these models, we would like to guess "who is the owner." That is, for each model, can you determine the person whose reasoning the model seems to capture?

Please write you judgment on each page. If you uncertain of the ownership of a model, you can give more than one guess.

Also, try to give your reasons for assigning models to individuals. What in your experience working with each individual leads you to assign each particular model to each particular person?

One of the models is yours. Can you tell which? If so, what clues you in?

Not all of the models presented here go with the Experts. One or more may represent the reasoning of an individual who is less-than-expert. See if you can determine that.

Not all of the models presented here are "real" models that would represent the reasoning of ANY domain Practitioner, whether expert or not. See if you can determine which of the models is "bogus."

In attempting to determine the ownership of a model, you may be confident, or uncertain, for any of a variety of reasons. For instance, you may have the suspicion that a given model represents the reasoning of some particular person, but you may not have had much opportunity to actually witness that person's reasoning style. Hence, you may be uncertain. As another example, you may feel that more than one model might represent the reasoning of a particular person.
Please share such reactions with the researcher.

If you have any questions at any time, please ask.

(reproduce the following set of cells, as needed)

MODEL ID	Whose Model is it?	Rate your confidence

	Check One: Name _____ A "bogus" model _____ Someone who is not an expert _____	Circle one: **1** - Highly Confident **2** - Confident **3** - Somewhat confident **4** - Somewhat uncertain **5** - Uncertain **6** - Very Uncertain
COMMENTS		

5.2.3.3. Round 3

Analyst	
Participant	
Date	
Start Time	
Finish Time	

E =EVENT, P = PROBE R = REPLY, A = ANALYSIS	**TIME**	**NOTES**
		(reproduce this table, as needed)

6. Critical Decision Method

6.1. General Description of the CDM Procedure

The CDM Procedure involves multi-pass retrospection. The expert is guided in the recall and elaboration of a previously-experienced case. The CDM leverages the fact that domain experts often retain detailed memories of previously-encountered cases, especially ones that were unusual, challenging, or in one way or another involved "critical decisions." The CDM works by deliberately avoiding generic questions of the kind, "Tell me everything you know about x," or, "Can you describe your typical procedure?" Such generic questions have not met with much success in the experience of many knowledge engineers. More fundamentally:

1. In complex sociotechnological domains of practice, there often is no general or typical procedure.
2. Scenarios may be "typical," or prototypical, but:
   There are likely to be many action sequence alternatives, even for routine tasks and routine situations.
   Procedures will vary as a function of style.
   Procedures will vary as a function of the equipment status.
   Procedures will vary as a function of practitioner skill level.

Thus, any uniformities observed in procedures may reflect problem differences only, and are likely to simply reflect the procedures taught in the schoolhouse and are prescribed by the operational forms and reporting procedures. Therefore, the CDM probes focus on the recall of specific, concrete experiences from the forecaster's experience

6.1.0. *Preparation*

- Training of elicitors to properly conduct the CDM procedure has involved a number of exercises, including experience in the role of interviewee, experience in the conduct of (mock) CDM procedures, and experience at preparing an interview guide. As in the exercise of any knowledge elicitation method, effective use of the CDM presumes that the elicitor is a skillful facilitator at the level of the social dynamic.

- The knowledge elicitor must become familiar with the domain—through analysis of research and technical documents, preliminary interviews (i.e., conversations with domain experts, on-site observations, etc.).

- The goals for knowledge elicitation are specified and defined operationally.

- Experts, in sufficient number, must be available and cooperative.

6.1.1. Step One: Incident Selection

The opening query poses a particular type of event or situation and asks for an example where the experts' decision making altered the outcome, where things would have turned out differently had s/he not been there to intervene, or where the experts' skills were particularly challenged. The incident must come from the person's own lived experience

as a decision maker. The goal in this Step is to help the participant identify cases that are non-routine, especially challenging, or difficult —cases where one might expect differences between the decisions and actions of an expert and those of someone with less experience, and cases where elements of expertise are likely to emerge. One should bypass incidents that are memorable for tangential reasons (e.g., someone died during the incident) and incidents that are memorable but did not involve the expert in a key decision-making role.

6.1.2. Step Two: Incident Recall

The participant is asked to recount the episode in its entirety. The participant is asked to "walk through" the incident and to describe it from beginning to end. The elicitor asks few, if any, questions, and allows the participant to structure the account.

6.1.3. Step Three: Incident Retelling

Once the expert has completed the initial recounting of the incident, the elicitor tells the story back, matching as closely as possible the expert's own phrasing and terminology, as well as incident content and sequence. The participant is asked to attend to the details and sequence. The participant will usually offer additional details and clarifications, and corrections. This sweep allows the elicitor and the participant to arrive at a common understanding of the incident.

6.1.4. Step Four: Timeline Verification and Decision Point Identification

The expert goes back over an incident account a second time. The expert is asked for the approximate time of key events. A timeline is composed along a domain-meaningful temporal scale, based on the elicitor's judgment about the important events, the important decisions, and the important actions taken. The timeline is shared with and verified by the expert as it is being constructed. The elicitor's goal is to capture the salient events within the incident, ordered by time and expressed in terms of the points where important input information was received or acquired, points where decisions were made, and points where actions were taken. Input information could include objectively verifiable events (e.g., *"The second fire truck arrived on the scene"*) and subjective observations (e.g., *"I noticed that the color of the smoke had changed; it was a lot darker than it had been when I arrived on the scene"*). In some cases, the markers for decision points are obvious in the expert's statements (e.g., *"I had to decide whether it was safe to send my firefighters inside the building"*). In other cases, the expert's statements suggest feasible alternative courses of action that were considered and discarded; or suggest that the expert was making judgments that would affect the course of decision making. The goal is to specify and verify decision points, points where there existed different possible ways to understanding a situation or different possible actions.

6.1.5. Step Five: Deepening

The elicitor leads the participant back over the incident account a third time, employing probe questions that focus attention on particular aspects of each decision-making event within the incident. The step typically begins with questions about the informational cues that were involved in the initial assessment of the incident. The elicitor focuses the

participant's attention on the array of cues and information available within the situation, eliciting the meanings those cues hold and the expectations, goals, and actions they engender. The elicitor then works through each segment of the story, asking for additional detail and encouraging the expert to expand on the incident account. This step is often the most intensive, sometimes taking an hour or more. Solicited information depends on the purpose of the study, but might include presence or absence of salient cues and the nature of those cues, assessment of the situation and the basis of that assessment, expectations about how the situation might evolve, the goals that were considered, and the options that may have been evaluated. The array of topics and specific probe questions that might be used in this sweep of the CDM are listed with the Boiler Plate forms (below). No single CDM procedure would be likely to include all of the topics or probes presented here; elicitors select some subset of these in accord with goals of the study.

6.1.6. Step Six: "What if?" Queries

The fourth sweep through the incident involves shifting the from the participant's actual experience of the event to a more analytical strategy. The elicitor poses various hypothetical changes to the incident account and asks the participant to speculate on what might have happened differently. In studies of expert decision making, for example, the query might be: *"At this point in the incident, what if it had been a novice present, rather than someone with your level of proficiency? Would they have noticed Y? Would they have known to do X?"* The elicitor can ask the expert to identify potential errors at each decision point and how and why errors might occur, in order to better understand the vulnerabilities and critical junctures within the incident. The purpose of the "what if" probes is to specify pertinent dimensions of variation for key features contained in the incident account. Klein, Calderwood, and MacGregor (1989) noted that the reasons for taking a particular action are frequently illuminated through understanding choices that were not made, or that were rejected.

6.2. Protocol Notes For the Researcher

- The goal is to develop a detailed account of a set of scenarios via the method of retrospection concerning "critical" decisions.
- The detailed account will include timelines and decision-point analyses.
- A time-line is a layout of the sequence of events involved in the analysis of a particular event or scenario.
- The fundamental dimension is the time course of the event, and laid onto this are milestones indicated by component events, situation assessments and decision points-actions.
- Situation assessments are based on informational cues and component events.
- Decision points involve triggering cues and situation assessments, determination that information is needed, hypothetical reasoning, action options, action goals, and action rationale.
- Do not plan to obtain detailed CDM accounts representing all possible tasks, scenarios, mission types, etc.
- Since practitioner skill levels and styles are both highly variable, and are themselves dependent on context, the analysis must focus at the functional level. If the analysis goes into too detailed a level, the recommendations may not address the most significant system design problems.
- Seek information about cognitive skills and requirements across the many different aspects of the work as performed by different types of practitioners, rather than trying to specify everything involved in each of a set of various particular types of procedures.
- Do not focus on specifying the details of each and every task/scenario, but instead focus on the variability in the tasks that practitioners face and the cognitive demands that are shared across tasks/scenarios.

- Conduct detailed CDM analyses of selected incidents only to insure that the general models of styles and cognitive activities are representative.

- Do not assume that HCC decision-aids will be developed that support activities for each of a set of particular scenarios or mission types.

- According to the literature on the CDM, sessions should last from between 1 - 2 hours. However, in some domains, the case stories that practitioners have to tell are rich indeed. The CDM sessions an last hours, and the process of transcription and analysis likewise can take hours. It is possible to split the CDM into two or three sessions. In a three-way split:

- Encourage the Participant to create sketches or diagrams to describe the situation or any of its aspects. The Participant might be able to provide graphic resources from archived data on the case. These can all be integrated into the final case study.

- The first session includes Step One (Incident Selection) and Two (Incident Recall).

- After these two steps, the elicitor transcribes the notes and prepares for the second session.

- The second session includes Steps Three (Retelling) and Four (Timelining).

- After these two steps, the elicitor transcribes the notes and prepares for the third session.

- The third session includes Steps Four (Timeline verification and Decision-Point analysis), Five (Deepening) and Six ("What-if" Queries).

- In a two-way split:
    - The first session includes Steps One and Two (followed by a brief break during which the elicitor creates the Timeline) and then Step Three,
    - The second session includes Steps Four, Five and Six.

6.3. Boiler Plate Forms

6.3.0. Cover Form

Analyst	
Participant	

Time in Interview			
Date	Date	Date	Date
Start	Start	Start	Start
Finish	Finish	Finish	Finish

Time in Transcription and analysis			
Date	Date	Date	Date
Start	Start	Start	Start
Finish	Finish	Finish	Finish

Total Time

6.3.1. Step One: Incident Selection

Start Time	
Break	
Finish Time	

Instructions
In this interview we want to talk about the interesting and challenging experiences you have had. We will provide you with guidance in the form of specific questions to support you in laying out some of your past experiences.

