ARI Contractor Report 2001-05

Commander's Adaptive Thinking Skills Tutor (CATS Tutor) – Phase I Final Report

William R. Murray and Michelle Sams
Teknowledge Corporation

This report is published to meet legal and contractual requirements and may not meet ARI's scientific or professional standards for publication.

August 2001

United States Army Research Institute for the Behavioral and Social Sciences

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) August 2001	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) December 15, 2000 – June 15, 2001
4. TITLE AND SUBTITLE Commander's Adaptive Thinking Skills Tutor (CATS Tutor) – Phase I Final Report		5a. CONTRACT OR GRANT NUMBER DASW01-01-P-0352
		5b. PROGRAM ELEMENT NUMBER 0602785A
6. AUTHOR(S) William R. Murray and Michelle Sams (Teknowledge Corp.)		5c. PROJECT NUMBER A790
		5d. TASK NUMBER 211
		5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Teknowledge Corporation 1810 Embarcadero Road Palo Alto, CA 94303		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue ATTN: TAPC-ARI-IK Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER Contractor Report 2001-05

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES: Contracting Officer's Representative: Dr. James W. Lussier.
This report is published to meet legal and contractual requirements and may not meet ARI's scientific or professional standards for publication.

14. ABSTRACT *(Maximum 200 words)*:

This report describes the results of a SBIR Phase I effort and plans for a Phase II development. This project will apply intelligent tutoring technology to teach battlefield command reasoning through deliberate practice in key scenarios and tactical simulations. It is called the Commanders' Adaptive Thinking Skills Tutor (CATS Tutor). The CATS Tutor selects the training activities and provides explanatory feedback based on an evolving model of the individual learner. It draws on an ontology to provide a deep knowledge base and reasoning capabilities to infer a learner's strengths and weaknesses. The primary focus is on inferring the learner's cognitive strategies by their behavior rather than through primary reliance on questioning and thinking aloud protocols. The tutor's instructional strategies build on techniques designed to teach expert cognitive strategies, that is, how to think rather than what to think. Report developed under the Small Business Innovation Research Program 99.2 contract for topic OSD-CR02.

15. SUBJECT TERMS: intelligent tutoring system, cognitive strategies, tactical simulations, ontology, adaptive thinking skills, commanders, SBIR report

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	30	Dr. James W. Lussier DSN: 464-2613

COMMANDER'S ADAPTIVE THINKING SKILLS TUTOR (CATS Tutor) - PHASE I FINAL REPORT

CONTENTS

PAGE

LIST OF FIGURES

Commander's Adaptive Thinking Skills Tutor (CATS Tutor)

Executive Summary

The primary operational challenge facing the U.S. military for the foreseeable future is that military planners cannot know *a priori* specifically where our military will be deployed, who our adversaries will be, and what weapons and systems will be necessary for a given mission. There is an increased emphasis on both the broader spectrum of military missions, such as peacekeeping, and a broader range of potential adversaries, such as terrorists. Requirements for future education and training systems must take due account of this expanding range of operational conditions and needs. Our forces must be highly adaptive learning forces that organize to meet threats effectively and rapidly.

In particular, commanders must be able to make and implement decisions in a timely, efficient, and effective manner, most often with very limited information in a constantly changing, complex, and dangerous environment. This capability is often referred to as cognitive readiness--ensuring that the warfighter is mentally prepared for accomplishing the mission and is performing at their optimal performance level.

The challenge is to understand what constitutes expertise in military command decision making and then how to effectively teach and improve the skills of less experienced commanders. Less skilled and experienced tacticians conceptually understand battlefield thinking habits, however these behaviors are often absent during realistic tactical problem solving (Lussier, 1999).

While nothing can truly substitute for live exercises and real-world experience, computers are a good medium for training a number of different tasks and skills. Computers can present an unlimited number of exercises and scenarios for skill practice with immediate review and feedback tailored to the individual. Computer-based training is particularly useful for acquiring and refreshing skills, and then real world training exercises can focus on honing those skills.

The Army Research Institute (ARI) is developing and testing a concept for teaching battlefield command reasoning through deliberate practice, with focus on improving thinking skills (Deckert, et.al., 1996). In deliberate practice, the learner exercises the same skill sets in a variety of situations until those skills are applied both flexibly and automatically. In one application of this research, ARI is translating this concept into computer-supported, case-based exercises called *Think Like a Commander*. This program presents tactical situations and asks the learner to reason, reach conclusions, and make decisions. However, this program requires *a live expert* to be present during the exercises to provide prompting, advice, feedback, and explanations, as needed.

The goal of this SBIR project is to make the type of training offered in the *Think Like a Commander* (TLAC) course more cost-effective and available to a larger number of officers as an application of distributed learning – that is, it will not require a live expert to serve as a tutor.
This project has both a product and a research focus. The product focus is on developing an intelligent tutoring system to teach battlefield command reasoning through deliberate practice. We call it the Commanders' Adaptive Thinking Skills Tutor (CATS Tutor). The practice is targeted to improve some aspect of the required thinking skills through various exercises and activities within tactical simulations. With enough deliberate practice and tailored guidance, the thinking skills will become automatic.

The tutor selects the training activities and provides explanatory feedback based on an evolving model of the individual learner. It draws on an ontology to provide a deep knowledge base and reasoning capabilities to infer a learner's strengths and weaknesses. The primary focus is on inferring the learners cognitive strategies by their behavior rather than through primary reliance on questioning and thinking aloud protocols. The tutor's instructional strategies build on techniques designed to teach expert cognitive strategies, that is, how to think rather than what to think.

This brings us to one of the key research components of the project -- how to model and evaluate the cognitive strategies of learners. The cognitive strategies for this domain are adaptive thinking skills in tactical situations. Adaptivity is required under the pressures of time, risk, and complexity (many interacting system elements). Aspects of these thinking skills have been elucidated in various ways: 8 thinking habits in the TLAC course (e.g., model a thinking enemy, consider effects of terrain), METT-TC (Mission, Equipment, Terrain, Troops, Time, and Civilians) and BOS (Battlefield Operating Systems, such as armor, artillery, engineering, and air support). The CATS Tutor is *not* specifically teaching the 8 principles of the Think-Like-a-Commander course, or any other skill organizer. It draws on a rich domain representation to help learners develop their own expert cognitive strategies.

Most intelligent tutoring systems have focused on evaluating whether learners have the requisite concepts, rules, procedural skills, or heuristics. Few have addressed cognitive strategies as the primary knowledge to be modeled and evaluated. A new approach to user modeling will be investigated where different cognitive strategies are simulated and the results (inferences drawn and actions taken) are partially matched to user actions in specific task situations.

The CATS Tutor is designed as a generic tutor for critical thinking skills. Its open, flexible framework enables training strategies to applied to a wide range of products and DoD needs. The tutor is not tightly integrated to the training content or the simulation. The tutor has 'hooks' to the tactical simulation, with information being passed (communicated) between the simulation and the tutor agent. With this type of communication design, other simulations can be more easily substituted for tutorial interactions supporting other topic domains.

The CATS Tutor provides repeatable, scenario-based skill practice that should transfer well to physical training sites and real world operations. In addition, the intelligent tutoring system pushes the training technology envelope by providing tailored guidance for a challenging domain, battlefield command reasoning.

Results of Phase I

During the Phase I effort, we accomplished the following objectives:

- Revised and refined our concept of commanders' adaptive thinking skills
- Investigated various approaches for the CATS Tutor
- Developed a detailed design of the CATS Tutor
- Developed a proof-of-concept implementation

A brief description of the Phase I approach and the results of these activities are discussed.

Revised and Refined Our Concept of Commanders' Adaptive Thinking Skills

Clarified that the purpose of this SBIR project is primarily intended to assess tools, methods, and techniques of applying advanced automated training technology to battle command thinking tasks and to demonstrate the feasibility or lack thereof of pursuing such training methods for this type of behavior. The focus of this project is on the thinking process not the training content (e.g., not focused on teaching the TLAC course or vignettes). The training system activities and assessments should focus on a commander's behavior (execution) rather than formulation of a specific plan.

The challenging aspects of the project will be to:

- mentor a commander's adaptive thinking skills in a manner that facilitates rather than inhibits the thinking process.
- develop valid measures and methods that evaluate a commander's thinking skills.
- design a training environment that challenges and engages the trainee in realistic activities.

More discussion on the concept is presented in the section, Phase II, Adaptive Thinking Skills.

Clarification and revision of the project's concept was based on:

Discussions

Discussions with ARI psychologists at the kick-off meeting helped tremendously to clarify the project's objectives. At various times during Phase I, we sent concept documents and exchanged e-mails with ARI and received further helpful guidance towards refinement of the concepts.

Participated in the BCT on-line forum. This is an on-line forum for questions about the Brigade Combat Team (BCT) tactical simulation and how it models tactics, weapons, and for any other questions related to BCT. Most contributors are active or retired military personnel, including one involved with WARSIM 2000 and a retired LTC with Armor experience. It is useful to hear how soldiers discuss BCT, especially those who have used JANUS and TacOps and participated in NTC and JTRC exercises. In general, they prefer BCT to JANUS and TacOps, and some say they have learned more with BCT than in some of their training rotations.

