DISCUSSION CONTROL IN AN AUTOMATED SOCRATIC TUTOR

Eric A. Domeshek	Elias Holman	Susann Luperfoy
Stottler Henke Associates, Inc.	Stottler Henke Associates, Inc.	Stottler Henke Associates, Inc.
Somerville, MA	Somerville, MA	Somerville, MA

Abstract

Socratic interaction is a mainstay of instruction in professions such as law, business, and education, and to a lesser extent, in the military. Socratic instruction comprises a set of techniques that help students acquire and practice professional-level reasoning and decision-making skills. Our observations of exemplary tactical instructors demonstrate these techniques in use with military commanders. Unfortunately, this is a difficult and expensive form of instruction. Accordingly, we are extending the general drive towards development of Intelligent Tutoring Systems (ITSs) that automate military instructional expertise, seeking to understand and address the specific requirements of Socratic tutoring.

Socratic tutoring is characterized by (a) the setting of a thought-provoking problem, (b) a student's attempt(s) to provide solutions to the problem, (c) the instructor's repeated exploration and challenging of the student's solutions, which (d) elicits incremental justification, elaboration, refinement, and revision of both the student's understanding of the situation under discussion and their proposed solution. The prototypical structure of Socratic tutoring sessions involves a series of tutor-generated questions and student-generated answers. Major issues, then, are how should an automated Socratic tutor control its participation in such a dialog, and how can the behavior of such a tutor be specified cost-effectively?

This paper describes results from a U.S. Army Research Institute (ARI) sponsored project that developed a prototype Socratic tutor for battlefield command reasoning skills. In this application, the problem situations were tactical decision games. The tutor's behavior was modeled after that of expert tactical instructors. We present examples of the tutor's behavior, characterize its general capabilities, explore the discussion control mechanisms it uses to produce this behavior, show how scenarios and dialog moves are scripted, and analyze the costs and benefits of our approach, including its relation to prior work and likely future directions.

Author's Biographies

Dr. Eric A. Domeshek is an Artificial Intelligence (AI) Project Manager at Stottler Henke Associates, Inc. He received his Ph.D. in Computer Science from Yale University, where his work focused on cognitive modeling and technology, most especially on development of Case Based Reasoning (CBR). While working as Research Faculty at the Georgia Institute of Technology, he helped launch the EduTech institute, and became involved in educational applications of AI and CBR. He continued to work on educational and training technology while on faculty at Northwestern University's Institute for the Learning Sciences. For the last five years, Dr. Domeshek has conceived and managed a variety of Intelligent Tutoring System (ITS) projects at Stottler Henke. His main ITS efforts currently focus on dialogue-oriented tutors such as the Socratic ITS described here.

Mr. Elias Holman is an Artificial Intelligence Software Engineer at Stottler Henke Associates, Inc. He received his BA in Music Technology from Oberlin College, and his MEd at the Harvard School of Education in the Technology in Education program. Over the last four years, Mr. Holman has worked on several ITS projects at Stottler Henke, as well as projects focused on web-based collaboration.

Dr. Susann Luperfoy is a Senior Scientist at Stottler Henke Associates, Inc. She has over fifteen years experience leading and executing AI research projects, organizing stakeholders in emerging technology, and managing the design and delivery of commercial software products. She has directed technical efforts for the Department of Defense, NIST, DARPA, and NSF in addition to corporate internal R&D projects. She taught graduate courses and guided dissertation research at Georgetown University for 8 years, and has authored and edited over 30 publications, given numerous invited lectures, and organized several workshops on AI topics. She has international experience living and working abroad, collaborating on contracts with the Royal Technical Institute in Stockholm and as postdoctoral Visiting Scientist at Advanced Telecommunications Research (ATR) in Kyoto, Japan. She was also Founding President of SIGDIAL an international special interest group for community sharing of dialog data, analyses of data, analysis techniques and tools, and development of standards for marking up dialog data.