Can you think of a case you were involved in that was particularly difficult or challenging? -- A time when your expertise was really put to the test?
Can you remember a situation where you did not feel you could trust or believe certain data, such as a computer model or some other product--a situation where the guidance gave a different answer than the one you came up with?
Can you remember a case when you went out on a limb?--Where the data might have suggested something other than what you thought or another practitioner might have done things differently, but you were confident that you were right?
(expand the cells and copy this table as needed)

6.3.2. Step Two: Incident Recall

Start Time	
Break	
Finish Time	

Instructions
This (selected event) sounds interesting. Could you please to recount it again in a bit more detail, from beginning to end.

(copy this table as needed)

6.3.3. Step Three: Incident Retelling

Date		
	Transcription	Retelling
Start Time		
Finish Time		

Instructions
Now what I'd like to do is read back your account. As I do so, please check to make sure I've got it all right, and feel free to jump in and make corrections or add in details that come to mind.

Elicitor's embellishments/modifications	**Participant's added details**
(expand the cells and copy this table as needed)	

6.3.4. Step Four: Timeline Verification and Decision Point Identification

Analyst drafts the timeline

Date	
Start Time	
Finish Time	

Verification and Decision Point Identification

Date	
Start Time	
Finish Time	

Instructions
Now I'd like to go through the event again and this time we will try to create a timeline of the important occurrences--when things happened, what you saw, the decisions or judgments you made, and the actions you took.

OSA = Observation / Situation Assessment	**D = Decision**	**A = Action**
Event	**OSA, D, A**	**Time**
(expand the cells and copy this table as needed)		

6.3.5. Step Five: Deepening

Date	
Start Time	
Finish Time	

Instructions
Now I want to go through the incident again, but this time we want to look at it in a little detail and I'll guide you with some questions.

PROBE TOPIC	PROBES
Cues & Knowledge	What were you seeing?
Analogues	Were you reminded of any previous experience?
Standard Scenarios	Does this case fit a standard or typical scenario? Does it fit a scenario you were trained to deal with?
Goals	What were your specific goals and objectives at the time?
Options	What other courses of action were considered or were available?
Basis of Choice	How was this option selected/other options rejected? What rule was being followed?
Mental Modeling	Did you imagine the possible consequences of this action? Did you imagine the events that would unfold?
Experience	What specific training or experience was necessary or helpful in making this decision? What training, knowledge, or information might have helped?
Decision-Making	How much time pressure was involved in making this decision? How long did it take to actually make this decision?
Aiding	If the decision was not the best, what training, knowledge, or information could have helped?
Situation Assessment	If you were asked to describe the situation to a relief officer at this point, how would you summarize the situation?
Errors	What mistakes are likely at this point? Did you acknowledge if your situation assessment or option selection were incorrect? How might a novice have behaved differently?
Hypotheticals	If a key feature of the situation had been different, what difference would it have made in your decision?

Event		**OSA, D, A**	**Time**
PROBE (Copy this set of cells for each event in the timeline)	RESPONSE		

6.3.6. Step Six: "What-if" Queries

Date	
Start Time	
Finish Time	

Instructions
Now I want to go through the incident one more time, but this time I want to analyze it a little and to do that I'll ask you some hypothetical questions.

PROBES
What might have happened differently at this point?
What were the alternative decisions that could have been made here?
What choices were not made or what alternatives were rejected?
At this point in the incident, what if it had been a novice present, rather than someone with your level of proficiency? Would they have noticed Y? Would they have known to do X?"
What sorts of error might have been made at this point? Why might errors have occurred here?

Event		**OSA, D, A**	**Time**
PROBE (copy this set of cells for each event in the timeline)	RESPONSE		

6.3.7. Final Integration (Re-telling, Timeline, Deepening, and What-if Steps)

In the final integration, the Researcher collapses all of the comments from the Timeline Verification, Deepening and What-if steps into the text for each entry in the verified timeline.

The result is a detailed case study expressed as a timeline, with each entry specified as being either OSA, D, or A. Any sketches the Participant made during any of the steps, or any graphic resources on the case that have been procured can be inserted at the appropriate places.

Event and Comments	**OSA, D, A**	**Time**
(expand the cells and copy this table as needed)		

6.3.8. Decision Requirements Table

The Researcher examines each cell in the Final Integration and culls from it all statements that refer to each of the categories in the DRT.

Cues and Variables	
Needed information	
Hypotheticals	
Options	
Goals	
Rationale	
Situation Assessment	
Time/effort	

7. Goal Directed Task Analysis

Debra G. Jones and Mica R. Endsley
SA Technologies

7.1. Introduction

In systems that involve a significant amount of cognitive work, conducted under dynamic and changing conditions, developing and maintaining situation awareness (SA) is a significant and critical task for decision makers, forming the basis for decision making and action. Developing systems to effectively support decision makers as they carry out their mission requires that these systems specifically support a wide range of SA requirements that are needed to support the many key decisions and goals of the decision maker. Good system design needs to support not just part of their decisions or tasks, but the full range of work that must be accomplished. To meet this need, a cognitive task analysis methodology for determining the SA requirements of decision makers has been developed (Endsley, 1993, 2000), which can be applied to a wide variety of domains. Key features of this method are that it (a) focuses not on just determining a list of data that is needed, but also identifies how that data is used and combined to form SA – specifically the comprehension and projection requirements associated with a job, and (b) provides a systematic approach for determining these requirements across the many varying goals of a job.

This method is based on obtaining a detailed knowledge of the specific goals the decision maker must accomplish and the decisions that people make in meeting each of these goals. As goals are a central organizing feature for information seeking and interpretation in complex systems, they form a foundation for this type of cognitive task analysis. A detailed goal structure provides a systematic foundation for insuring that the full range of decisions and information requirements involved in a person's job are considered in the analysis process. The goals, decisions and information requirements identified through this analysis provide the foundation for both the creation of new systems and the evaluation of current systems. The ability of a system to support decision makers in successfully accomplishing their mission depends on how well the system supports their goals and information requirements.

This type of information typically can not be accessed easily via a traditional task analysis, as traditional task analyses focus more on the physical tasks and functions that must be accomplished and the processes by which these functions and tasks are completed. This focus is restrictive in that it often does not identify the real information the decision maker needs to accomplish goals, nor does it identify how the decision maker integrates information to gain an understanding of the situation. In addition, they tend to assume that work is sequential and linear, when in fact much of situation awareness involves a constant juggling back and forth between multiple goals and processing of information in a very non-linear fashion. Thus, although traditional task analyses can provide information that is vital to the design process, alone they are insufficient for fully identifying the informational requirements of decision makers interacting with complex, dynamic environments.

A Goal Directed Task Analysis (GDTA), addresses these issues by identifying the goals a decision maker must achieve in order to accomplish a mission, the decisions that must be made in order to accomplish these goals, and the specific information that is needed to support these decisions (Endsley, Bolte, & Jones, 2003) (See Figure 1).

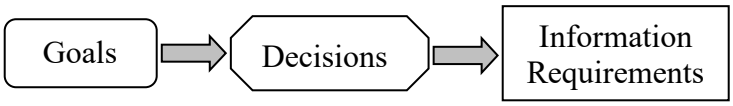


Figure 1: Elements of the GDTA

The GDTA documents what information decision makers need to perform their job and how they integrate or combine information to address a particular decision. The GDTA is not developed along task or procedural lines, is not sequenced according to a timeline, and does not reflect goal priority, as priorities change dynamically depending on the situation. Rather, the GDTA focuses on what information decisions makers would ideally like to know to meet each goal, even if that information is not available given current technology. The ideal information is the focus of the analysis, as focusing only on what current technology can provide would induce an artificial ceiling effect that would obscure much of the information the decision maker would ideally like to know. Further, the means an operator uses to acquire information are not the focus of this analysis as methods for acquiring information can vary from person to person, from system to system, from time to time, and with advances in technology.

The GDTA delineates the specific SA requirements a decision maker needs and determines the nature and format of how this information must be integrated in order to achieve each goal relevant to successful task completion. The purpose behind the creation of a GDTA is to identify the information the decision maker really needs and uses. Once this information has been identified, systems can be evaluated to determine how well the current design meets these needs, and future designs can be created that take these needs into account from the beginning.

7.2. Elements of the GDTA

The GDTA has three main components: Goals, Decisions, and SA Requirements.

7.2.1. Goals

The goals a decision maker has with respect to a particular task, mission, or operation are one element of the GDTA. Goals are higher-order objectives essential to successful job performance. The highest level goal reflects the *overall goal* of the decision maker. For example, the overall goal for the Army Operations Officer is to "Plan and Execute the Mission". Associated with this overall goal is a set of *main goals* that must be achieved in order to accomplish the overall goal. Although the number of main goals varies depending on the complexity and breadth of the domain under consideration, typically three to six main goals describe the overall goal. Finally, each main goal may have a varying number of *subgoals* associated with it. For more complex domains, subgoals may also have any number of associated sub-goals. The goals increase in specificity as they move down the hierarchy (Figure 2).

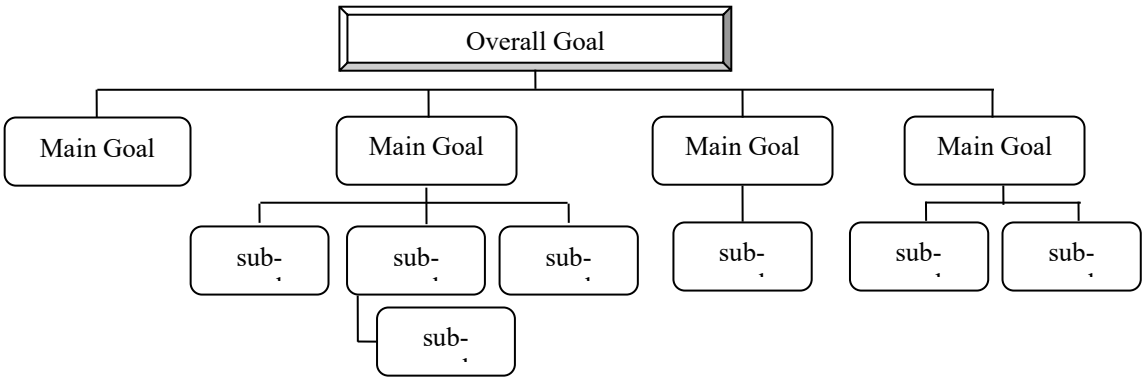


Figure 2: Example levels of goals within the GDTA

7.2.2. Decisions

The next element of the GDTA involves the decisions the decision maker must make in order to achieve a particular goal. Decisions are associated with a specific goal, although a similar decision may play into more than one goal. These decisions are essentially the questions the decision maker must answer in order to achieve a specified goal. These questions require the synthesis of information in order to understand the situation and how it will impact its associated goal. For example, for the Army brigade intelligence officer goal "Determine impact of environment on friendly forces" one of the questions that needs to be answered in order to achieve this goal is "what is the impact of weather on friendly forces?".