Classroom Observation and SME Input

We observed two CGSC classes at Ft Leavenworth in March 2001. LTC Prevou led the trainees in discussing two TLAC vignettes. LTC Prevou was very engaging and effective, getting all trainees to participate, in contrast to some of the other videotaped sessions we viewed. After the class, we talked with LTC Prevou about our initial concept plans and received some valuable feedback and ideas. We also received valuable advice and feedback from Rex Michel and Major Trevor Marshall.

Relevant Materials

We obtained and evaluated various materials to develop our understanding of the domain and project concepts.

- Digital files containing 5 TLAC vignettes, TLAC briefing slides by Dr. Lussier (provided by ARI)

- VHS tape of several CGSC classroom sessions covering TLAC vignettes (provided by ARI)
- CD ROM on Objective Force (provided by ARI), which contained a very useful discussion of adaptive training and the Adaptive Thinking Training Methodology and its application of deliberate practice to thinking and reasoning tasks.
- ARI tech report on commander's reactions to the TCDC (Tactical Commander's Development Course) (Lussier and Litavec, 1992).
- JANUS materials available on STRICOM web-site, for comparison with BCT.
- The Operations and Training SMART book: Guide to Operations and the Battlefield Operating Systems" (Wade, 1999a) (a book we found at the CGSC bookstore which covers many military key concepts in a way that is more readable than the Field Manuals).
- "The Battle Staff" (Wade, 1999b) (addresses MDMP, orders, and plans in readable detail)
- FM-101-5, Operational Symbols and Graphics (Dept of Army, 1997)
- *Think Like a Grandmaster* (Kotov, 1971) and *Secrets of Chess Training* (Dvoretsky, 1991) to gain better insight to the origins of Dr. Lussier's approach to adaptive thinking skills.
- *Armored Cav* by Tom Clancey, a nonfiction guide to an ACR.
- *The Art of War* (Sun Tzu, 1971). Classic text on military thinking.
- *On War* (Clausewitz, 1993). Another classic text on both strategy and tactics.
- *The Art of Maneuver: Maneuver-Warfare Theory and AirLand Battle.* (Leonhard, 1994).
- *Strategy.* B. H. (Hart, 1991). Another classic on military strategy.
- *Jane's Tanks and Combat Vehicles Recognition Guide.* 2nd Edition (Foss, 2000). Useful guide to help visualize vehicles such as T-72s, T-80s, T-90s, FIST-Vs, BMPs, BRDM-2 (ATGW), 2S19 152mm, and all the many others used in BCT and the contemporary combined arms warfare scenarios it depicts.
- *Breaching 101.* Mike Robel, retired Armor LTC, member WARSIM 2000 project. Useful guide for mistakes in breaching and how it's supposed to be done. Reprinted by permission in the BCT User's Guide manual.

Investigated Various Approaches for the CATS Tutor

We considered various approaches for the CATS Tutor, and shaped our design by our evolving concept of how best to coach adaptive thinking skills for this domain. We provided PPT mockups and concept papers to ARI for feedback and suggestions. We modified our design based on critical issues pointed out by SMEs (at ARI and CGSC) and had to constrain the plan for what could be accomplished within the scope of a Phase II development effort.

A first inclination would be to try to emulate the human tutorial interactions that occur in the TLAC class. We explored and then ultimately rejected this approach based on several factors. The focus of this project is on evaluating and guiding the thinking processes of commanders. We were interested in approaches that inferred these thinking processes through evaluation of behavior rather than through primary reliance on questioning and 'thinking aloud protocols'. In particular, we were interested in approaches that leveraged capabilities that computers are good at and that could be extended to other domains.

Realistic human tutorial dialog requires a robust natural language understanding (NLU) system. From our experience with various types of NL approaches and projects, simpler NL approaches can support mostly scripted dialog interactions and full-up NLU just isn't available yet. By full-up NLU, we mean a system that has a comprehensive, robust parser, semantics, ontology and knowledge base, and dialog management capabilities. It is only this full-up NLU approach that has promise for natural conversational and tutorial dialog. Currently this type of NLU system is not robust, has uneven syntactic and semantic coverage and falls off the 'conversational cliff' quite easily. Continued research in NLU will result in more robust systems, but we believe that is 5-10 years out.

Recently there has been a great deal of attention paid to latent semantic analysis (LSA). LSA has been touted as being able to understand and evaluate natural language input. However, LSA performs no functions towards 'understanding'. It relies on a statistical approach to compare the 'equivalence' of two sets of words, a user's input compared to statistically compiled volumes of text or exemplar answers. Since the comparison method is a heuristic, LSA is in essence a 'black box' approach. LSA cannot provide a common-sense explanation of why the input is similar or dissimilar, determine what the user's misconceptions might be, etc. We believe that LSA is best used for information retrieval functions rather than attempting to evaluate what a trainee does and does not know.

Our position is that there is a greater possibility of project success and effective tutoring with a focus on inferring the thinking processes of commanders through their interactions in scenarios, studies, and exercises within tactical simulations. We will leverage existing simulations and focus the development of the tutoring system on advances in user modeling, instructional strategies, and domain reasoning on a knowledge base. We would restrict user NL input to a modified NL query system with the tutor deriving an appropriate response from the domain knowledge. We include more detail and rationale for our approach in the section, Phase II Work Plan.

Developed a Detailed Design of the CATS Tutor

During Phase I we developed, modified, and refined a design for the CATS Tutor. We propose building an intelligent tutoring system for critical thinking skills applicable across multiple Army domains. The tutor leverages existing high-fidelity simulations that have been independently developed by subject-matter experts over a period of years. Its instructional strategies build on techniques designed to teach expert cognitive strategies, that is, how to think rather than what to think. It uses a shared upper ontology, a mid-level Army and Armor-specific ontology, and a domain-specific lower-level ontology to provide a deep knowledge base and reasoning capabilities, the majority of which can be shared across similar domains.

Such an intelligent tutoring system is unique in several respects. A new kind of instructional strategy and user model is required. Instead of only modeling the facts and rules of the domain, as in most intelligent tutoring systems, this tutor must model the cognitive strategies, which are one step further removed.

Deliberate practice is used to promote automaticity in certain kinds of thinking and is a key aspect of this strategy. A second key component is the use of case studies where trainees are placed in realistic situations. They must first determine the problem (i.e., threats), and then take action to address them, and to achieve the mission goals. The case studies are presented to the trainee using the simulation, and then a limited play-out of the problem follows, with the trainee playing against the computer's simulated opponent. During this play-out, the tutor tracks what the trainee does in the simulation, looking for omissions or actions indicating that the threats were correctly or incorrectly perceived or acted upon. The tutor can also freeze the simulation and ask questions about the trainee's reasoning to assist in its assessment.

The simulation part of the CATS Tutor is also used in conventional ways, as in earlier intelligent tutoring systems. The simulation can be used to demonstrate the thought process of subject matter experts. For example, a problem is presented and the solution played back along with an expert (e.g., a Colonel or General's) thinking and decisions about the problem. Scaffolding can be provided by offloading one or more aspects of the problem solving aspects onto the computer. For example, the computer can handle one or more battlefield operating systems while the trainee handles the rest (e.g., computer handles artillery and air, trainee handles everything else). This scaffolding approach reduces the cognitive load on the trainee while allowing them to see their skills used in the context of a complete problem solution.

The ability to share knowledge, instructional plans, and plug into different simulations to build tutors for different, but related subject-matter areas and training tasks also requires changes to the ITS architecture. It must be capable of supporting a distributed domain representation where a qualitative model resides in the tutor and a separate quantitative and graphical model resides in the attached simulation. Concepts, rules, heuristics, and strategies (e.g., tactical strategies) are encoded in ontologies that are not domain-specific. These ontologies are the substrate over which the tutor's knowledge bases (for domain expertise, user modeling, and pedagogical control) are built. The ontologies can be reused across multiple related domains and for multiple related training tasks.

For the CATS Tutor we do not intend to develop our own tactical simulations. It makes more sense to leverage existing, high fidelity simulations that can be used to illustrate and exercise commanders' thinking skills. Often a company with some capability in intelligent tutoring systems will undertake to build *both* the ITS part of a system and a simulation component. The result may be a passable simulation but these typically cannot match the fidelity that subject matter experts can provide if they build a simulation over a period of years. Furthermore the simulation is custom-built into the ITS architecture and is not readily separable, which makes reuse of the tutorial component more difficult.

Our design approach is to build the ITS component separate from the simulation component and then hook into the simulation. The subject matter expertise and teaching strategies used in CATS can then be applied to different simulations. Some modifications will be required of course but much of the knowledge and the instructional plans can be shared, assuming the target skills are in the critical thinking /adaptive thinking skills area and the simulation is a tactical simulation..

Details of the CATS Tutor design are included in the section, Phase II Work Plan. This section includes descriptions of the functional capabilities and technical design for each of the following aspects:

- learner model
- adaptive instructional strategy
- meta-level controller
- a knowledge base
- response to learner queries about the domain
- case library, tactical simulation
- tutor-simulation integration

Developed a Proof-of-Concept Implementation

Developed a proof-of-concept implementation to demonstrate feasibility and proposed functions. This prototype was shown to ARI at the Phase I final briefing meeting on June 6, 2001.