DISCUSSION CONTROL IN AN AUTOMATED SOCRATIC TUTOR

Eric A. Domeshek
Stottler Henke Associates, Inc.
Somerville, MA

Elias Holman
Stottler Henke Associates, Inc.
Somerville, MA

Susann Luperfoy
Stottler Henke Associates, Inc.
Somerville, MA

INTRODUCTION

In an earlier paper (Domeshek, Holman, & Ross, 2002) we presented arguments in favor of using Socratic tutoring to help teach critical battlefield command reasoning skills. In addition, we presented work on an experimental Intelligent Tutoring System (ITS) (Ong & Ramanchandran, 2000)—an early attempt to automate aspects of exemplary Socratic tutoring in the context of Tactical Decision Games (TDGs) (see, e.g. Schmitt, 1994). Our system aimed to understand, critique, and discuss students' proposed courses of action. In this paper, we focus on how the system manages interactive tutorial discussions.

Socratic instruction is a kind of teaching interaction typically applied in high-level professional education (e.g. law and business) and most often characterized by its external form: the teacher asks questions, and the student answers. These surface features actually reflect the deeper nature of Socratic instruction: it is an interactive and highly engaging form of scaffolded constructivist pedagogy. The question and answer format keeps the student engaged, but lets the teacher lead. The questions are posed in a sequence that leads the student to reconstruct the logic of expert situation analysis and decision-making for themselves.

This paper is structured as follows:

- We introduce the tutor and an example scenario,
- We explore the discussion control mechanisms it uses to produce its main behavior,
- We present examples of the tutor's behavior, along with a look at the authored structures that control that behavior, and
- Finally, we analyze the costs and benefits of our approach, including its relation to prior work and likely future directions.

TUTOR OVERVIEW AND SAMPLE SCENARIO

Each scenario starts with a short briefing, which the program gives as an automated slide presentation with voice-over narration. Figure 1 shows the narrative that launches one of the system's scenarios: "Enemy Over the Bridge" (EOTB). The centerpiece of the narration (and the ensuing interaction) is a map. The map is typically abstracted to the level of a sketch showing a limited range of territory, and is accompanied by icons that lay out the situation to be discussed. In addition, the friendly and enemy task organizations are presented as wire-diagrams, to the extent they are relevant to the problem (and to the extent the enemy organization is known).

At the end of the briefing, the student is transitioned to the system's main interaction screen. Figure 2 shows that screen at the point where the student has begun to describe their proposed course of action. The screen has three major parts: (1) the map, (2) the force-structure wire-diagram tree and timeline, and (3) the forms-based input/output area.

The interactive situation map, located in the upper right, takes up the most space on the screen. Initially it shows the situation as described at the close of the briefing. This map can be manipulated by both the student and the automated tutor. Icons representing forces can be dragged to new positions. Checkpoints can be laid down and used to describe taskings. Locations and forces can also be referred to by pointing with the mouse.

Beneath the map is the combined force-structure wire-diagram tree and timeline view. A standard tree-display can be expanded to show what is known about blue and red force-structures and their relationships. Each element represents a military unit, showing its name and an appropriate icon. If not already present on the situation-map, those icons can be dragged and dropped on the map to suggest where the corresponding unit should be located. To the right of the force-structure tree is a set of timelines that show major scenario events and taskings assigned to particular units. Like the map, the timelines can also be used to fill in details of proposed taskings by pointing and clicking with the mouse. Controls at the bottom of the screen allow you to scroll forwards and backwards in time, and to set an appropriate scale for the display.

The final part of the system's main display—taking up the entire left column—is the form-based input/output area. The top pane (the Transcript pane) accumulates a textual transcript of all major events that happen during the tutoring session. Not surprisingly, that pane starts out blank. The next pane down (the Prompt pane) shows only the most recent output from the tutor. Initially the prompt pane contains the tutor's request that the student enters "orders and reports" in response to the described situation. The bottom pane (the Input pane) provides a changing set of buttons, menus, and forms that offer the main way for the student to participate in dialog with the tutor. This area usually contains an array of buttons labeled with standard options such as 'Yes' 'No' or 'I Don't Know', followed by a cascading set of drop-down menus (here labeled "Choose from Main Menu"). These menus offer ways to say more complicated things to the tutor—to make statements or to ask questions:

- Students can make statements about facts in the military world such as the capabilities of weapons-systems, vehicles, and military units.
- Likewise, students can ask questions about the same sorts of facts.
- Students can also make statements about the facts of the current scenario—talking about terrain properties, or about military units, their activities, plans, and knowledge.
- Finally, students can also ask questions about the scenario, which will often take them back to parts of the scenario's introductory briefing.