7.2.3. Situation Awareness Requirements

The final element of the GDTA involves the information needed to answer the questions that form the decisions. These information needs are the decision maker's situation awareness requirements. Situation awareness (SA) can be defined as "the perception of the elements within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1988, 1995). From this definition, 3 levels of situation awareness can be identified: Level 1 which involves the most basic data that is perceived, Level 2 which involves an integration of Level 1 data elements, and Level 3 which involves projecting how the integrated information will change over time. The SA requirements analysis identifies and documents relevant information at all three of these levels.

7.3. Documenting the GDTA

Two main structures are utilized to document the GDTA: the goal hierarchy and the relational hierarchy.

7.3.1. Goal Hierarchy

The primary goal hierarchy begins with the overall operator goal, from which the major goals are identified and the subgoals essential for successfully achieving the major goal defined. The structure and depth of the primary goal hierarchy will depend on the characteristics of the domain under consideration. The information needed develop this goal hierarchy as well as for subsequent components of the GDTA are generated through a variety of mechanisms (e.g., interviews, document reviews, etc) which will be discussed later.

The goal hierarchy is made up of the overall goal, the main goals, and the subgoals associated with a particular task or domain. These goals and subgoals are numbered in outline style for organizational purposes; the numbers are not intended to denote priority or sequential processes. The main goals are numbered 1.0, 2.0, 3.0, etc., and the associated sub-goals are numbered 1.1, 1.2, etc. (Figure 3). Although the GDTA can be completed without this numbering system, using

this convention has been found to facilitate discussion and revision of the GDTA. An example of a GDTA for the Army S2 is shown in Figure 4.

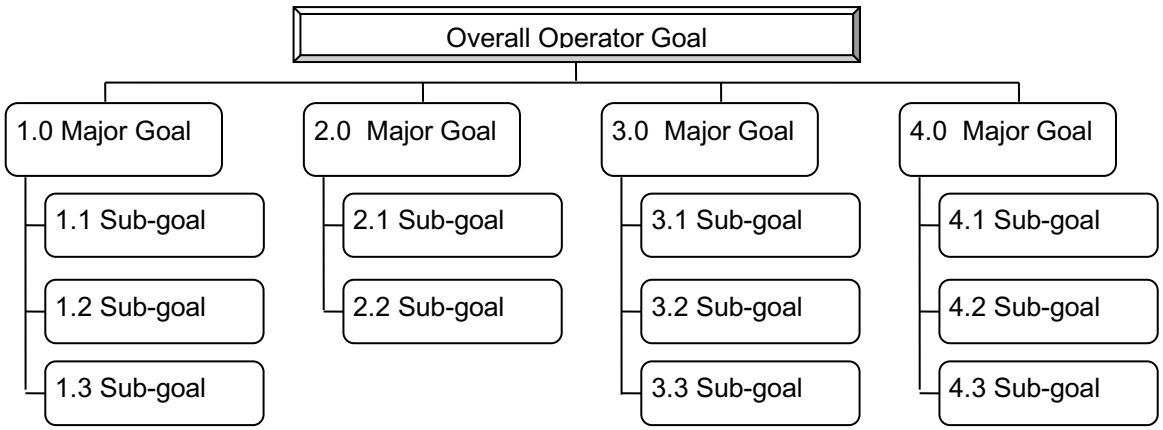


Figure 3: GDTA goal hierarchy

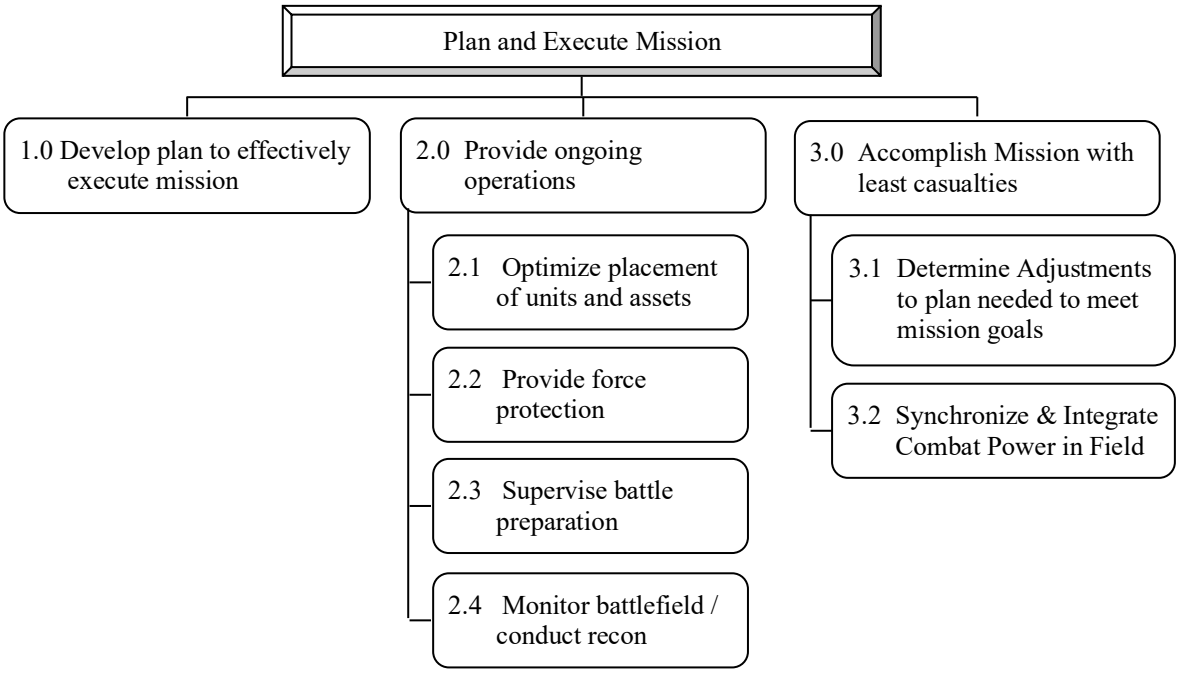


Figure 4: Army Operations Officer Goal Hierarchy

7.3.2. Relational Hierarchy

The relational hierarchy shows the relationship between the goals, the subgoals, the decisions related to each subgoal, and the SA requirements relevant for each decision (See Figure 5). Each of the elements in the relational hierarchy has a corresponding shape that, although not necessary for the completion of a GDTA, does facilitate information transfer and comprehension. The exact shape of the GDTA with respect to the number of goals and related sub-goals depends on the complexity of the domain and the goal being analyzed. The GDTA is a flexible tool. The structure can be varied however needed to most accurately represent the relationship between the

various goals, subgoals, decisions, and SA requirements essential to the domain under consideration.

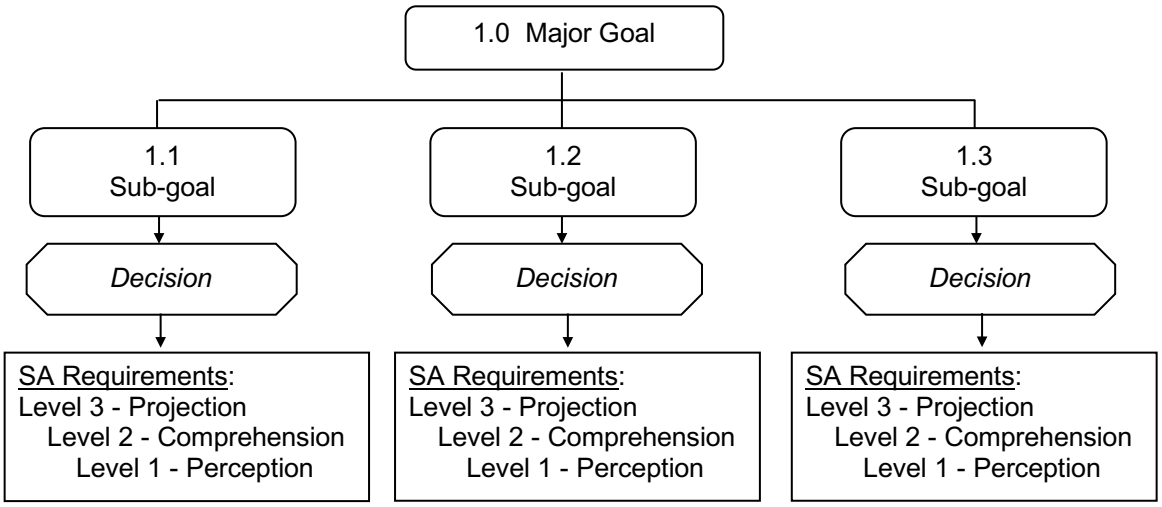


Figure 5: Relational Hierarchy

7.3.3. Identifying Goals

Items that are appropriate for designation as goals are those items that require operator cognitive effort and that are essential to successful task completion. They are higher-level items as opposed to basic information requirements. They are a combination of sometimes competing subgoals and objectives which must be accomplished in order to reach the person's overall goal. The goals themselves are not decisions that need to be made, although reaching them will generally require that a set of decisions and a corresponding set of SA requirements be known.

At times, a subgoal will be essential to more than one goal. In these cases, the sub-goal should be described in depth at one place in the GDTA, and then in subsequent utilizations, it can be "called out", referencing the earlier goal rather than repeating the goal in its entirety. Cross referencing these callouts helps maintain the clarity of the GDTA, improve cohesiveness, support consistency, and minimize redundancies. Utilizing callouts instead of repeating information is beneficial to the interview process as well as these subgoals can be discussed at one point rather than inadvertently taking up valuable interview time reviewing the same subgoals.