The Phase I prototype, which we will call CATS-1, has a basic learner model (currently derived from simple scoring methods). Trainees can engage in practice exercises, critiquing studies, and tactical simulations. Trainees can review a summary of their strengths and weaknesses based on the tutor's scoring method.

For the Phase I tactical simulations, we used the Brigade Combat Team (BCT) simulation program developed by CPT Patrick Proctor and now offered at a nominal cost ($35) through his company, ProSim. CPT Proctor based the development of BCT on his experiences working with JANUS at Fort Hood and as an artillery officer and at JTRC and NTC rotations. Many of the scenarios in BCT mirror NTC and JTRC exercises. In fact, JTRC has contracted ProSim to develop a special version of BCT for their use. (personal communication, April 2001)

Brigade Combat Team is a simulation of combined arms warfare. The game pits two opposing units of up to brigade/regimental size against each other in simulated combat. BCT allows a user to take the role of the friendly forces while it plays the OPFOR. It randomly selects from multiple enemy COAs stored with each scenario. The icons, which the user manipulates, represents sections or team of vehicles, which, when grouped together, form the companies and battalions of the unit. Just as in modern land warfare, the user fights with and against units consisting of a wide variety of vehicles. These include armor, infantry, artillery, engineers, air defense, and aircraft. These units must be synchronized and massed at the key point on the battlefield to win. The cybernetic battlefield is a digitized elevation map of actual terrain. Some of the battlefields ... include the Mojave Desert (site of the US Army National Training Center), Kuwait, and North Korea" (BCT Manual, p. 6).

We selected the BCT simulation for two main reasons. First, it is the most realistic PC-based tactical simulation at the brigade level currently available. It models terrain line of sight better than any previous PC game, including TacOps. It is a *real-time* simulation (1X, 2X, 4X, or 8X real-time) of combined arms warfare. Note that this real-time aspect will be important in helping trainees acquire an intuitive feel in how fast the battlefield changes, and help in learning to visualize the battlefield. BCT uses standard Army OPORDs and maps. It also models artillery better than any other PC game (not surprisingly, as Captain Proctor is an artillery officer). The second major reason is the availability and enthusiasm of Captain Proctor for working with us. To take full advantage of the simulation within the ITS context, we will need to be able to develop hooks to more tightly control the interaction and data collection between the simulation and the tutor.

The CATS-1 prototype includes a few examples of each of the following types of training activities:

Focused Exercises for Individual Thinking Habits

This is a part-task training activity, with focus on just one of the thinking habits at a time. Trainees predict interactions/outcomes for a selected 'slice' of a tactical situation.

Critiquing Studies

This is a training activity primarily focused on learning by observing, with elements of participation via the critiquing process. Trainees critique the decisions of other commanders.

Tactical Studies (Simulation Participation)

This is a whole task training activity where trainees make decisions within various tactical simulations that require application of multiple thinking skills.

Figure 1. CATS top level GUI

A brief description of these training activities, as implemented in CATS-1 follows:

Focused Exercises for Individual Thinking Habits

Trainees are presented with information and graphics about a scenario. They must predict the interaction or outcome for a given 'slice' of that situation. Currently, predictions are selected from a list of possible textual alternatives, or by selecting from pictures showing different possible battlefield situations. After the prediction, the tutor provides feedback and the trainee can opt to view the tactical situation being automatically played out to observe the outcome. The exercise can be tried out with different predictions as often as the trainee wishes. The battle outcome may be different each time (due to the stochastic nature of the simulation) but the key principle will be illustrated.

Note: We had considered using video clips of the simulation but decided that the variability in simulation results may help the trainee see the range of possibilities and further increase the acquisition of an intuitive feel for the battlefield mechanics illustrated.

Practice exercises for two of the eight thinking habits were implemented for CATS-1.

> *Consider all elements/systems available to you and your enemy and their interactions.* The trainee is given two different kinds of elements or systems (usually one friendly and one enemy, but it could be larger combinations, or friendly engineering assets versus enemy obstacles). The commander must predict the most likely interactions (selects from a list of alternative possibilities). Trainee receives feedback and can view the simulation play out.
>
> *Exhibit visualizations that are dynamic, proactive, and flexible.* The trainee is given a current state of the battlefield, along with the planned routes, current speed, etc. of each unit. The trainee must forecast what the battlefield will look like at some specified time in the future (trainee selects a picture 'snap shot' of the battlefield from a group of alternatives). Again the simulation can be played out if desired, or just the correct answer shown. Trainee receives feedback and can view the simulation play out.

Critiquing Studies

In these exercises the trainee critiques the performance of *another* (fictional) commander. The trainee is presented with a view of the battlefield and the plans of the fictional commander (e.g., planned unit paths, planned fire missions). The trainee selects critiques (from a list of alternatives) for one or more units. For example, if the trainee wants to critique a fire mission, he selects the artillery unit involved and types in an appropriate critique. After entering all unit critiques, the tutor provides explanatory feedback. As there may be 40 to 80 units on the battlefield, the trainee is encouraged to focus on the most important critiques (e.g., the top 5).

A case library stores a pool of scenarios that were played out in BCT and saved. The cases illustrate errors of commission and omission. Errors are detailed specific to the scenario (e.g., not obscuring the other side of a breach to protect engineering assets used in breaching). The errors are also mapped to higher order principles (e.g., not protecting your breaching assets). These higher order principles are taken from each of the organizers for thinking skills (METT-TC, BOS, and TLAC). While there are some overlaps in principles among these three organizers, there are also some principles that appear to be unique. Our goal was to include a broad range of higher order principles to give the trainee ample feedback to develop their own cognitive organizers.

For example, a single scenario may illustrate:

- mistake made in the use of terrain (METT-TC) or (TLAC)
- mistakes made in the use of infantry, armor, and aviation (BOS)
- not following the principle, "Model a thinking enemy" (TLAC)

The tutor has a small number of critiques associated with each case. Each critique, in turn, is associated with one or more units on the battlefield and a unique mistake category. Scoring is based on the mistakes/critique statements correctly identified.

<u>Tactical Studies</u>

In CATS-1 the studies are just set up and the trainee can play them out, receiving pre-stored hints on request. However, CATS-1 does not have the required level of integration with the simulation to allow it to monitor the trainee's actions. Furthermore, it also needs explicit rules to help it assess those actions. Both will be supplied in CATS-2; there will be both tutor monitoring of user performance and tutor-generated hints and feedback.

Phase II Technical Objectives and Approach

The product goal is to develop an intelligent tutoring system to improve battlefield command reasoning through deliberate practice. The primary research goal is to model and evaluate the cognitive strategies of learners. The achieve these top level goals, the Phase II effort will accomplish the following technical objectives:

- Develop learner models representing their cognitive strategies
- Model key aspects of the domain
- Develop training activities that facilitate the development of adaptive thinking skills
- Provide meaningful feedback
- Give appropriate responses to trainee- initiated questions about the domain
- Tailor instructional strategy to each individual soldier (based on their strengths and weaknesses
- Have persistence (maintain a student model and keeps track of each soldier's progress to enable stopping and resuming sessions at any time).

Phase II Work Plan

This Phase II plan addresses both the research and product orientation of ARI's solicitation.

The product focus is on developing an intelligent tutoring system to teach battlefield command reasoning through deliberate practice in key scenarios and tactical simulations. The CATS Tutor selects training activities and provides explanatory feedback based on an evolving model of the individual learner. It draws on an ontology to provide a deep knowledge base and reasoning capabilities to infer a learner's strengths and weaknesses. The primary focus is on inferring the learner's cognitive strategies by their behavior rather than through primary reliance on questioning and thinking aloud protocols. The tutor's instructional strategies build on techniques designed to teach expert cognitive strategies, that is, how to think rather than what to think.

This brings us to one of the key research components of the project -- how to model and evaluate the cognitive strategies of learners. Most intelligent tutoring systems have focused on evaluating whether learners have the requisite concepts, rules, procedural skills, or heuristics. Few have addressed cognitive strategies as the primary knowledge to be modeled and evaluated. A new approach to user modeling will be investigated where different cognitive strategies are simulated and the results (inferences drawn and actions taken) are partially matched to user actions in specific task situations.

The CATS Tutor is designed in an open, flexible framework so that its training strategies can be applied to a wide range of products and DoD needs. The tutor is not tightly integrated to the training content or the simulation. The tutor has 'hooks' to the tactical simulation, with information being passed (communicated) between the simulation and the tutor agent. With this type of communication design, other simulations can be more easily substituted for tutorial interactions supporting other topic domains.

First, we will discuss some of the research concepts about adaptive thinking skills, as these are the primary drivers of the design and challenges of the CATS Tutor.

Adaptive Thinking Skills

The cognitive strategies for this domain are adaptive thinking skills in tactical situations. Adaptivity is required under the pressures of time, risk, and complexity (many interacting system elements). These thinking skills have been elucidated in various ways: 8 thinking habits in the TLAC course (e.g., model a thinking enemy, consider effects of terrain), METT-TC (Mission, Equipment, Terrain, Troops, Time, and Civilians) and BOS (Battlefield Operating Systems, such as armor, artillery, engineering, and air support). The CATS Tutor is *not* specifically teaching the 8 principles of the Think-Like-a-Commander course, or any other skill organizer. It draws on a rich domain representation to help learners develop their own expert cognitive strategies.