The initial tutor request asks the student to enter orders and reports. The system does not yet have a good form defined for entering reports, so we focus here on issuing orders. There are several ways the student can assign a task to a unit. The most common approach is to use the mouse to drag the icon representing the unit to the location on the map where the student wants it to go. In response, the tutor opens the tasking form in the Input pane, and primes several fields of the form including **who** is receiving the order, **where** they are supposed to go, and the **route** by which they are supposed to get there. The other six fields of the tasking form are blank: the **role** the unit is supposed to play in the larger operation, the **start** and/or **end** times for the action, the specific **task**, the **enemy** force that is the target, and the **intent** behind the tasking. Figure 2 shows the result of dragging alpha company up the road and into the assembly area, as part of specifying an attack on the enemy mech located there.

Enemy Over the Bridge

You command a balanced tank-mech battalion task force consisting of 2 tank companies (A and B), two mech infantry companies (C and D), a scout platoon and a mortar platoon, plus your medical, support and maintenance platoons organized into a company train.

You are fighting a capable, mechanized enemy equipped with T-62s and BMPs and supported by towed and self-propelled artillery.

Host-nation forces hold the bridge and the river line to your west. You have been told the river is unfordable. Reconnaissance elements are operating west of the river. In 48 hours, the division begins a major offensive west across the river to destroy enemy forces in zone, with the main effort in your brigade's zone. Your battalion will spearhead the brigade's attack.

You have been instructed to occupy the assembly area shown on the map east of Hamlet in preparation for the 0500 attack the morning after next. You are moving to the assembly area as shown. At 0100 your scout platoon, which is forward reconnoitering the route to the assembly area, makes the following report:

> *Battalion, be advised have just made contact with a host-nation motorized reconnaissance patrol that was operating west of the river but about 2 hours ago was forced east across the river under fire. They came across the bridge and then via Wharton Farm. The reconnaissance patrol leader reports there is no sign of friendly forces holding the river line or the bridge and that enemy mech infantry and some tanks have been moving east across the bridge for almost 2 hours. He says he counted 10 T-62s in the last half hour; does not know how much mech. He says he has reported this twice to his higher headquarters. Over.*

A few minutes later the scout platoon leader adds the following:

> *Be advised we've got enemy mech infantry occupying our assembly area in strength. I say again they are enemy and not host-nation forces. I've got a solid visual on several BMPs. Don't know the size, but I estimate at least a company. They seem to be still moving into the area, over.*

Suddenly, you start to see artillery impacting in the woods just north of Alpha Company at the head of the battalion column.

Moments later, you hear automatic weapons fire from the direction of the assembly area. "*We're in contact!*" the scout platoon commander shouts over the radio.

What do you do?

Figure 1. Narrative Introduction to the "Enemy Over the Bridge" Scenario.