Performing a task is not a goal, in part because tasks are technology dependent. A particular goal may be accomplished by means of different tasks depending on the systems involved. For example, navigation may be done very differently in an automated cockpit as compared to a non-automated cockpit. Yet, the SA needs associated with the goal of navigation are essentially the same (e.g., location or deviation from desired course). Although tasks are not the same as goals, the implications of the task should be considered. A goal may be addressed by a task the subject matter expert mentions during the interview. By focusing on goals rather than tasks, future uses of the GDTA will not be constrained. In some future system, this goal may be achieved very differently than by the present task – for example, the information may be transferred through an electronic network rather than with a verbal report.

Goal names should be descriptive enough to explain the nature of the subsequent branch (i.e., the related subgoals, decisions, and SA requirements) and broad enough to encompass all elements related to the goal being described. Further, the goals should be just that – goals, not tasks nor information needs. For example, physical tasks are not goals: rather they are things the operator

must physically accomplish, such as filling out a report or calling a coworker. Rather, goals require the expenditure of higher-order cognitive resources such as predicting an enemy course of action (COA) or determining the effects of an enemy's COA on the outcome of the battle.

7.3.4. Defining Decisions

The decisions that are needed to effectively meet each goal in the goal hierarchy are listed beneath the goals to which they correspond (see Figure 5). Decisions reflect the need to synthesize information in order to understand how that information is going to affect the system both now and in the future. Although decisions are presented in the form of questions, not all questions qualify as decisions. Questions that can be answered "yes/no' are not typically considered appropriate decisions in the GDTA. For example, the question "Does the commander have all the information needed on enemy courses of action?" can be answer with yes/no and does not qualify as a decision. On the other hand "how will enemy courses of action affect friendly objectives?" requires more than a yes/no answer and is a decision that is pertinent to the person's goals. Further, if a question's only purpose is to discern a single piece of information, it is not a decision, rather it is an information requirement and belongs in the SA requirements portion of the hierarchy.

When a single goal or subgoal has more than one relevant decision, these decisions can be listed separately or together within the goal hierarchy. These decisions (and their corresponding SA requirements) can be listed separately or bunched in the hierarchy, depending on personal preference and the size of the goal hierarchy. One advantage for listing them separately is that the SA requirements for each decisions can be easily discerned from each other. This aids in the process of insuring that all information needed for a decision is present. If after final analysis the SA requirements for several decisions show complete overlap, they can then be combined in the representation for conciseness.

7.3.5. Delineating SA requirements

Decisions are posed in the form of questions, and the associated SA requirements provide the information needed to answer the questions. To determine the SA requirements, each decision should be analyzed individually to identify all the information the operator needs to make that decision. The information requirements should be listed without reference to technology or the manner in which the information is obtained. When delineating the SA requirements, be sure to fully identify the item for clarity – for example, instead of 'assets', identify 'friendly assets' or 'enemy assets'. Although numerous resources can be utilized to develop an initial list of SA requirements (e.g., interview notes or job manuals), once a preliminary list is created, the list needs to be verified by domain experts.

Often, experts will express many of their information needs at a data level (e.g., altitude, airspeed, pitch). Further probing may be required to find out why they need to know this information. This probing will prompt them to describe the higher-level SA requirements – how this information is used. For example, altitude may be used (along with other information) to assess deviations from assigned altitude (Level 2 SA) and deviations from terrain (Level 2 SA). When the expert provides higher order information requirements (e.g., "I need to know the enemy strengths"), further probing will be required to find all of the lower-level information that goes into that assessment (e.g., ask "What about the enemy do you need to know to determine their strengths?").

The SA requirements are listed in the GDTA according to its level of SA by using an indented stacked format (see Figure 5). This format helps ensure the SA requirements at each level are considered and generally helps with the readability of the document. However, the three levels of SA are general descriptions that aid in thinking about SA. At times, definitively categorizing an SA requirement into a particular level will not be possible. For example in air traffic control the amount of separation between two aircraft is both a Level 2 SA requirement (distance now) and a Level 3 SA requirement (distance in the future along their trajectories). Further, not all decisions will include elements at all three levels – in some cases, a Level 2 requirement may not have a corresponding Level 3 item, particularly if the related decision is addressing current, rather than future, operations.

At times, a series of SA requirements will be essential for numerous goals. In these cases, this frequently used subset of SA requirements can be turned into callout blocks with a descriptive title. For example, weather issues at all levels of SA may be of concern across many subgoals in the hierarchy. To decrease complexity and reduce redundancy, a weather callout block can be created. This block can be called within the SA requirements section of multiple subgoals as 'Weather' and fully defined at the end of the GDTA (See Figure 6). Consistency is important for the SA requirements, and SA requirement callout blocks can be useful to ensure that each time a frequently used item is referenced all of its relevant information requirements are included.

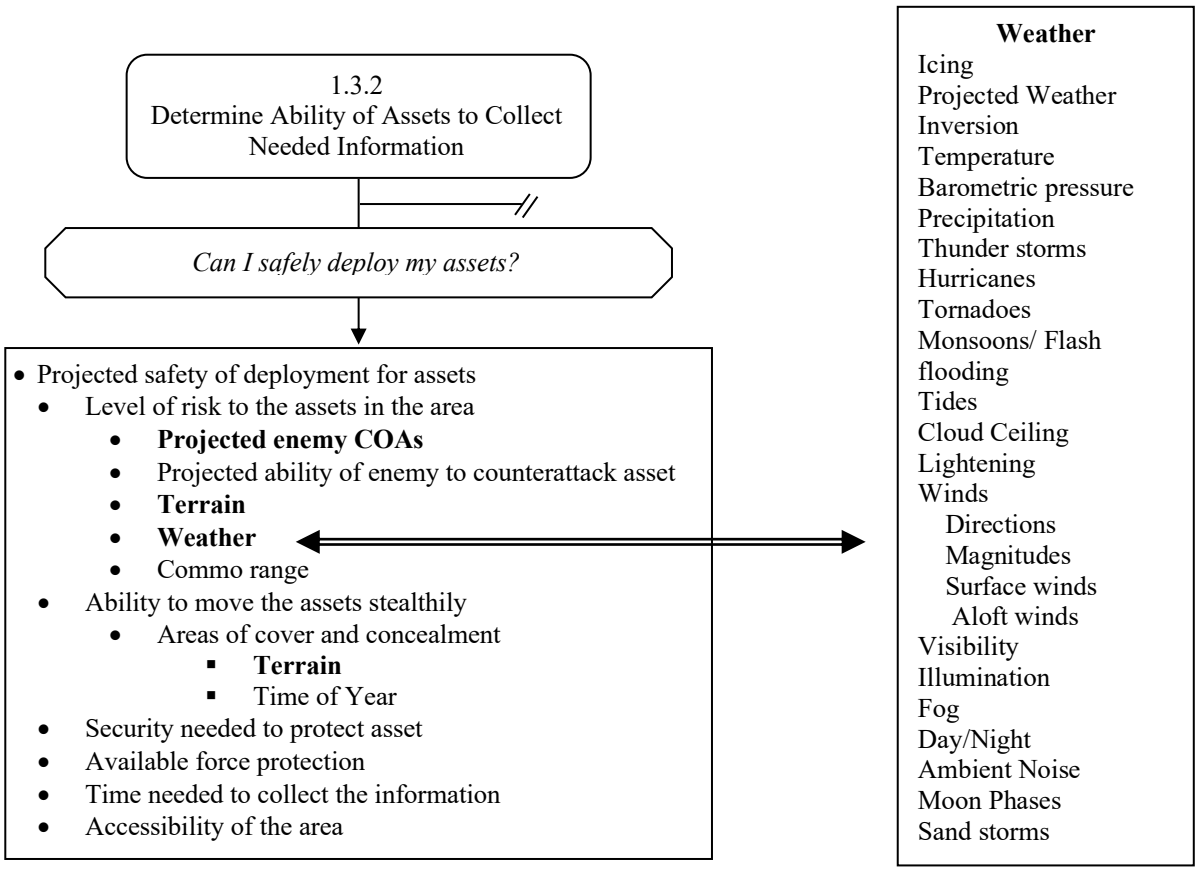


Figure 6: Subset of Army Intelligence Officer GDTA showing Weather Call-out.

7.4. GDTA Methodology

7.4.1. Step 1: Review Domain

Prior to beginning a goal-directed task analysis, a general understanding of the domain of interest should be developed. This understanding can be gained by reviewing relevant documents pertaining to the domain of interest (e.g., job descriptions, job taxonomies, manuals, performance standards). This review accomplishes several things. First, it provides an overview of the domain and the nature of the decision maker's job. Second, it provides a background for understanding the lingo the interviewee is accustomed to speaking. Finally, the interview should go more smoothly (and the interviewer's credibility enhanced) if the interviewer can "talk the talk." The interviewer must be cautious, however, not to develop a preconceived notion of what the decision maker's goals are likely to be and, as a result, to seek only confirming information in the interview.

7.4.2. Step 2: Initial interviews

Interviews with subject matter experts (SMEs) are an indispensable source for information gathering for the GDTA. Whenever possible, each SME should be interviewed individually. When more than one person is interviewed at a time, the data collection may be negatively impacted if one participant is more vocal than the other or if the participants have dissenting opinions regarding relevant goals and decisions. Each interview begins with an introduction of the purpose and intent of the data collection effort and a quick review of the interviewee's experience. Next, the SME is asked about the overall goals relevant for successful task completion. As the SME explains the overall goals, make note of topics to pursue during the interview. This process of noting potential questions as they come to mind should be followed throughout the entire interview. Relying on memory for questions may work for some, but for most the questions will be forgotten (but crop up again after the interview when the practitioner is trying to organize the GDTA).

Obtaining time with domain experts is often difficult, so maximizing the time when it is available is essential. Several suggestions can be offered to this end. First, conduct the interview with two interviewers present to minimize downtime; one interviewer can continue with questions while the other interviewer quickly reviews and organizes notes or questions. The presence of two interviewers will also have a benefit beyond the interview; constructing the goal hierarchy is easier when two people with a common frame of reference (developed in part from the interview session) work on the task. Next, limit the number of interviews performed on any given day. Time is needed to organize information gleaned from each interview and to update the charts to insure that subsequent sessions are as productive as possible. Interviewing too many people without taking time to update notes and charts can negatively impact not only the interview sessions but also the resulting data quality. Finally, interviewer fatigue can be a considerable factor if too many interviews are scheduled without a break or if the interviews last too long. Two hours per session is generally the maximum recommended.