The training focus for adaptive thinking skills is on "*how* to think" as opposed to "*what* to think". A key approach is through the use of deliberate practice. The practice is targeted to improve some aspect of the required thinking skills through repeated practice in varying scenarios. With enough deliberate practice, the thinking skills will become automatic. ARI applied the principles of deliberate practice to the domain of the tactical thinking skills for a brigade level commander's course at CGSC, called "Think Like a Commander". Some of the origins and principles behind this work are based on Soviet chess training techniques.

Mark Dvoretsky, in his book Secrets of Chess Training advocated the use of studies for training. These chess studies may be artificial puzzles, rather than actual game positions. They are deliberately designed ("composed") to exercise particular chess (thinking) skills. Dvoretsky states that "... the game of chess is a lot richer and more difficult than a study, although there are ideas that may be expressed in studies more fully and effectively than in actual games" (Dvoretsky, p. 169). Dvoretsky targets studies to the trainees to improve the areas they are relatively weak on. An example of this kind of exercise selection: "It is quite common for a player, even an excellent tactician, only to see his own possibilities and to underestimate his opponent's counterplay. For this player, we can pick out special studies whose center of gravity lies not in finding one's own combinations, but in taking account of some unexpected resources of the opponent's at the proper time" (Dvoretsky, p. 148).

A key similarity between both domains, chess and tactical command, is decision-making under time pressure. For example, from Kotov, "As a rule, however, a grandmaster will not start checking a second time all the variations which he has already examined. This is an unforgivable waste of time..., and moreover shows a lack of confidence in one's analysis." Another example of the importance of time in a cognitive strategy for chess: "We repeat the rule: candidate moves must be established straight away and they must be clearly enumerated. This task cannot be split into parts by examining one move fully and then looking for the next one. This brings disorganization to your thinking. Without knowing how many candidate moves there are, you could devote too much time to one of them and when finish examining its ramifications find that you just don't have enough time for the rest."

Of course, there are some important differences between chess and tactical thinking. The enemy's position and strength is completely known in chess, but not known in most tactical situations. The discrete nature of chess moves and limitation on legal moves allows a deep analysis of variations and game-tree kind of analysis that is much less applicable to tactical analysis. Tactical warfare is continuous (real-time, no turn-taking) and there are no limits on how weapons can be used other than those imposed by physics and rules of engagement. The most important difference is that the fear of death or injury and the responsibility for other lives is missing from any game or simulation, but a real presence in combat situations.

The CATS Tutor is being designed as a generic tutor for critical thinking skills. It provides opportunities to practice and refine these skills, with tailored guidance. From our various readings and research, we suggest that these following principles need to be included in an adaptive thinking skills tutor.

<u>Focus on Task Performance, Not Talking About the Task</u>

We believe that attempting to emulate the exact human tutorial interactions that occur in the TLAC portion of the CGSC TCDC class is not the most effective design approach for the intelligent tutoring system. Effective human tutoring requires a skillful mix of nonverbal perception, tact, knowledge, trust, and timing that (we believe) will not appear in computers for some decades to come. If it were possible to develop a system, it would more likely teach trainees how to *talk* about tactical situations, but not necessarily how to perform in tactical situations.

Therefore we believe that the best use of an intelligent tutoring system is with a highly realistic and interactive simulation. Tactical studies (analogous to problem studies in chess) could be used to place the trainee in a situation requiring their assessment and response. The commander's tactical decision could be played in the simulation but need not be played out in its entirety. The ITS would tailor its guidance and feedback based on an evaluation of a commander's decisions.

Deliberate and Repeated Practice

Deliberate practice requires exercises and studies that promote automaticity (e.g., perceptual chunking). It is not enough for trainees to be given explicit feedback and explanation about tactical decision errors. They must also have repeated practice at solving the same and similar problems. The training system should provide many hours of practice with varied exercises and tactical studies.

"Experience and the constant analysis of the most varied positions builds up a store of knowledge in a player's mind enabling him often at a glance to assess this or that position. It is this erudition that helps a grandmaster to choose the right move without deep thought." [Kotov, P.80]

Time Management During Decision-Making

Another key characteristic of adaptive thinking in tactical situations is that it is time-constrained. The trainee must modify their thinking strategies based on the available time. Rarely, can a commander make a leisurely exhaustive study of all possibilities of a situation. The system must include time constraints to impose a realistic factor in the evaluation and training.

Training Adapted to the Individual

A 2-sigma level of improvement in instructional effectiveness is found in one-on-one human tutorial interactions (Bloom, 1984). Tutorial interactions tend to include more active participation by the learner in solving problems, and guidance and feedback tailored to the individual. We believe the best approach for this project is to design a training system that includes active participation by the learner (via simulation-based training studies) and tailored tutoring by the ITS.

Design of the CATS Tutor

The CATS Tutor will focus on adaptive thinking skills for brigade and battalion level commanders. It will consist of an intelligent tutoring system that controls a tactical simulation. The CATS Tutor will ask questions, give advice, and track the commander's decisions and actions within the various training activities (exercises, critiques, tactical studies).

The tutor will include a library of cases (exercises/tactical situations/scenarios) developed with the help of subject matter experts. Training activities will cover a broad spectrum of critical thinking skills. There will be sufficient cases to demonstrate customized case selection and to provide coverage for a subset of target user categories (e.g., different levels of experience, difference branches).

The Phase II implementation will be a major extension to the Phase I prototype. The key extensions include:

- richer learner model
- more adaptive instructional strategy
- meta-level controller
- an ontology (knowledge base)
- responds to learner queries about the domain
- larger case library, tactical simulation

- tutor-simulation integration

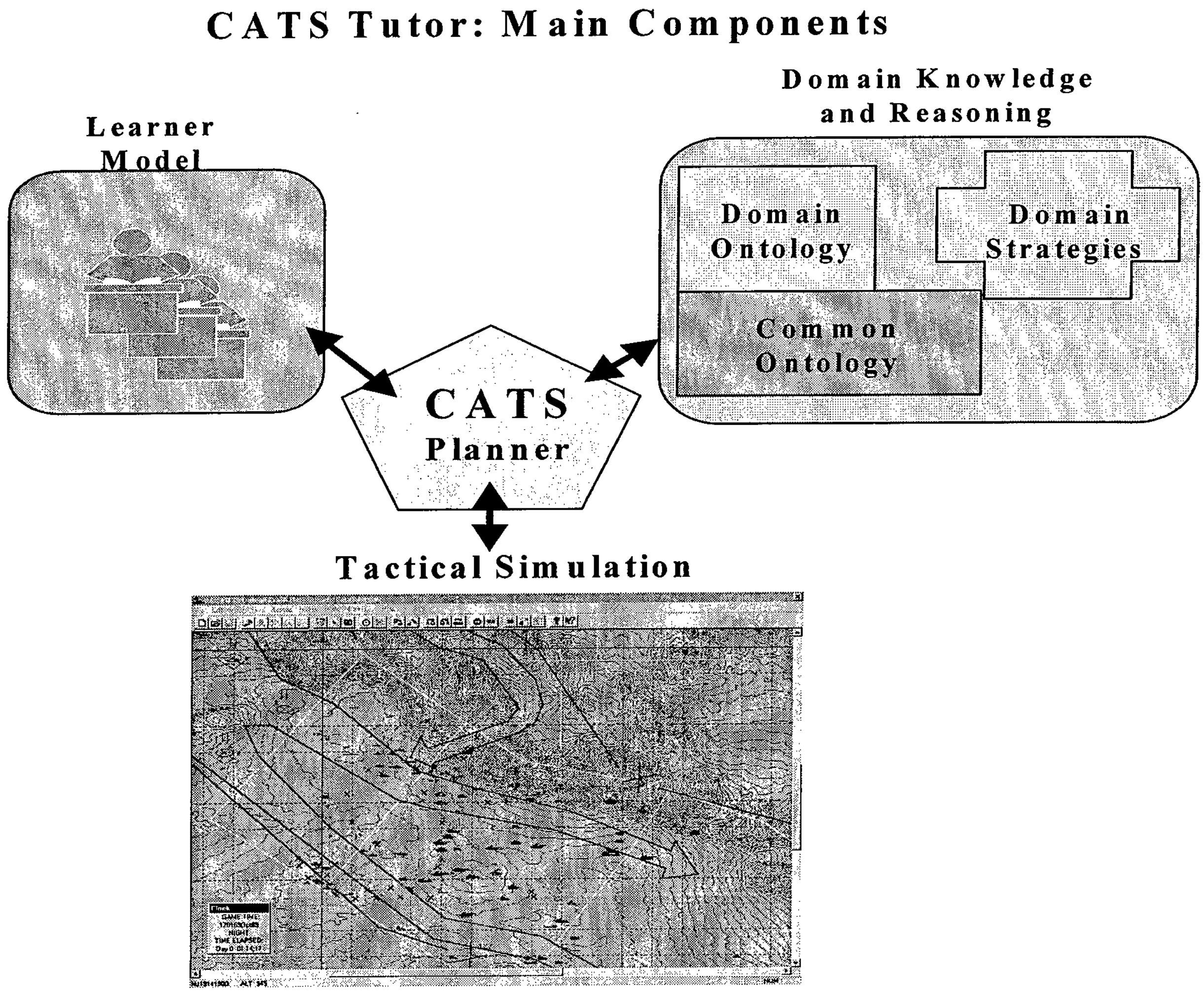

Figure 2. CATS Tutor: main components

<u>Richer Learner Model</u>

We used a simple scoring method for the 'learner model' in the Phase I prototype. Two kinds of learning will be modeled in the user model for our Phase II implementation: cognitive strategy and degree of automaticity.