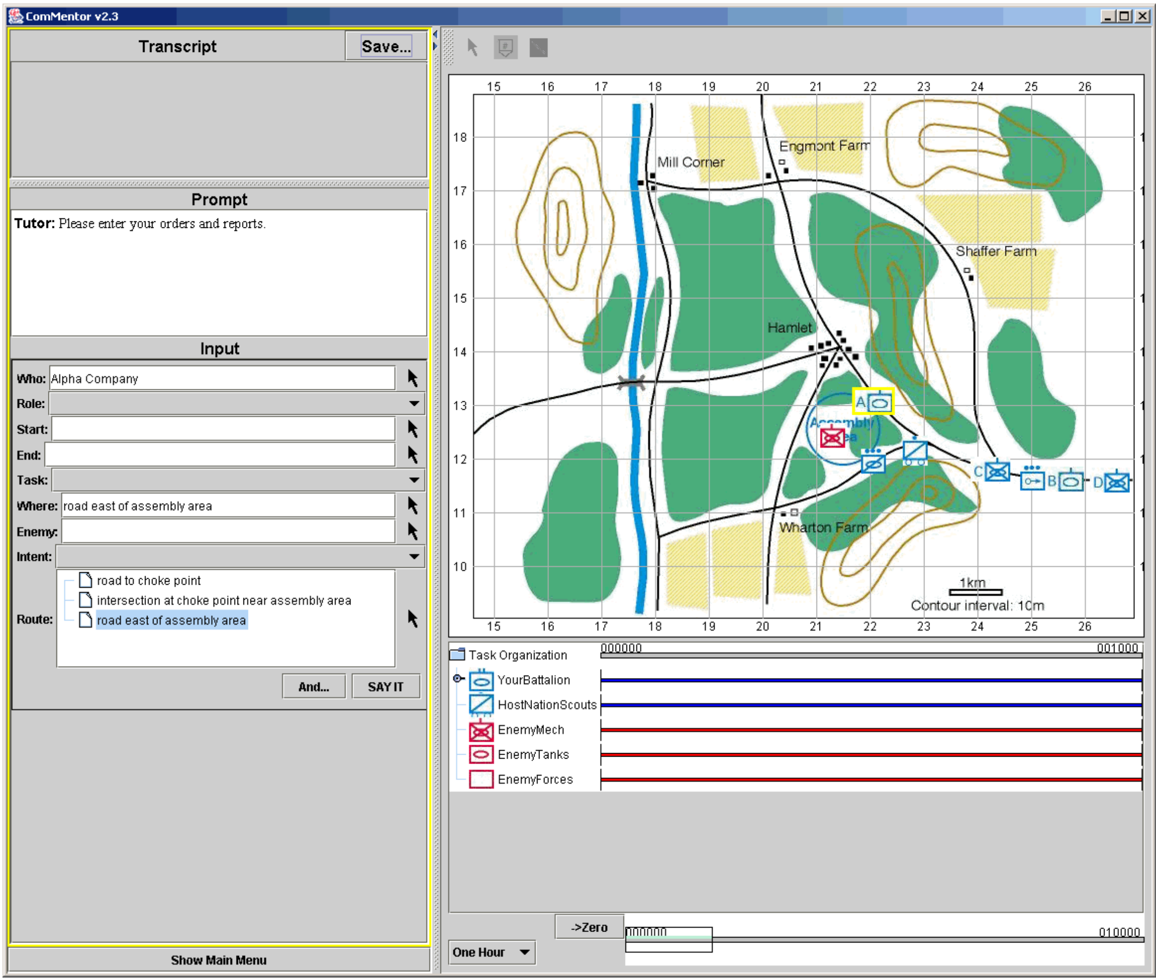

Figure 2. Main ITS Screen After Dragging Alpha Company Up the Road.

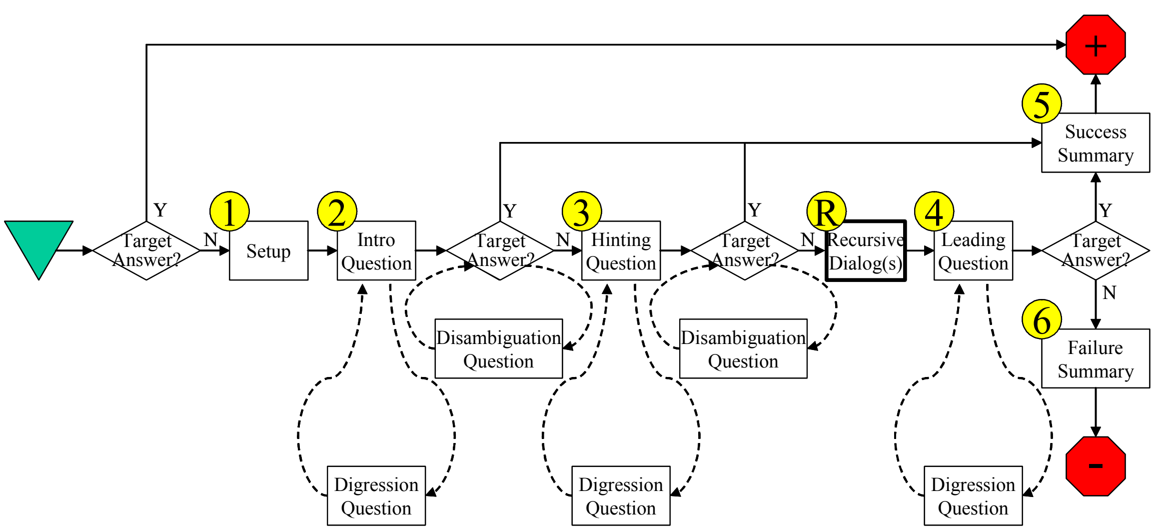

Figure 3. Flowchart of Processing for a Dialog Node.