7.4.3. Step 3: Develop the goal hierarchy

Once the initial interviews have been completed, the task becomes one of organizing all the disparate pieces of information collected during the interview into a working preliminary goal structure that will allow for adequate portrayal of the information requirements. Determining the overall goal is usually fairly straightforward. The art of the process comes in when delineating the goals essential for the success of the overall goal. Typically more questions are raised than answered during the first attempts to develop a goal hierarchy. Nonetheless, creating a preliminary goal structure, even if incomplete or sketchy, is essential and will aid in focusing future data collection efforts.

Although each practitioner will develop a unique style for developing the goal hierarchy, one approach is to begin by reorganizing the notes from the interviews into similar categories. This categorization can be done by beginning each new category on a separate page and adding statements from the notes to these categories as appropriate. Sorting information in this manner helps illuminate areas that constitute goals and may make it easier to create a preliminary goal hierarchy. Although defining an adequate goal hierarchy is the foundation of the GDTA, in the early stages this hierarchy will not be perfect, and an inordinate amount of time should not be spent trying to make it so. Further interviews with experts will most likely shed new light that will require that the goal hierarchy be revamped by adding, deleting, or rearranging goals. Nonetheless, developing even a preliminary goal structure at this point in the process allows for a baseline for future iterations, helps in the process of aggregating information, and helps direct information gathering efforts during the next round of interviews.

7.4.4. Step 4: Identify decisions and SA requirements

Once a preliminary goal hierarchy has been developed, the associated decisions and SA requirements can be developed to the extent possible given the amount of data collected. Notes taken during the interview can be preliminarily linked with the associated decision and goals. Further, after completing the process of organizing the interview notes into the relational hierarchy, existing manuals and documentation can be referenced to help fill in the holes. Caution should be exercised however, since these sorts of documents tend to be procedural and task specific in nature. Information in the GDTA is concerned with goals and information requirements, not current methods and procedures for obtaining the information or performing the task. Although operators often don't explicitly follow written procedures found in the related documentation, evaluating the use of this information in the hierarchy can help to insure the GDTA is complete and can spark good discussions of what information the decision maker actually needs to achieve a goal.

7.4.5. Step 5: Additional interviews / SME review of GDTA

Once a draft GDTA is created, it can serve as a tool during future interview sessions. One way to begin an interview session is to show the primary goal hierarchy to the interviewee and talk about whether the goal hierarchy captures all the relevant goals. After this discussion, one section of the GDTA can be selected for further review, and each component of that section (I.e., goals, subgoals, decisions, and SA requirements) can be discussed at length. The draft can be used to probe the interviewee for completeness (e.g., "what else do you need to know to assess this?") and to determine higher-level SA requirements ("how does this piece of data help you answer this question?"). Showing the entire GDTA to the participant is generally not a good idea, as the apparent complexity of the GDTA can be overwhelming to the participant and consequently counter-productive to data collection. If time permits after discussing one section of the GDTA, another subset can be selected for in-depth discussion.

7.4.6. Step 6: Revise the GDTA

After each round of interviews is complete, the GDTA should be revised to reflect new information. New information gleaned from iterative interviews allows for reorganizing and condensing of the goals within the structure to create better logical consistency. For example, when several participants bring up a low-level goal as an important facet of the job, the goal may need to be moved to a higher place in the hierarchy. Furthermore, as more information is uncovered concerning the breadth of information encompassed by the various goals and subgoals, the goals and subgoals can often be refined to create a more descriptive and accurate category. Often goals will seem to fit more than one category. In these cases, the goals can be represented in detail in one place in the GDTA and referenced (i.e., "called out") to other places within the hierarchy as needed.

As the GDTA is reviewed further, commonalities become apparent that can be used to assist in refining the charts. Consequently, after defining a preliminary set of goals, the goal structure should be reviewed to determine if goals that originally seemed distinct, albeit similar, can actually be combined into a single category. Combining similar goals will reduce repetition in the analysis. For example, goals involving planning and re-planning rarely need to be listed as separate major goals; often representing them as branches under the same major goal is sufficient to adequately delineate any distinct decisions or SA requirements. When combining goals that are similar, all of the information requirements associated with the goals should be retained. If the information requirements are very similar but not quite the same, separate subgoals may be in order, or perhaps future interviews will resolve the inconsistency. If several goals seem to go together, this clustering might be an indication that the goals should be kept together under the umbrella of a goal one level higher in the hierarchy.

7.4.7. Step 7: Repeat steps 5 & 6

Additional interviews with subject matter experts should be conducted and the GDTA revised until a comprehensive GDTA has been developed. After discussing a set of GDTA charts with one participant, the charts should be updated to reflect the changes before being shown to another participant. Showing a chart to another participant before it has been edited may not be the best use of time; interviewer notations and crossed out items on the charts can be confusing to the participant and thus counterproductive.

7.4.8. Step 8: Validate the GDTA

To help insure that the final GDTA is as complete and accurate as possible, it should be validated by a larger group of subject matter experts. Printouts of the final GDTA can be distributed to experts with instructions on how to interpret it, and the SMEs asked to identify missing information or errors. Needed corrections can then be made. In some cases a particular expert will report that he or she does not consider certain information or do particular subsets of the GDTA. This feedback does not necessarily mean that these items should be eliminated, however. If other experts report using that data or performing those goals, then these components should remain part of the GDTA. Some people will have had slightly different experiences than others and simply may have never been in a position to execute all possible subgoals. As the GDTA will form the basis for future design, the full breadth of possible operations and information requirements should be considered.

An additional way to validate the GDTA is through observation of actual or simulated operations by experienced personnel in the position. While it can be difficult to always know exactly what the operators are thinking (unless they are instructed to 'think aloud,' performing a verbal protocol), these observations can be used to check the completeness of the GDTA. It should be possible to trace observed actions, statements, and activities back to sections of the GDTA. If people are observed to be performing tasks that there is no apparent reason for in the GDTA, or looking for information that is not identified in the GDTA, follow up should be conducted to determine why. Any additions to the GDTA that are needed should be made based on these validation efforts.

7.5. Simplifying the Process

Creating a GDTA can be a seemingly overwhelming task at times. Several suggestions can be offered to simplify this process.

- Organize the notes into the preliminary GDTA as soon as possible after the first interview while the conversation is still fresh in memory.
- Once a draft of the GDTA has been created, number the pages to facilitate discussion and minimize confusion. During a collaborative session, locating page 7 is easier than locating Subgoal number 4.2.3.1.

- Using paper copies of the GDTA when reorganizing the GDTA is often easier than trying to edit it in electronic format, as the pages can be arrayed across the table, elements compared, and changes notated quickly. Once changes have been decided, notate all the changes on a clean copy for review before updating the electronic format.
- As soon as the reorganization process allows, make a new primary goal hierarchy to provide structure for other changes.
- While the GDTA is being developed, make notes of any questions that arise concerning where or how something fits. This list can be brought to the next interview session for clarification.
- Do not get too concerned about delineating the different SA levels; these are not concrete, black and white items, and it is conceivable that a particular item can change SA levels over time. Thinking about the different SA levels is mainly an aid to help in the consideration of how information is used.
- If the same thing is being considered at various places to achieve essentially the same goal, combine them. For example, the questions 'How does the enemy COA affect ours?' and 'How does the enemy COA affect battle outcome?' are essentially the same and can be combined.
- During the interview listen closely for predictions or projections the person may be making when reaching a decision. Try to distinguish between what they are assessing about the current situation (e.g., 'where is everyone at?' – the Level 1 SA requirement "location of troops") and what they are projecting to make a decision, (e.g., "where will the enemy strike next?" – the Level 3 SA requirement "projected location of enemy attack"). Distinguishing between these types of statements will assist the practitioner in following up on higher order level SA requirements and documenting lower level SA requirements.

7.6. Conclusion

Creating a comprehensive GDTA for a particular job can be difficult. It often takes many interviews with subject matter experts; it is not uncommon for it to take anywhere from 3 to 10 sessions, depending on the complexity of the domain. Even then, there can be a fair degree of subjectivity involved on the part of the analyst and the experts. These problems are common to all cognitive task analysis methods, and GDTA is no exception. Nonetheless, the GDTA methodology provides a sound approach to delineating the SA requirements decision makers have with respect to specific goals. Understanding these SA requirements will assist designers in evaluating and designing systems to ensure that the decision maker's efforts to build and maintain a high level of situation awareness are supported to the maximum extent possible.