Determining a learner's cognitive strategy is a meta-level matching process. By this we mean that the evidence is based on student actions and conclusions. We need to model a time-limited reasoning process since real decision-making is made under time pressure. In this particular application, we would model time-pressured decision-making as the selective application of rules and facts according to a meta-level reasoning strategy where resources are limited. In other words, we either limit the number of rule applications, the amount of working memory, and/or the time available for computation in such a way that not all rules can be applied.

Suppose that altogether there are about 1200 rules that would cover all the various factors (e.g., terrain, unit interactions) for any battlefield situation. Suppose that 200 rules are applicable for a specific battlefield situation. The rules could make all sorts of deductions (e.g., about whether a path is covered, when a shell will splash, etc). A single rule can be potentially instantiated (applied) to an infinite number of paths in a single battlefield situation.

Once we elaborate all the rules that might be applicable in our selected set of scenarios, we model different cognitive strategies using these rules. Some of the strategies will be more and some less effective. One strategy might be to always apply the terrain rules, then the path rules on the single most likely COA, then rules having to do with fire missions, etc., until time (to decide) runs out. Another strategy might apply quick-and-dirty rules from each category, form initial hypotheses about what to do and then apply rules from each category until time runs out.

Each strategy would make different decisions for a given situation. The tutor would attempt to identify the student's strategy by matching it to one of these strategy models.

Of course this task is not so easy. The subject matter experts can help us by telling us how different officers with different backgrounds are most likely to try to solve a problem. We can attempt to model these cognitive strategies and then see how successful they are in determining user cognitive strategies. An interesting research application of the CATS Tutor would be to collect performance data from commanders with differing levels of expertise, and use that data to form the differing cognitive strategy models.

The determination of the learner's cognitive strategy will be based largely on inference from their decisions and actions. Where there is ambiguity, the tutor could ask them clarification questions, but we would like to keep these questions to a minimum. Instead, we prefer to rely on evidence gathered over multiple decision points and cases.

Our user model need not be perfect, just useful enough to assist us in guiding the user to the next training activity and to provide relative measures of strengths. For example, it may appear that the commander is most likely using strategy A, but may be using strategy B. If strategy A and strategy B both indicate deficits in "Modeling a thinking enemy", then the tutor can recommend exercises in that venue without further distinction between the two strategies.

We also want to include a degree of automaticity in the learner model. Time to make a decision is one measure of automaticity. The tutor could compare the trainee's time performance to a standard, resulting in a multiplier (e.g., 0.5 for slower performance, 2 for performance twice as fast, etc.). For example, suppose the trainee scores 0.5 on visualizing the battlefield but 2.0 on considering terrain (i.e., they take twice as long as the standard in solving battlefield visualization exercises but only half as long in solving terrain effect exercises.) The tutor can adjust the resources allocated in a particular model of a cognitive strategy by these multipliers.

Obviously this approach needs more refinement and testing. We need to interview subject matter experts to see to what extent a small number of common cognitive strategies can be captured and whether we can trace a path from novice to expert to track the trainee's improvements. We need to weigh the value of trainee self-assessment (e.g., "Did you consider what the enemy can infer from its DRT team?", "How much time do you think you spent (1) evaluating terrain, (2) visualizing the battlefield, ... ") and see whether we need a normative measure of scores, obtained of many trainees, rather than just a relative measure. The latter is more difficult but is appealing in allowing a trainee to compare their skills to others.

More Adaptive Instructional Strategy

Most adaptive computer-based training systems control interactions via a pre-scripted storyboard. Intelligent tutoring systems are more dynamic in their control and can include plans developed and revised on the fly. Even though deliberate practice provides us with an overall instructional strategy, an ITS still needs to make many more detailed decisions: for example, what degree of user control to provide, how to handle student questions and requests, how to manage the tutor's time, etc. It can do this reactively, without planning, but more coherent behavior occurs if planning is incorporated into the control mechanism.

The approach we plan to use for Phase II is to represent an instructional strategy as a set of instructional plans. Each plan is a set of steps. Each step represents constraints on desired pedagogical actions, usually expressing a preferred category of instructional actions. Each step also indicates a goal, which when achieved, terminates the step. A plan also has an overall goal. If the goal is achieved during the execution of the plan then the remaining plan steps need not be executed. If the plan fails to achieve the goal then the plan failure is noted and a separate kind of instructional plan (a repair plan) handles the failure. The failure can be handled in many different ways according to the situation (time remaining, user model, desired instructional objectives, etc.). One way is to attempt to achieve the goal of the plan by other means. Another way, used after repeated failures, would be to try to mark as failed a higher-level instructional goal and then to proceed trying to handle that goal. Finally, a goal can be deferred or human assistance requested.

This type of instructional planner better enables:

Instructional strategies based on the learner model. The user model will be used to plan a sequence of tactical studies, practice exercises, and study of expert examples. Mission play might also be recommended if these resources are available. In general the plan generated will depend both on the user model and the resources available (or allowed).

Variable levels of prompting and feedback. CATS-2 will also have hints available for all exercises. Initially these will be pre-stored, and later these will be supplemented with hints generated from the CATS-2 knowledge base. More specific feedback will also be available as to what tactical guidelines appear to be violated.

The level of hinting and feedback can be selectable (by instructors or trainees, if desired). Users can specify that hints are only given by request (the default), according to a predefined hinting strategy (e.g., give a hint after *n* seconds then every *m* seconds thereafter but only give the final hint upon request), or are not available at all. Similarly, feedback can be restricted to only what the user has done correctly (as in CATS-1) or can be extended to include mistakes perceived by the tutor. Finally, the trainee may be allowed to nullify (have the tutor retract) perceived errors if the tutor has misperceived them.

Tolerance to different degrees of learner control. The tutor can entirely guide the training, strongly suggest certain paths, or only provide guidance when asked.

Meta-Level Controller

An instructional plan is not much use without an interpreter. CATS-2 will use an agenda-based interpreter derived from prior work on dynamic instructional planners and black board-based planners. The basic idea is to provide an agenda of actions that can be performed and to repeatedly choose the best action given the current plan step designating the most desirable actions to perform. This allows the tutor to do the best (actions) with what it has (the resources available right now).

An Ontology (Knowledge Base)

Most training systems know nothing about the domain they are teaching. They simply retrieve material and multimedia. A key feature of intelligent tutoring systems is the capability to reason about the domain through the use of a knowledge base. These knowledge bases represent concepts and rules that use these concepts. A drawback to the typical approach for knowledge-based systems is that they are usually hand-crafted for the specific domain and are not easily extended to handle new concepts or domains. They also tend to provide a shallow layer of knowledge without the underlying concepts and rationale to back up the rules. Thus, they tend to be brittle when common-sense reasoning is required, or reasoning outside of the intended domain.

An emerging approach is the have an ontology as the foundation for the knowledge base. An ontology represents basic (common) concepts, such as gravity and time. Ontologies promote extensibility by providing a common starting point on which to build domain-specific knowledge bases. This approach allows more focused effort on developing the domain-specific aspects rather than spending time representing common concepts. The ontology also provides for deeper reasoning by providing a more richly modeled set of underlying concepts for domain-specific concepts.

We propose building the knowledge in the CATS Tutor using an ontology. We will use an upper level ontology for general terms, a middle-level ontology for Army specific terms, and a lower level ontology for domain specific terms. An upper level ontology effort, the Standard Upper Ontology (SUO), is currently being led by Teknowledge, with input from the knowledge representation community. It is being vetted by the IEEE standards, and will be in the public domain. Teknowledge is also developing an Army Reference Ontology (ARO) in a separate Phase II SBIR project. As part of this SBIR effort, we will add the middle and lower level knowledge to support the scenarios in the CATS Tutor.

The primary benefit that an ontology and knowledge-based approach brings, is that the tutor will be able to reason about the domain. This enables it to "observe" what the trainee is doing and "understand" the implications of decisions and actions instead of directly elicit this information from the trainee, such as through questioning. The "observation" aspect is accomplished via transfer of information and data from the simulation to the tutor. The "understanding" aspect is through application of the tutor's domain knowledge to the observations.

The simulation program collects and passes activity information, such as the number of friendly / enemy vehicles moving through a particular area (e.g., a key pass or other critical terrain). The simulation also monitors various "design rules" (e.g., is the infantry outrunning the artillery? is breaching being performed in the best possible manner?). It also can record actions the trainee never takes (never uses FASCAM ammunition, never uses helicopters for scouting, never uses engineering assets). All these types of information are 'observation' clues for the tutor.

Domain knowledge is needed to interpret the results for two reasons. First, the data from the simulation is too low level and needs to be abstracted. For example, the simulation may tell the tutor that a T-72 is rolling up on a BFV in defilade and is approximately 2000 meters away from the BFV and in its line of sight. The tutor may have a rule that an a TOW missile can destroy a tank within a range of 3750 meters if the vehicle carrying it is stationary and can see the target. In this case the rule applies, but the tutor must first infer that a BFV normally carries TOW missiles, the missile supply is not exhausted, the BFV can see the T-72, the T-72 is within range, and the BFV can fire (i.e., is not operating under a "Hold Fire" order at the time). The simple reasoning that the T-72 is a tank (and thus the rule applies given the other conditions) is an example of the abstraction required.