TUTOR DISCUSSION MECHANISMS

Once a set of orders has been entered, the student clicks 'Done' and the system begins to respond to what has been said. When a Scenario is created, the author defines a set of Evaluations to watch for expected good and bad actions typically taken by students. The system matches patterns associated with Evaluations against a Scenario situation model which is continually updated by the student's input. Evaluations have two kinds of patterns: trigger patterns, and refinement patterns. Trigger patterns determine when the tutor will judge an Evaluation to be relevant, and so queue up its Dialogs for discussion. Refinement patterns are generally subsets or variants of the trigger pattern that prompt the tutor to ask clarification questions.

The Evaluations authored along with the Scenario are organized into trees under grouping structures called Scenes; Scenes thus provide one way of breaking an extended interaction into smaller chunks that focus on different issues. The Evaluations from the current Scene whose trigger patterns are satisfied by student input have their accompanying Dialog nodes placed on another structure called the Tutor Agenda. Then, when the tutor needs to start a new Dialog (e.g. when the Scene starts, or when a previous Dialog has finished) it pulls the first Dialog node from the front of the Agenda.

When added to the Agenda, Dialog nodes are placed in a position that corresponds, to the extent possible, to the Evaluation's position in the Scene's Evaluations tree. The idea is that the author should be able to control (or express preferences about) the order of Evaluation-related Dialogs by reordering the Evaluation tree. If, however, an Evaluation from earlier in the tree is triggered after some later Evaluation, and the Dialog for the later Evaluation has already started, then the newly triggered Evaluation must wait until the tutor has a chance to get around to the new topic.

The purpose of most Dialog nodes is to check if a student understands some point about the Scenario, and to lead them to see (and say) that point if they don't spontaneously get it. For instance, a Dialog node might represent an assessment of the red forces facing the student, where the available evidence ought to suggest a *mechanized battalion*. Here, the student succeeds on that node if they say they are facing a mech battalion. If when the node is pulled from the Agenda the student has already stated a correct characterization of the opposing force, then the node succeeds immediately; otherwise, processing of the node is initiated.

When a Dialog node is initiated the system enters a standard cycle that makes use of a set of up to six dialog Presentations, and possibly invokes a set of recursive Dialogs. All six Presentations are optional, as are the recursive Dialogs. The six Presentations are: (1) Setup, (2) Introductory Question, (3) Hinting Question, (4) Leading Question, (5) Success Summary, and (6) Failure Summary. When present, each of the three questions normally provides an opportunity for the student to say something to the system in response (or to ask a Digression Question of their own).

The standard sequence is illustrated by the flowchart in Figure 3. In this flowchart, processing starts at the triangle and ends at the octagons (stop-signs); the octagon with a '+' sign indicates the student is credited with having succeeded at the node, while the one with a '-' sign indicates the student is judged as having failed at the node. Diamonds represent tests, each of which has two outward bound arrows, one for when the test succeeds (labeled 'Y') and one for when the test fails (labeled 'N'). Finally, rectangles represent processing steps. The rectangles labeled 1, 5, and 6 represent the simple playing of a Presentation for the student. The rectangles labeled 2, 3, and 4 represent the playing of a Presentation and the requesting of input from the student. The rectangle labeled R represents the invocation of a recursive set of Dialog nodes. The dotted lines and rectangles represent optional paths for when the student either asks a Digression Question of their own, or gives an answer that triggers one of the tutor's follow-up Disambiguation Questions.