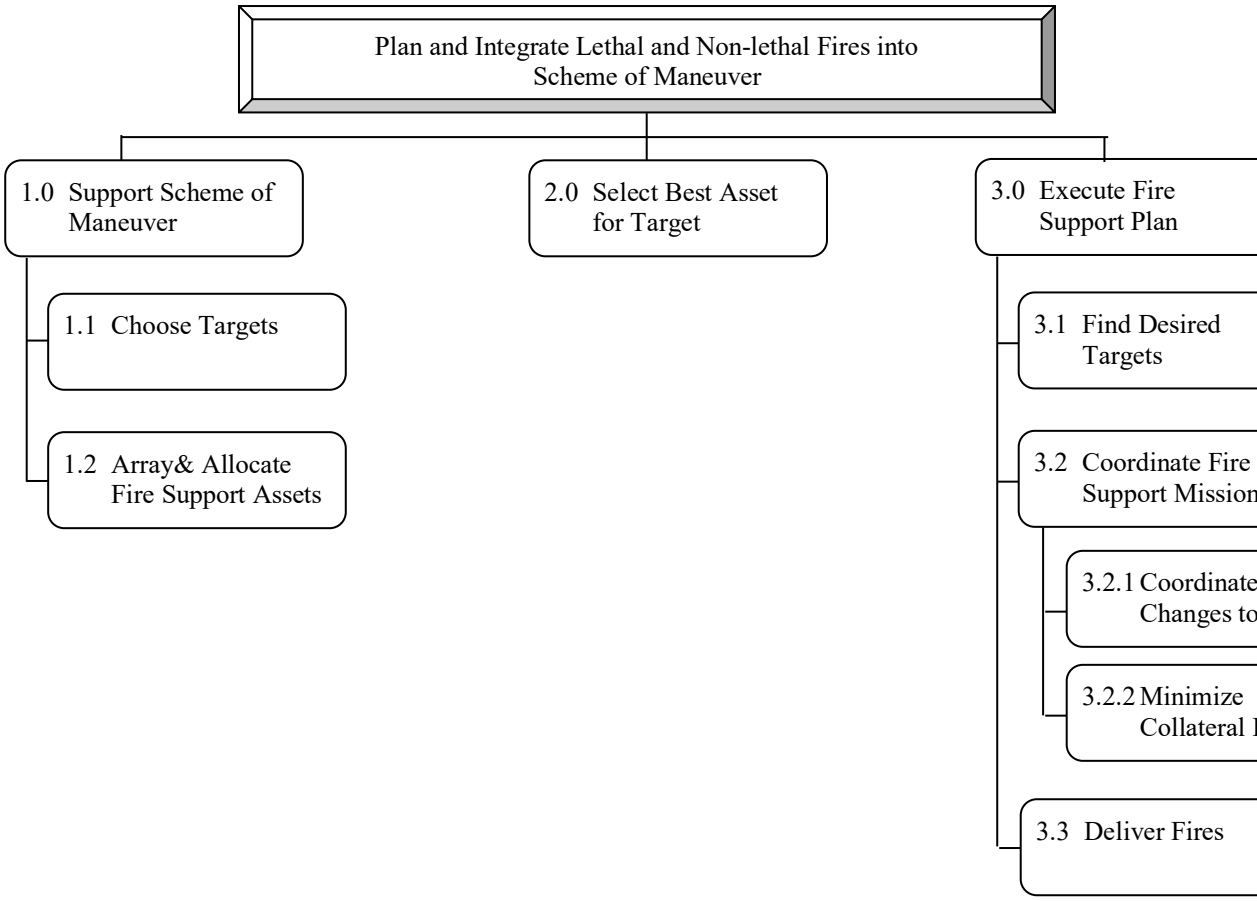
The resulting analysis from this effort feeds directly into efforts to design systems to support situation awareness, in that the structure provides

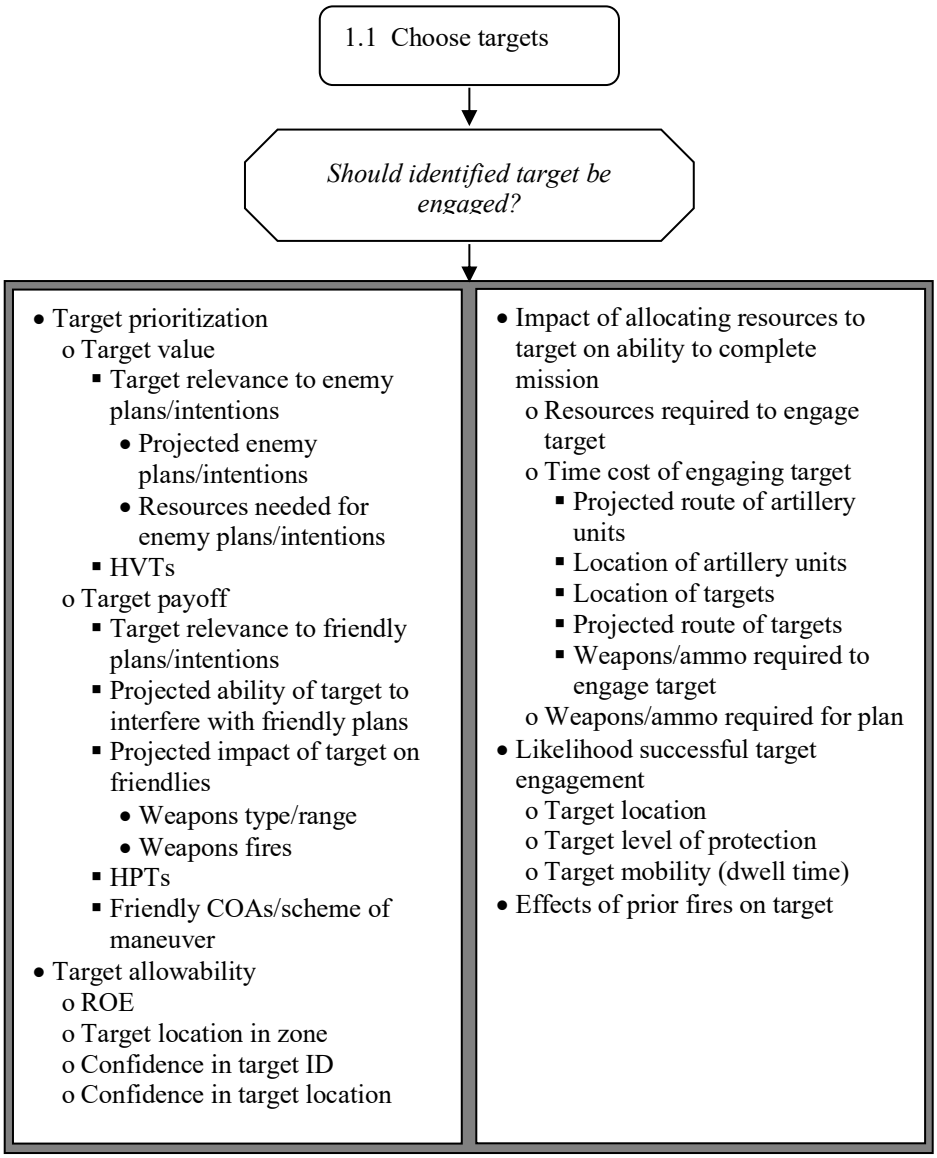
(a) a clear delineation of what the individual really needs to know, allowing designers to integrate data in meaningful ways on displays,
(b) key information on what information needs to be grouped together to support decisions,
(c) guidance as to how information and decisions relate to common goals, a critical organizing feature for system design, and
(d) information on critical cues/situations that dictate necessary shifts in priority between situation classes and goals.

This information is used directly in SA-Oriented Design (Endsley, Bolte, & Jones, 2003) to create systems that support SA, based on a set of clearly defined design principles. In addition, the SA requirements identified through the GDTA can be used to create objective metrics for evaluating the degree to which different design concepts and technologies are successful in supporting the SA of decision makers (Endsley, 2000). Overall, the GDTA provides a useful

tool for assessing and organizing the situation awareness requirements associated with cognitive work in a wide variety of domains and applications.

Army Fire Support Officer GDTA


Plan and Integrate Lethal and Non-lethal Fires into Scheme of Maneuver
1.0 Support Scheme of Maneuver
2.0 Select Best Asset for Target
3.0 Execute Fire Support Plan
1.1 Choose Targets
1.2 Array& Allocate Fire Support Assets
3.1 Find Desired Targets
3.2 Coordinate Fire Support Mission
3.2.1 Coordinate Changes to
3.2.2 Minimize Collateral
3.3 Deliver Fires


1.1 Choose targets
Should identified target be engaged?
• Target prioritization
o Target value
▪ Target relevance to enemy plans/intentions
• Projected enemy plans/intentions
• Resources needed for enemy plans/intentions
▪ HVTs
o Target payoff
▪ Target relevance to friendly plans/intentions
▪ Projected ability of target to interfere with friendly plans
▪ Projected impact of target on friendlies
• Weapons type/range
• Weapons fires
▪ HPTs
▪ Friendly COAs/scheme of maneuver
• Target allowability
o ROE
o Target location in zone
o Confidence in target ID
o Confidence in target location
• Impact of allocating resources to target on ability to complete mission
o Resources required to engage target
o Time cost of engaging target
▪ Projected route of artillery units
▪ Location of artillery units
▪ Location of targets
▪ Projected route of targets
▪ Weapons/ammo required to engage target
o Weapons/ammo required for plan
• Likelihood successful target engagement
o Target location
o Target level of protection
o Target mobility (dwell time)
• Effects of prior fires on target

1.2 Array Assets to Best Advantage

How accurate is my representation of the battlefield?

- Confidence in information
  - o Verified information
  - o Assumptions
- Projected friendly movement / actions (Bde, Adjacent Bde)
  - o Current friendly maneuver/actions
    - ▪ Friendly location/strength
    - ▪ Routes of movement
    - ▪ Observer location
- Projected friendly movement / actions (Artillery Battery)
  - o Current maneuver/actions
    - ▪ Location/strength
    - ▪ Weapons
    - ▪ Ammo type/quantity
    - ▪ Radar types/location
  - o Coverage of battlefield
- Projected enemy movement/action
  - o Current enemy maneuver/action
    - ▪ Enemy location/strength
    - ▪ Enemy weapons
    - ▪ Enemy weapon coverage area
    - ▪ Enemy knowledge of friendlies (all)
    - ▪ Assumptions about enemy
    - ▪ Enemy doctrine
- Special areas of interest
  - o Fire support coordination measures
  - o Maneuver boundaries

How does the enemy threaten the scheme of maneuver?