The second reason domain knowledge is needed to interpret the results is that many of the rules are sensitive to the context of the mission and situation. We need a knowledge of various aspects of the situation and mission to determine if an action or lack of an action indicates an error on the part of commander. For example, suppose the commander never uses smoke against targets over the course of three missions, even though smoke is available both from M109A6 ammunition and from mortars. If the smoke could have been used effectively to blind enemy vehicles in a situation where that would contribute to the mission, and the commander did not achieve the results by some other means, then the tutor will flag not using smoke as a possible error of the commander. That may in turn signal a lack of understanding of the enemy (in the METT-TC framework), lack of understanding the interaction between friendly and enemy units (TLAC), or a lack of understanding of artillery systems (BOS). But if the enemy vehicles had thermal sights then none of these conclusions apply, as they would have been able to see through the smoke with the sights. More generally, to check a simple rule (e.g., Does the commander let his forward line of troops outrun his artillery?) may require inferring where the FLOT is, which units are artillery, and what their ranges are.

In addition, the tutor's ability to reason about the domain, based on the ontology and knowledge base, enables it to dynamically form responses and explanations to a trainee's query. More on this aspect in the next section.

Responds to Learner Queries About the Domain

In CATS-2, the trainee can ask questions about tactical studies and missions. Our focus for this capability will be on the tutor's reasoning about the domain and formulating relevant responses from the knowledge base. We do not expect to handle free text queries. Instead, menus will be used to construct queries or a context-free grammar will be used to parse a limited vocabulary.

Because we are building a knowledge-based tutoring system, the tutor will be able to dynamically reason about the domain, not just retrieve closely matched stored facts. It will not be an approach like most help systems that use quasi-NL queries on a set of FAQs or hints. A knowledge-based approach has a much greater likelihood of actually answering the posed question.

For example, some of the questions the tutor should be able to handle include:

> "Can the Enemy artillery hit us?
> Domain Model Rule: (Friendly position vs. [Enemy position, artillery range, 2/3 rule])
>
> "How soon can unit X get to position Y?"
> Domain Model Rule: (Unit X position vs. [position Y, geographical terrain, unit X vehicle speed and mobility])
>
> "Can we get to position X before daylight?"
> Domain Model Rule: (Our position vs. [position X, time before daylight, our vehicle speed and mobility])
>
> "How much should I lead the enemy column by, given their current speed and the time to splash for DPICM rounds?"

Domain Model Rule: (Calculating time for fire mission & time to splash, calculating how far enemy column will move given current vehicle speed and terrain being traversed)

"Will smoke be effective against these enemy units?"

Domain Model Rule: (Weather, speed of wind, dispersal of smoke)
Domain Model Rule: (Thermal imaging units, knowledge of enemy weapons)
Domain Model Rule: (Smoke generators available, artillery available, size of artillery shells, smoke produced by mortar vs. 155mm vs. smoke generators)

Larger Case Library, Tactical Simulation

After interviewing SMEs we will develop/obtain a wider selection of cases that are sufficiently challenging and appropriate for officers learning to be brigade or battalion-level commanders.

In Phase II we can continue using Brigade Combat Trainer (BCT) or use its successor, Armored Task Force (ATF). BCT models tactical engagements at the battalion and brigade level. ATF has a simplified interface, better graphics, and support for engagements at the company and platoon level. ATF will have additional enhancements to increase the realism of tactical situations, such as modeling different kinds of night vision equipment, both turret and vehicle facing, different kinds of suppressive effects of artillery, use of laser designators from Kiowa helicopters for PGM (Copperhead) ammunition, etc.

Tutor-Simulation Integration

Most simulations are built as stand-alone tools and are not easily incorporated into other software programs. For this project, we plan to leverage and incorporate an existing tactical simulation into our intelligent tutoring system. The objectives are to:

- use the simulation as part of its graphical user interface
- control the simulation for tutorial purposes (e.g., disabling certain commands, stopping the clock and asking questions about the user's reasoning, tracking what the user does)
- maintain enough separation from the simulation so that improved versions or alternate simulations can be substituted for the original, yet provide the appearance of seamless integration.

The CATS-1 proof-of-concept simply uses Quickeys, a program utility intended to allow adding hot-keys to programs and macros that can operate across several programs. Quickeys is sufficient to show the concept but a greater degree of integration is required for Phase II. For example, the CATS-2 tutor will need access to unit information that is currently presented only in pop-up boxes in BCT.

In Phase II we will investigate the use of *software connectors*, a software means of safely integrating programs that were initially intended to run standalone (Balzer and Goldman, 1999). We will also work with ProSimCo to allow a greater degree of tutor access to their simulations through software connectors.

Software connectors are system-level traps to DLL calls. A DLL (dynamic link library) is a kind of shared software component that is a building block of operating systems and programs. Operating system DLLs provide window management facilities (e.g., opening and closing windows, rearranging their order). Program-specific DLLs provide a modular decomposition of the program to facilitate upgrades and sharing of code among different parts of the program. A group within Teknowledge has been working in the area of software connectors, and we can build on this experience to provide a more principled and more efficient means of connecting simulations to tutors.

Related Work

Relevant Prior Work By Dr. William Murray (P.I. for this SBIR project)

Dr. Murray's work has focused in the areas of simulation-based tutors, blackboard-based intelligent tutoring systems, dynamic instructional planning, and user/student models. Dynamic instructional planning is an approach to controlling an intelligent tutoring system whereby plans are selected or generated and then executed. These plans can be revised as opportunities arise and changes occur in the resources and the user model. The planning is *dynamic* as it occurs during instruction, in real-time. Traditional computer-based training follows a pre-stored plan and other learning environments, such as collaborative teaching approaches or standalone simulations, have no tutorial plan at all. His work in user modeling is on investigating alternative approaches to representing a tutor's inherent uncertainty in interpreting incomplete and changing student performance data, along with possibly conflicting assessment data from the student and instructor.

A few examples of Dr. Murray's work includes:

Simulation Tutors

The Mark 45 Lower Hoist tutor used a STEAMER-like simulation of an electrical-hydraulic-mechanical assembly for loading shells into the Mark-45. Using the simulation, the tutor taught the layout and configuration of component parts of the assembly, how they operate normally, and how they operate under faulted conditions. Trainees could then practice troubleshooting unknown faults and the tutor could evaluate their reasoning (e.g., is the student's hypothesized fault consistent with the measurements taken so far?).

Instructional Planners, a Language and Interpreter For Instructional Plans

A blackboard-based instructional planner was used to control the Mark-45 Lower Hoist tutor, and has been used in other projects and applications. A blackboard architecture is a software architecture for problem solving, real-time control, and component integration. Language frameworks were used to express both domain-level actions that can apply across similar domains and to express meta-level reasoning in the blackboard system.

Part of the Blackboard Instructional Planner work was the development of BB1 language frameworks (languages for specifying operators and solutions to a category of tasks) for instructional tutoring applications, specifically with the use of simulations. Subsequent work with the CLOS (Common Lisp Object System) and Common Lisp version of BB1 led to a deeper understanding of the implementation of BB1 and language frameworks, and a knowledge of how to simplify the general blackboard system to an easier-to-use and easier-to-author instructional-plan interpreter for an agenda-based control mechanism. This mechanism is the one that will be used in CATS-2. The relevance of this work in blackboards and language frameworks is that it provides a model and shows prior experience in representing instructional strategies as meta-level reasoning plans, and actions as part of an ontology.

User/Student Modeling

An endorsement-based approach to user modeling was implemented in a production system with a TMS (truth-maintenance system). A TMS is a bookkeeping mechanism to ensure consistency between facts, rules, and inferences that result. It was used to model default beliefs that could later be changed when new data arrived and to model data that was to be withdrawn when it had "aged" beyond a certain point and was no longer relevant. An IJCAI paper describes this work (Murray, 1991).

Other projects have utilized Bayesian analysis techniques, which have become increasingly popular as a user modeling approach within ITS. Simple algorithms were developed for propagating updates within tree representations and techniques representing different kinds of performance data (e.g., tasks, multiple-choice questions, self-assessments, etc.) (Murray, 1991).

Relevant Work by Others

Intelligent Tutoring Systems That Build On Simulations

SOPHIE (Xerox Palo Alto Research Center). SOPHIE was perhaps the earliest intelligent tutoring system to appropriate a separate standalone simulation, in this case to provide part of its domain expertise. SOPHIE used the SPICE circuit simulator to evaluate student hypotheses about circuit faults to see if they were consistent with earlier measurements. It could also evaluate whether a measurement was redundant using the circuit simulator. In addition to evaluating student hypotheses, the circuit simulator was used to simulate part replacements and the effect that resulted from them. (Sleeman and Brown, 1982).