The rule for processing of recursive Dialogs depends on the particular Dialog node. A node may specify that its children should be handled using one of the options AND, OR, AT-LEAST, AT-MOST, or BETWEEN. The default rule is AND; in this case, all of the Dialog node's children will be executed. For an OR node, the children will be executed until either the student succeeds at one of them, or the tutor runs out of nodes. AT-LEAST is similar to OR, except some number more than 1 may be specified, and the system will keep attempting child nodes until the student succeeds on the designated number (or the tutor runs out of nodes to try). AT-MOST nodes and BETWEEN nodes (which combine AT-LEAST and AT-MOST conditions) will only start to make sense once the system begins to factor real-world clock time into its tutoring decisions.

One other complication affects the processing of Dialog nodes. Not only may a Dialog node have a target answer pattern, it may also have an entry test pattern. A Dialog entry pattern functions much like an Evaluation trigger pattern, in that the Dialog will not be attempted until the entry condition is satisfied. This can lead to Dialog nodes being executed in an order other than the one specified by the author. The primary use of this feature to date has been to construct option-evaluation Dialogs where pairs of child-nodes are authored to discuss different possibilities. One node in each pair has a given option as its target answer, and the discussion at that node tries to get the student to recognize the option exists. Meanwhile the other node of the pair has that same option as its *entry* condition so that only once the student mentions the option does the system start in on an evaluation discussion.

A distinguishing feature of ITSs is the extent to which they model individual student's knowledge and skills, and how they adapt based on those dynamic models. In this Socratic ITS, the Student Model is updated based on which Evaluations a student triggers, and based on which of the Dialog nodes that become active are exited successfully or unsuccessfully. Curriculum points can be linked to Evaluations in a positive or negative sense. If an Evaluation triggers, the student gets credit for mastery of its positive Curriculum points and/or loses credit for its negative Curriculum points. If an Evaluation *could have triggered, but didn't*, then credit/blame is reversed—its positive Curriculum points get blamed and its negative Curriculum points get credit. Curriculum points can also be linked to Dialog nodes, but only in one sense. If the Dialog node exits successfully, then the linked Curriculum points get credited. If the Dialog node exits with failure, then the linked Curriculum points get blamed.

We organized the system's curriculum using two different schemes. The first was rooted in the Think Like a Commander themes (Lussier, Ross, & Mayes, 2002) slightly expanded by adding issues to which our SMEs devoted considerable time. The second started from an abstract task-based breakdown of an officer's reasoning processes, and moved at the next level to an analysis based on work products and attributes of those products. Figure 4 shows the top levels of both those breakdowns. In all, there are currently over 150 Curriculum points defined in the system.

TLACThemes
  KnowTheFacts
  CommunicateEffectively
  FocusOnMissionAndHighersIntent
  ConsiderTheBigPicture
  ConsiderEffectsOfTerrain
  MakeEffectiveUseOfResources
  ModelAThinkingEnemy
  PrioritizeAndTakeActionInTime
  PlanFlexiblyAndForContingencies
  VisualizeAChangingSituation
	GeneralTasks
	  CommunicationTasks
	    IssuingOrder
	    IssuingReports
	    IssuingRequests
	  AnalysisTasks
	    AnalysisOfSituationAwareness
	    AnalysisOfTerrain
	    AnalysisOfForces
	    AnalysisOfImpact
	    AnalysisOfTerrainEvolution
	    AnalysisOfForcesEvolution
	    AnalysisOfAlternatives
	  PlanningTasks
	    PlanRelevance
	    PlanEfficacy
	    PlanEfficiency
	    PlanFlexibility
	    StandardPlanComponents
	  TimeManagementTasks
	    TimeAllocation
	    TimeSequencing
	    EventCoordination
	    ContextualTimeFactors

Figure 4. Top Levels of System's Curriculum Tree.

DISCUSSIONS AND DIALOG NODES

Here we examine a snippet from a Socratic ITS session transcript in detail, while looking at the underlying data and processes; we also look at some possible alternative behaviors that might have been exhibited in response to other student inputs. The segment in Figure 5 comes from near the start of the session and deals with the red situation. Note that student utterances appear awkward because their text comes from fill-in-the-blank forms, rather than being typed directly as it appears here. We use Figure 5 to explore Dialog node processing. The next section considers Evaluation node processing.