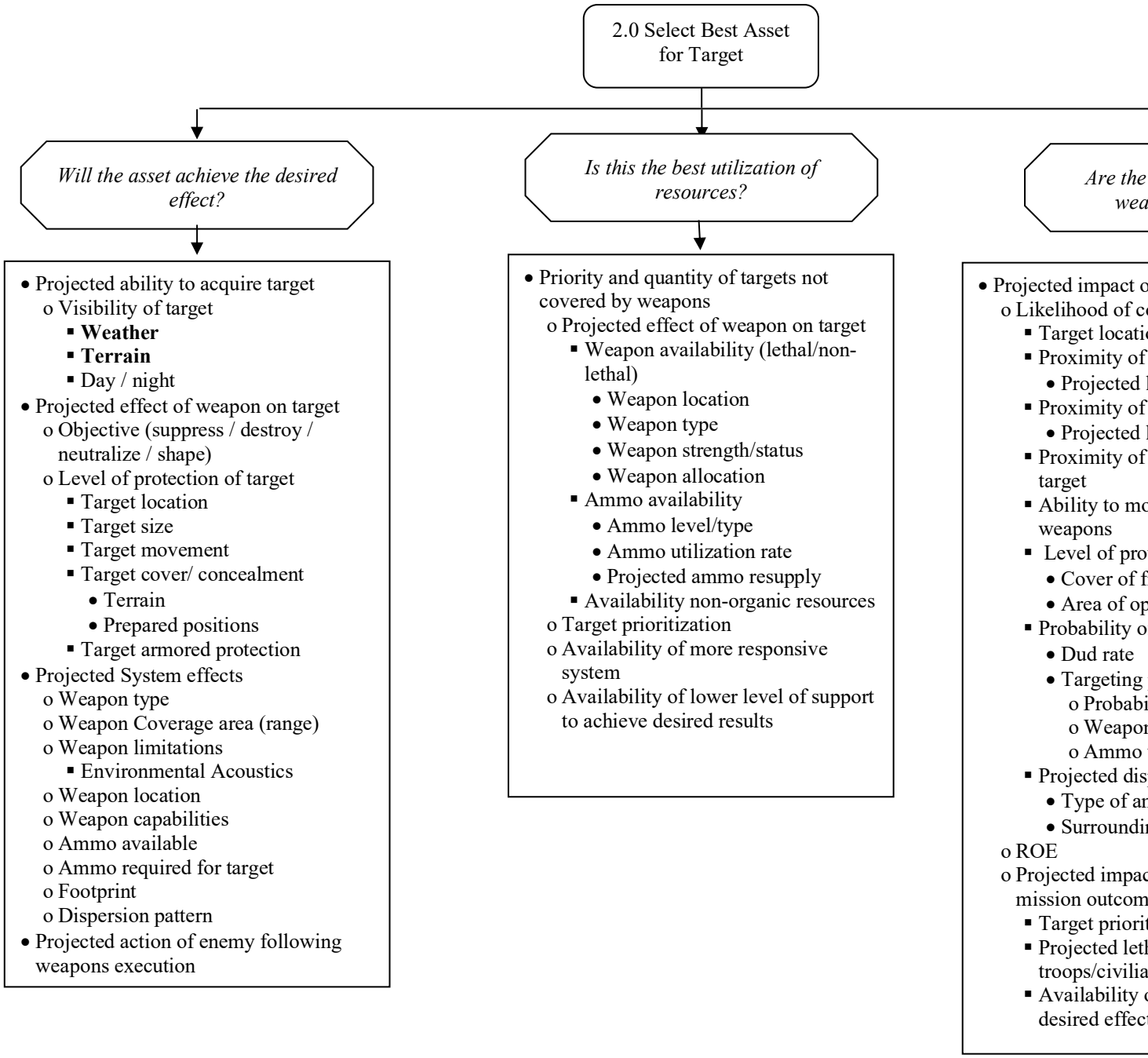
- Projected Impact on Friendly COA and maneuver
- CO Intent
- **Terrain**
- Projected friendly (all) movement / action
  - o Current friendly maneuver / action
    - ▪ Location / strength
    - ▪ Routes of movement
- Projected enemy movement / action
  - o Current enemy maneuvers/action
    - ▪ Enemy location / strength
    - ▪ Enemy weapons
    - ▪ Enemy weapon coverage area
    - ▪ Enemy knowledge of Friendlies (all)
    - ▪ Assumptions about enemy
    - ▪ Enemy doctrine

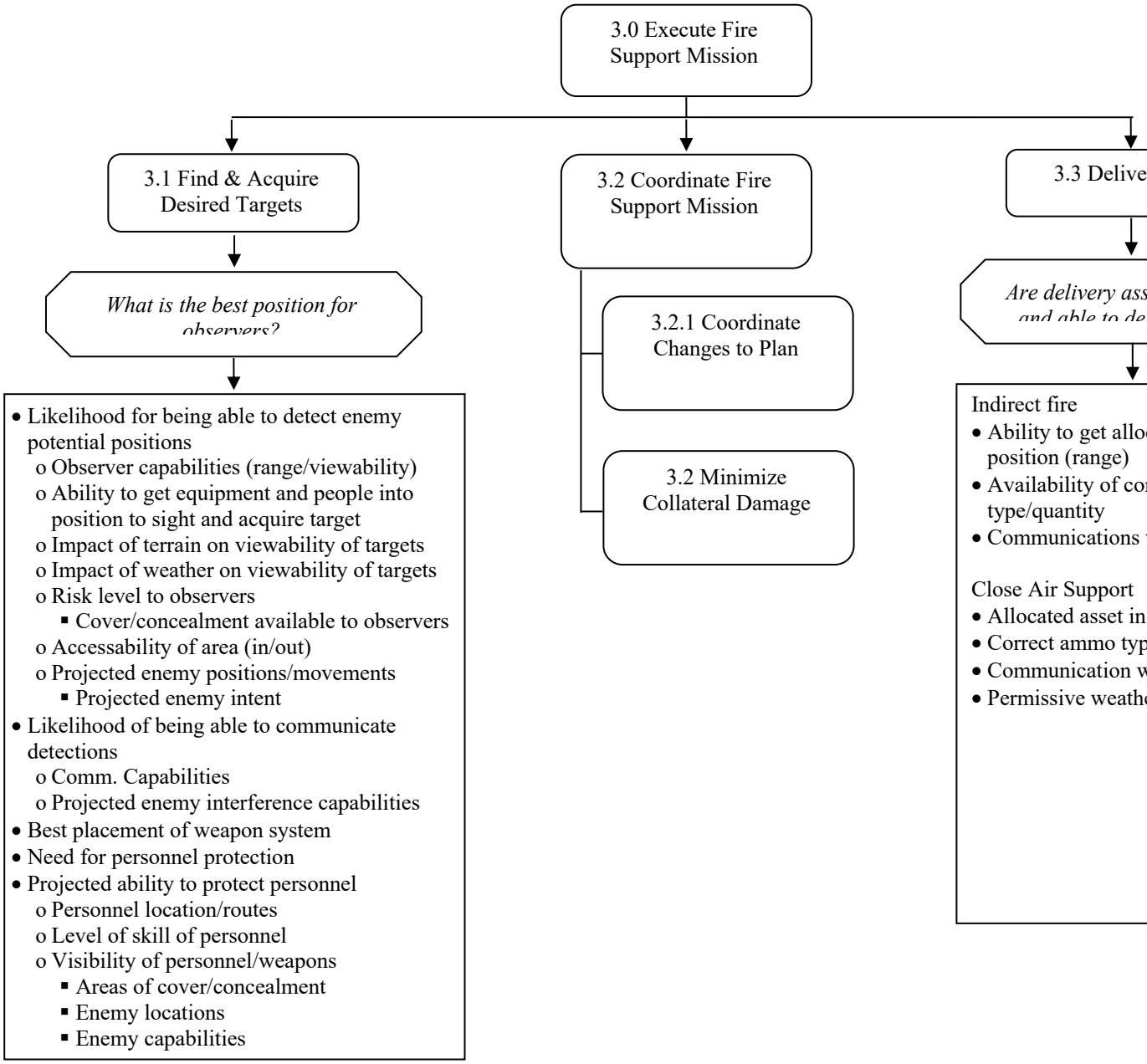
Is plan supportable with artillery?

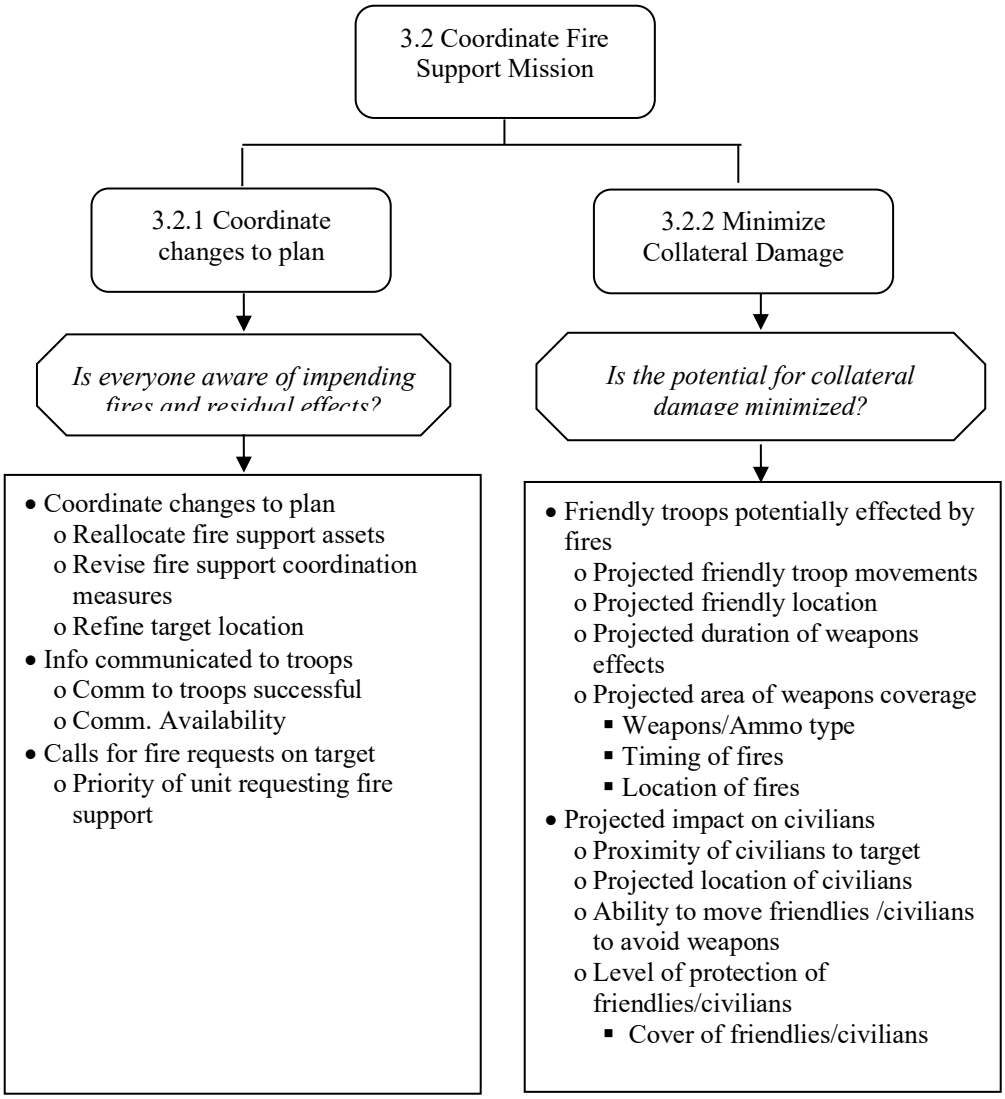
- CO Intent
- Projected impact of weapon on enemy
  - o Range and location
  - o Ammo capabilities
  - o Level of protection of enemy
  - o Timing of plan (ability to meet with weapons)
- Projected availability of other friendly (non-organic) assets
  - o Supporting fires (artillery, mortars, naval gunfire, air support)
  - o Aircraft availability
  - o Support artillery
    - ▪ From higher
    - ▪ From adjacent
  - o Electronic Warfare