STEAMER (NPRDC and BBN). The SPICE circuit simulator was a numerical simulation of analog circuits. STEAMER provided the first well-known interactive graphical simulation of a complex system (a naval steam-plant used in propelling a ship) (Hollan, Hutchins, and Weitzman, 1984). By allowing the student's to interact with a device simulation they could simulate opening and closing valves, introducing leaks, etc. The simulation was comprised of models of components and their connections. These models were also designed to generate explanations of behavior. Although we have not proposed it here, we could similarly use the CATS-II knowledge base to generate explanations of tactical behavior from the interactions of tactical units.

RBT (University of Massachusetts). One of the most commercially successful examples of the use of a simulation is the RBT (Recovery Boiler tutor) simulation (Wolf, et.al., 1986). The simulation was developed by the paper pulp mill industry to mathematically simulate Kraft recovery boilers using differential equations. The Kraft recovery boilers are large multi-story boilers for recovering paper pulp from recycled paper products. Unfortunately, under certain critical conditions the boilers can explode. The simulation by itself lacks tutorial explanations and guidance, but could be used for training by human instructors.

Dr. Woolf added an intelligent tutoring system that communicated with and used the existing plant simulation. It was used to teach critical scenarios that can lead to boiler explosion and how to deal with these crises. The combination of the tutor and simulation has been so successful that they are now required for all new paper pulp plants with Kraft recovery boilers.

The connection between RBT and the current work is that RBT shows how an existing subject-matter expert developed simulation can be incorporated and leveraged by an intelligent tutoring system. Furthermore, RBT can be used to teach time-critical operations and thinking. If the operator performs the right action at the wrong time (e.g., too early or too late) it is ineffective or counterproductive. The importance of time and synchronizing operations is a recurring element in tactical thinking, too.

Similarly to RBT, we propose building an intelligent tutoring system platform that is plug-compatible to SME-developed tactical simulations.

<u>SHERLOCK (LRDC).</u> SHERLOCK is another ITS-simulation success story, this time incorporating its own built-in simulation. SHERLOCK simulates avionics test equipment in normal and faulted modes of operation, and troubleshooting probes that can be performed on the equipment. Trainee technicians who used SHERLOCK for 40 -80 hours performed as well as technicians who had been on the job for several years (Lesgold, et.al., 1992).

<u>Strap-on tutors (CMU).</u> As can be seen simulations have commonly been either attached to or part of ITS systems. Ritter and Koedinger (1997) proposed having a separate tutor component that can "strap onto" an existing application. CATS-II adopts this idea, but in a more ambitious fashion. Ritter's work used applications such as Excel that are designed for external control; CATS-II assumes the simulation was not so designed.

<u>Intelligent Tutoring Systems with Natural Language Approaches</u>

<u>SCHOLAR (Carbonell, 1970).</u> SCHOLAR was built by Carbonell to teach geography with a Socratic question and answer style. For example, it would propose counterexamples to a student's overgeneralization. WHY (Collins, 1975) also provided a natural language interface and discussion about how rainfall occurs. SOPHIE itself used a simple (semantic grammar) natural language interface.

<u>AUTO-TUTOR (Univ of Memphis).</u> AutoTutor simulates the discourse patterns and pedagogical strategies of a typical human tutor (Graesser, et.al., 1999). The tutor attempts to identify the student input via a 'speech act classification' and dialog is controlled by a curriculum script. Student responses to tutor-posed questions are evaluated for accuracy using latent semantic analysis (LSA). LSA is a statistical approach to comparing text fragments (e.g., a query to a text document, or one document to another). LSA may be able to discriminate whether two text fragments are similar or dissimilar, but is unable to analyze why there is a difference. This is because LSA is a heuristic approach based on abstracted characteristics (e.g., infrequently used words or word combinations) with no real understanding. Consequently it can misinterpret matches, leading to false negatives or false positives. Rather like a grammar checker, it can be a useful tool (e.g., in information retrieval or search engine applications) but should not be interpreted as exhibiting any true kind of understanding.

<u>MENTOR/TUTOR, a Dialog Agent System (Teknowledge and Amber Consortium).</u> The goal of this program is to provide authorable, dialog-enabled intelligent agents for tutoring and performance support systems. Users interact with intelligent agents who carry out strategies and goals and can engage in mixed-initiative dialog via a natural language understanding and generation system. Non-programmers can author new domains and scenarios and create new conversational agents. The dialog system is authorable by non-computational linguists. This

project attempted a full-up NLU and generation dialog capability. The dialog system (developed by Amber) had a lexicon, parser, semantics analyzer, knowledge base and ontology, and dialog management. While the dialog system shows a proof-of-concept, it is less than robust. Main shortfalls were some parser limitations, and a shallow ontology and reasoning capability.

A problem with most natural language interfaces is that they provide uneven coverage of sentences or queries, leading to frustration among users. In many domains it is also more natural to interact with graphical diagrams (e.g., electronics, geometry, physics) than to try to explain desired actions.

Soviet Chess-Training: Training Techniques To Teach Expert Cognitive Strategies

Soviet chess training techniques are relevant as they are the basis for Dr. Lussier's development of ATTM (the Adaptive Thinking Training Methodology). Two influential books revealing these training techniques are Kotov's Think like a Grandmaster (Kotov 71) and Dvoretsky's Secrets of Chess Training (Dvoretsky 91).

Kotov provides examples of how a grandmaster thinks. He also talks about specific strategies that they apply in their thinking process. For example, they will not check variations twice and they will enumerate candidate moves before exploring variations in depth. He discusses reasons for blunders (e.g., overconfidence) and the importance of time and avoiding time trouble in making decisions.

There are many parallels to the training proposed in CATS-II. In CATS-II there will be tactical cases played out by experts to model expert tactical thinking for novices. Time will also be a critical factor that trainees will have to learn to deal with and they have to learn to avoid time trouble. "Time trouble" arises in battlefield situations when there is not enough time to perform the full MDMP (military decision making process) and a commander must rely on their intuitive judgments of friendly and enemy COAs to decide which to evaluate in more detail.

Dvoretsky's book introduces the training technique of having players play out studies. The studies are chess positions deliberately composed to train one part or another of a chess player's reasoning capabilities (e.g., analysis of variations, intuitive analysis of positions, etc.). Ideally they are played out against a human trainer, although they need not be if one is not available.

The tactical studies of CATS-II are similar to the chess studies Dvoretsky discusses. The tactical studies, authored by SMEs, will be designed to teach different thinking habits in the TLAC principles. Dvoretsky says that "playing these studies out is a very good form of training" [p. 161] and CATS-II allows this form of practice by allowing the trainee to play out studies against the programmed opponent of the tactical simulation.

Dvoretsky also emphasizes that there are some things that cannot be taught solely by studies, and for these more subtle aspects actual play is required: "... the game of chess is a lot richer and more difficult than a study." Similarly, CATS-II will let the trainee practice entire mission scenarios if desired. In tactical decision making such practice is more important than in sports such as tennis or golf where games are easily played out. In contrast, in tactical decision making, commanders infrequently (e.g., once a year on a training rotation) get to practice their art in realistic conditions, so practice in targeted exercises needs to be supplemented with practice in more complete situations (e.g., missions) to provide sufficient practice both in individual skills, in blending the skills together, and in strategic thinking where the resources in one tactical engagement must be balanced against their need in future engagements in the same mission.

Conclusion

This Phase I effort contributed some innovative concepts for an intelligent tutoring system that focuses on cognitive strategies. The CATS Tutor requires a new kind of learner model that determines meta-level thinking (cognitive strategies). It needs to determine what do commanders spend time thinking about? What do they forget to attend to? Is too much time spent in areas that are trivial or which can be automaticized?

Most other user models assume a single cognitive strategy, or ignore the problem altogether, and attribute all mistakes to lack of knowledge, incorporation of misconceptions, or slips. In contrast, one novel aspect of the CATS user model is that we will assume that all mistakes derive from inappropriate cognitive strategies or lack of automaticity leading to time trouble. The ARI Newsletter (Lussier, 2001) points out that "Field Grade Army officers, have a great deal of knowledge in the military domain but do not always use the knowledge effectively".

In addition, we propose to develop learner models of different types and levels of cognitive strategies (e.g., intermediate with artillery expertise, novice with good terrain skills).
If we were to key the initial learner model to the user's background (e.g., a captain from Artillery would receive an intermediate with strong artillery background), then this approach bears some similarity to cognitive stereotypes (Rich, 1979). Our approach differs in the modeling of cognitive strategies as opposed to sets of preferences or categories of background knowledge. We expect the differences between the strategies used by learners will show up most in situations where time-pressure is applied. [We suspect that the data collected with the CATS Tutor will be most interesting.]

Ontologies are just beginning to be used in intelligent tutoring systems and are another active area of research at this time. CATS will further this research by developing a deep knowledge base for each of its domains, built on a middle-level ontology of concepts and axioms describing Army, and tactical knowledge, and leveraging an upper-level ontology (the Standard Upper Ontology) that acts as a general theory of the world.

The challenging questions here are to show how such upper, middle, and lower-level ontologies should be merged, demonstrate the claim that much deeper reasoning does occur, and that extensibility to similar domains is greatly simplified. Although it might seem intuitively obvious that more knowledge would be better, practical problems in the control of reasoning must be solved to use the knowledge effectively. In other words, CATS must have a meta-level strategy for controlling its own resources, just as the trainees must learn a meta-level strategy suited for their domain.