How car

- Projected im
  - o Time cons
  - o Time avai
  - o Time requ
- Projected abi
- Projected tim concealment
- Projected fri
  - o Current m
    - ▪ Locatio
- Projected ene
  - o Current en
    - ▪ Enemy
    - ▪ Enemy
    - ▪ Enemy
    - ▪ Enemy
- **Terrain**
  - o Cover and
    - ▪ Positio
    - ▪ Maskin
  - o Foliage ef
  - o Munitions
  - o Protection
  - o Impact on
  - o Impact on
  - o Impact on
  - o Weather
  - o Ability to
  - o Line of sig
  - o Ability to
- Area of oper


2.0 Select Best Asset for Target
Will the asset achieve the desired effect?
Is this the best utilization of resources?
Are the wea
• Projected ability to acquire target
o Visibility of target
▪ Weather
▪ Terrain
▪ Day / night
• Projected effect of weapon on target
o Objective (suppress / destroy / neutralize / shape)
o Level of protection of target
▪ Target location
▪ Target size
▪ Target movement
▪ Target cover/ concealment
• Terrain
• Prepared positions
▪ Target armored protection
• Projected System effects
o Weapon type
o Weapon Coverage area (range)
o Weapon limitations
▪ Environmental Acoustics
o Weapon location
o Weapon capabilities
o Ammo available
o Ammo required for target
o Footprint
o Dispersion pattern
• Projected action of enemy following weapons execution
• Priority and quantity of targets not covered by weapons
o Projected effect of weapon on target
▪ Weapon availability (lethal/non-lethal)
• Weapon location
• Weapon type
• Weapon strength/status
• Weapon allocation
▪ Ammo availability
• Ammo level/type
• Ammo utilization rate
• Projected ammo resupply
▪ Availability non-organic resources
o Target prioritization
o Availability of more responsive system
o Availability of lower level of support to achieve desired results
• Projected impact o
o Likelihood of c
▪ Target locati
▪ Proximity of
• Projected
▪ Proximity of
• Projected
▪ Proximity of target
▪ Ability to mo weapons
▪ Level of pro
• Cover of f
• Area of op
▪ Probability o
• Dud rate
• Targeting
o Probabi
o Weapon
o Ammo
▪ Projected dis
• Type of ar
• Surroundi
o ROE
o Projected impac mission outcom
▪ Target priorit
▪ Projected letl troops/civilia
▪ Availability desired effec


3.0 Execute Fire Support Mission
3.1 Find & Acquire Desired Targets
3.2 Coordinate Fire Support Mission
3.3 Delive
What is the best position for observers?
3.2.1 Coordinate Changes to Plan
3.2 Minimize Collateral Damage
Are delivery ass
and able to de
• Likelihood for being able to detect enemy potential positions
o Observer capabilities (range/viewability)
o Ability to get equipment and people into position to sight and acquire target
o Impact of terrain on viewability of targets
o Impact of weather on viewability of targets
o Risk level to observers
▪ Cover/concealment available to observers
o Accessability of area (in/out)
o Projected enemy positions/movements
▪ Projected enemy intent
• Likelihood of being able to communicate detections
o Comm. Capabilities
o Projected enemy interference capabilities
• Best placement of weapon system
• Need for personnel protection
• Projected ability to protect personnel
o Personnel location/routes
o Level of skill of personnel
o Visibility of personnel/weapons
▪ Areas of cover/concealment
▪ Enemy locations
▪ Enemy capabilities
Indirect fire
• Ability to get allo
position (range)
• Availability of co
type/quantity
• Communications
Close Air Support
• Allocated asset in
• Correct ammo typ
• Communication w
• Permissive weath


3.2 Coordinate Fire Support Mission
3.2.1 Coordinate changes to plan
3.2.2 Minimize Collateral Damage
Is everyone aware of impending fires and residual effects?
Is the potential for collateral damage minimized?
• Coordinate changes to plan
o Reallocate fire support assets
o Revise fire support coordination measures
o Refine target location
• Info communicated to troops
o Comm to troops successful
o Comm. Availability
• Calls for fire requests on target
o Priority of unit requesting fire support
• Friendly troops potentially effected by fires
o Projected friendly troop movements
o Projected friendly location
o Projected duration of weapons effects
o Projected area of weapons coverage
▪ Weapons/Ammo type
▪ Timing of fires
▪ Location of fires
• Projected impact on civilians
o Proximity of civilians to target
o Projected location of civilians
o Ability to move friendlies /civilians to avoid weapons
o Level of protection of friendlies/civilians
▪ Cover of friendlies/civilians

8. References

Burton, A. M., Shadbolt, N. R., Hedgecock, A. P., and Rugg, G. (1987). A formal evaluation of a knowledge elicitation techniques for expert systems: Domain 1. In D. S. Moralee (Ed.), *Research and development in expert systems, Vol 4*. (pp. 35-46). Cambridge: University Press

Burton, A. M., Shadbolt, N. R., Rugg, G., and Hedgecock, A. P. (1988). A formal evaluation of knowledge elicitation techniques for expert systems: Domain 1. Proceedings, *First European Workshop on Knowledge Acquisition for Knowledge-Based Systems* (pp. D3.1-21). Reading, UK: Reading University.

Chase, W. G., and Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 5*, 55-81.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement, *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97-101). Santa Monica: Human Factors Society.

Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *International Journal of Aviation Psychology, 3*, 157-168.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley and D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 147-174). Mahwah, NJ: LEA.

Endsley, M. R., Bolte, B., and Jones, D. G. (2003). *Designing for situation awareness: An approach to user-centered design.* London: Taylor and Francis.

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

Fox, J. (Ed.) (1987). *The essential Moreno.* New York: Springer.

Gilbreth, F. B. (1911). *Motion study*. New York: Van Nostrand.

Gilbreth, F. B., and Gilbreth L. M. (1919). *Fatigue study: The elimination of humanity's greatest unnecessary waste*. New York: MacMillan.

Hoffman, R. R. (1986). "Extracting experts' knowledge." Report, U.S. Air Force Summer Faculty Research Program.

Hoffman, R. R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. *AI Magazine, 8*, 53-67.

Hoffman, R. R. (1998). How can expertise be defined? Implications of research from cognitive psychology. In R. Williams and J. Fleck (Eds.), *Exploring expertise: Issues and perspectives*. London: Macmillan.

Hoffman, R. R., Coffey, J. W., and Ford, K. M. (2000). "A Case Study in the Research Paradigm of Human-Centered Computing: Local Expertise in Weather Forecasting." Report on the Contract, "Human-Centered System Prototype," National Technology Alliance.

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

Hoffman, R. R., and Lintern, G. (2005). Eliciting knowledge from experts. In A. Ericsson, N. Charness, P. Feltovich, and R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance.* New York: Cambridge University Press.

Hoffman, R. R., Militello, L., and Eccles, D. (2005). “Perspectives on Cognitive Task Analysis.” Report to the Advanced Decision Architectures Collaborative Alliance, Army Research Laboratory, Adelphi, MD.

Hoffman, R. R., Shadbolt, N., Burton, A. M., and Klein, G. A. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes*, *62*, 129-158.

Hoffman, R. R., Trafton, G., and Roebber, P. (2005), *Minding the weather: How expert forecasters reason*. Cambridge, MA: MIT Press.

Johnston, N. (2003). The Paradox of Rules: Procedural Drift in Commercial Aviation. In R. Jensen, (Ed), *Proceedings of the Twelfth International Symposium on Aviation Psych* (pp. 630-635) Dayton, OH: Wright State University.

Kidd, A. L., and Cooper, M. B. (1985). Man-machine interface issues in the construction and use of an expert system. *International Journal of Man-Machine Studies, 22*, 91-102.

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

Kolodner, J. L. (1991). Improving decision making through case-based decision aiding. *The AI Magazine, 12*, 52-68.

Koopman, P., and Hoffman, R. R., (November/December 2003). Work-Arounds, Make-Work, and Kludges. IEEE: Intelligent Systems, pp. 70-75.

McClelland, D. C. (1998). Identifying competencies with behavioral-event interviews. *Psychological Science, 9*, 331-339.

McDonald, N., Corrigan, S. and Ward, M. (2002). Cultural and Organizational factors in system safety: Good people in bad systems. *Proceedings of the 2002 International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero 2002)* (pp. 205-209). Menlo Park, CA: American Association for Artificial Intelligence Press.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising, 19*, 39-52.

Prerau, D. (1989). *Developing and managing expert systems: Proven techniques for business and industry*. Reading, MA: Addison Wesley.

Rasmussen, J., Pejtersen, A. M., and Schmidt, K. (1990). *Taxonomy for cognitive work analysis*. (Report No. RIS∅-M-2871). Roskilde, Denmark: Riso National Laboratory.

Renard, G. (1968). *Guilds in the Middle Ages*. New York: A. M. Kelley.

Senjen, R. (1988). Knowledge acquisition by experiment: Developing test cases for an expert system. *AI Applications in Natural Resource Management, 2*, 52-55.

Slade, S. (1991). Case-based reasoning: A research paradigm. *The AI Magazine*, 42-55.

Stein, E. (1992). A method to identify candidates for knowledge acquisition. *Journal of Management and Information Systems, 9*, 161-178.

Stein, E. (1997). A look at expertise from a social perspective. In P. J. Feltovich, K. M. Ford, and R. R. Hoffman (Eds.). *Expertise in context: Human and machine* (pp. 181-194). Cambridge, MA: MIT Press/AAAI Books.

Thorndike, E. L. (1920). The selection of military aviators. *U.S. Air Service, 1 and 2*, 14-17; 28-32, 29-31.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work.* Mahwah, NJ: Erlbaum.

Wood, L. E., and Ford, K. M. (1993). Structuring interviews with experts during knowledge elicitation. In K. M. Ford and J. M. Bradshaw (Eds.), *Knowledge acquisition as modeling* (pp. 71-90). New York: Wiley.