Representing meta-level strategies is a challenge for the CATS Tutor in two respects: 1) in controlling its own application of knowledge and computational resources (pedagogical domain), and 2) in capturing alternative strategies used by novices and experts (tactical domain). We plan to use a blackboard instructional planner, but have not yet seen it applied in this unique ITS domain.

Another area of innovation is to explore the best approach for integrating subject-matter expert developed simulations to qualitative reasoning systems such as intelligent tutoring systems. A typical SME-developed tactical simulation will mix a graphical depiction with a stochastic simulation of hit and kill probabilities for different unit and vehicle types, modified by weather, vehicle facing, visibility, etc.

Software connectors are one approach to integrating such systems originally developed for standalone use. These are traps to calls in the software that allow tutorial processes to intervene to log user behavior, provide hints, disable a set of commands, etc. The Phase I approach to CATS Tutor used simple hot-key and macro definition software to provide a simple integration capability. In Phase II we will use more efficient approaches developed and tested for security applications (e.g., secure email attachments) and for software monitoring in quality of service applications.

CATS Tutor as a Research Tool

The CATS Tutor is intended to be a training tool but could also be used as research tool. It is already designed to present different conditions and collect data. With some extensions and modifications, it could also be used to conduct research to determine and test:

- the adaptive thinking skills that characterize experts
- the factors that play key roles in the development of adaptive thinking skills
- the validity of models representing the cognitive strategies of experts and varying levels of non-experts
- effective training approaches to improve adaptive thinking skills
- how learner models can be best used to tailor training approaches
- extensible domain representations in ontologies
- effective use of ontologies in intelligent tutoring systems (ITS)
- representing meta-level control strategies for ITS
- methods for integrating an ITS with extant simulations

Extensions of Critical Thinking Tutors for Commanders

The CATS Tutoring system could be used to develop a family of tutors that share background knowledge and a common set of instructional strategies. Tutors to teach basic critical thinking skills can be introduced at an earlier level of army training at lower echelons than brigade or battalion level.

The CATS Tutor could provide training for specific staff officer positions. The S2 and G2 positions (intelligence officers) could learn OPFOR doctrine and practice fighting against Blue force (friendly) opponents. The CATS Tutor would then monitor how well trainees applied OPFOR (opposing force, enemy) doctrine and tactical reasoning, and look for oversights made with respect to OPFOR thinking.

Specific tutors for teaching critical thinking skills for particular branches within the Army (i.e., Infantry, Armor, Engineering, and Aviation) could also be developed. A tutor to teach adaptive thinking skills in the context of command of Armor would be an appropriate outgrowth of the CATS Tutor. Such a tutor might use the Steel Beasts simulation of the M1A2 to improve critical thinking skills of a commander at the tank, section, platoon, or company level.

Each tutor would share a common knowledge base and add its own specific depth of knowledge for an Army branch, staff position, or other training situation. The knowledge added to one tutor could be shared by the other tutors. Thus, over time, all the tutors would become more knowledgeable. This improvement in reasoning capability leads to improvement in the tutor's ability to answer questions, to monitor trainee performance in tactical studies, and in the ability to simulate student behaviors of different backgrounds, which can be useful in helping determine the cognitive strategy a trainee is using at any time.

A more ambitious extension would be to apply the CATS Tutor to joint training situations or for other services. The design of the tutor supports such extensions, but would require the development of middle-level ontologies and tactical simulations tailored to naval and air warfare.

References

AlphaTech, Inc. (1998, August 21). Evaluating and Critiquing Courses of Action. [Technical Memo HPKB-COA-3]. Burlington, MA: Author.

AlphaTech, Inc. (1998, August 21). Notation and Draft Grammar for Course of Action Sketches and Statements, and for Products of Mission Analysis. [Technical Memo HPKB-COA-1]. Burlington, MA: Author.

Balzer, R. & Goldman, N. (1999). Mediating Connectors, Robert Balzer and Neil Goldman. Proceedings of the 19th IEEE international Conference on Distributed Computing Systems Workshop, Austin, Texas May 31 - June 5, 1999, 73-77.

Bassi, L.J., Cheney, S., & Van Buren, M. (1997, November). Training and Development. Training Industry Trends (1997).

Bloom, B. S, (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4-16.

Carbonell, J.R. (1970). Mixed-Initiative Man-Computer Instructional Dialogues. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Clausewitz, C. (1984). On War. (M. Howard and P.Paret, Trans.). New York: Alfred A. Knopf. University Press. (Original work published in 1832).

Cohen, P., Schrag, R., Jones, E., Pease, A., Lin, A., Starr, B., Gunning, D., & Burke, M. (1998, Winter). The DARPA High-Performance Knowledge Base Project. AAAI Magazine.

Collins, A., Warnock, E.H., & Passafiume, J. (1975) Analysis and synthesis of tutorial dialogues. In Bower, G. (Ed.), The Psychology of Learning and Motivation (vol. IX). New York: Academic Press.

Deckert, J.C., Entin, E.B., Entin E.E., MacMillan, J. & Serfaty, D. (1996, January). Military Command Decision Making Expertise. ARI Research Note 96-15. Alexandria, VA: U.S. Army Research Institute.

Department of the Army (1997, September 30). Army Field Manual, FM 101-5-1. Washington, DC: Headquarters, Department of the Army.

Dvoretsky, D. (1991). Secrets of Chess Training. London: B. T. Batsford Limited.

Fletcher, D. (1996). Does this stuff work? Some findings from applications of technology to education and training. Proceedings of the Conference on Teacher Education and the Use of Technology Based Learning Systems. Warrenton, VA: Society for Applied Learning Technology.

Foss, C. (2000). Jane's Tanks and Combat Vehicle Recognition Guide. New York: HarperCollins Publishers.

Graesser, A.C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: A simulation of a human tutor. Journal of Cognitive Systems Research, 1, 35-51.

Hart, B. H. (1991). Strategy. (Rev. ed.). New York: Penguin Books.

Hollan, J.D., Hutchins, E.L., & Weitzman, L. (1984). STEAMER: an interactive inspectable simulation-based training system. AI Magazine, Vol 5, No. 2, 15-27.

Leonhard, R. (1994). The Art of Maneuver, Maneuver-Warfare Theory and AirLand Battle. Novato, CA: Presidio Press.

Lesgold, A. M., Lajoie, S. P., Bunzo, M., & Eggan, G. (1992). SHERLOCK: A coached practice environment for an electronics troubleshooting job. In J. Larkin & R. Chabay (Eds.), Computer assisted instruction and intelligent tutoring systems: Shared issues and complementary approaches (pp. 201-238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Li, J., Condoravdi, C. & Pease, A. (1999). From Visual to Logical Representation: A GIS- based Sketching Tool for Reasoning about Plans. Unpublished Teknowledge internal working paper, HPKB project.

Lussier, J.W. & Litavec, D. (1992). Research Report 1628. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lussier, J.W. & Saxon, T.F. (Nov 1994). Critical Factors in the Art of Battle Command. (ARI Study Report 95-01). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lussier, J. (1999, Spring). Think like a Commander. ARI Newsletter.

Lussier, J. (2001) Think like a Commander course. [PowerPoint slides]. Fort Leavenworth, KS: School for Command Preparation, Command and General Staff College and Alexandria, VA: U.S. Army Research Institute [Producer and Distributors].

Lussier, J. (2000). Adaptive Thinking Mentor's Guide. [CD-ROM] Army Transformation: Training for Full Spectrum Dominance.

Kotov, A. (1971). Think Like a Grandmaster. London: B. T. Batsford Limited.

Murray, W.R. (1991). An Endorsement-based Approach to Student Modeling for Planner-controlled Tutors. Proceedings of the Twelfth International Joint Conference on Artificial Intelligence. August 24 - 30, 1991. (pp. 1100 – 1106).

Murray, W.R. (1999). An Easily Implemented, Linear-time Algorithm for Bayesian Student Modeling in Multi-level Trees. Proceedings of the 9th International Conference on Artificial Intelligence in Education. Burke, VA: IOS Publications.

ProSIM Company. (1999). BCT Brigade Combat Team. User's Guide. Apex, NC: Author.

Rich, E. (1979). User modeling via stereotypes. Cognitive Science, 3, 355-366.

Ritter, S. & Koedinger, K. R. (1997). An architecture for plug-in tutoring agents. Journal of Artificial Intelligence in Education, 7 (3/4), 315-347.

Robel, M. K. (1999). Breaching 101. In BCT Brigade Combat Team User's Guide. (pp.117-122).

Sleeman, D.H. & Brown, J.S. (Eds.). (1982). Intelligent Tutoring Systems. London: Academic Press. [Note: Chapter 11 is on SOPHIE-I, II, and III].

Tzu, S. (1971). The Art of War. (S. Griffith, Trans.). New York: Oxford University Press. (Original work published circa 500 B.C.).

Wade, N. M. (1999). The Operations and Training SMARTbook. (Rev. Ed.). Lakeland, FL: The Lightning Press.

Wade, N. M. (1999). The Battle Staff SMARTbook. Lakeland, FL: The Lightning Press.

Woolf, B.P., Blegen, D.; Jansen, J.J., and Verloop, A. (1986). Teaching a complex industrial process. Proceedings of the National Conference on Artificial Intelligence, Philadelphia, PA, II, 722-728.