Tutor	Let's start by talking about the red forces and their situation.
Tutor	What sort of force do you think you're facing in your area east of the river?
Student	Enemy Mech is located at area east of the river
Tutor	Can you characterize what sized unit you are seeing here?
Student	company is located at area east of the river
Tutor	How much do you think you're facing in terms of armored forces?
Student	10.0 t-62s is located at area east of the river
Tutor	Yeah, the host nation recon patrol said they had counted 10 T62s cross the bridge in the last half hour.
Tutor	How long has the enemy been moving forces across the bridge?
Student	Move has been going on for 2.0 Hour
Tutor	Good. The host nation recon patrol reported they were forced back across the river under fire 2 hours ago. So enemy forces have been moving in for something like 2 hours.
Tutor	With reports of at least a company of mech, and most of a company of tanks, would you say you might be facing a battalion?
Student	Yes
Tutor	Yes. Given that you've seen at least a company of enemy mech in what was supposed to be your assembly area, and given reports of at least 10 T-62s, plus the fact they've been moving over the bridge for at least 2 hours, it seems reasonable to assume that there's something on the order of a red battalion already east of the river.
Tutor	What do you think the enemy is up to?

Figure 5. Sample EOTB Transcript Snippet.

This excerpt represents the tutor's initiation of the discussion associated with an Evaluation called Eval_RedPicture. Eval_RedPicture is unique in EOTB in that it is the only Evaluation that has *no* trigger pattern to satisfy. That means it will always be discussed in every student session. Further, given that it appears early in the list of top-level Evaluations, it will always be discussed early in a session.

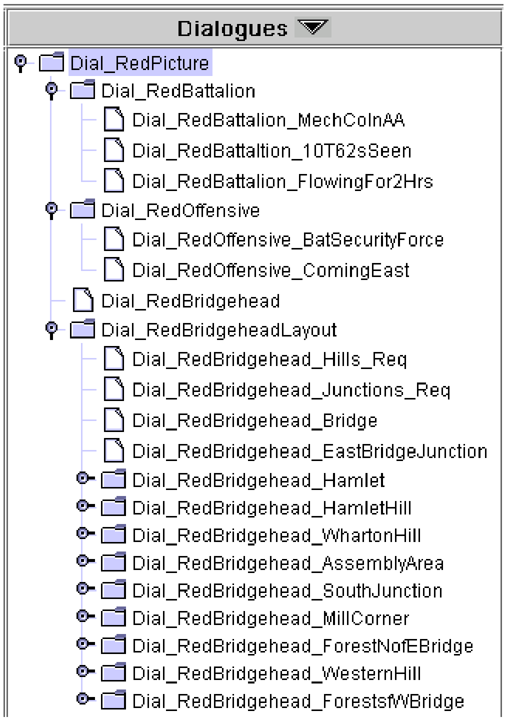

Figure 6. EOTB Dialog Tree for Eval_RedPicture.

Figure 6 shows much (though not all) of the tree of Dialog nodes for Eval_RedPicture. This Evaluation has a single root Dialog node Dial_RedPicture, which has 4 sub-nodes: Dial_RedBattalion, Dial_RedOffensive, Dial_RedBridgehead, and Dial_RedBridgeheadLayout. Figure 7 is the first of a set of diagrams that pick out the main contents of various EOTB Dialog nodes—in this case the contents of Dial_RedPicture (that is, the root of the tree shown in Figure 6). These diagrams show the entry and target patterns associated with a node, as well as the six Presentations: Setup, Introductory Question, Hinting Question, Leading Question, Success Summary and Failure Summary.

The Dial_RedBattalion Dialog node (see Figure 8) aims to get the student to acknowledge that they are likely facing a battalion-sized enemy element east of the river. Dial_Red-Offensive's target pattern (see Figure 9) is the assessment that this battalion-sized element is probably the advance guard of an enemy offensive. Dial_RedBridgehead's target pattern (see Figure 10) is the realization that Red is probably trying to form a bridgehead east of the river to secure the crossing. Dial_RedBridgeheadLayout does not itself have a target pattern, but it does contain a large number of sub-nodes ready to elicit and then discuss a variety of possible positions Red might try to use for the bridgehead.

Dial_RedPicture	
Entry:	
Target:	
Setup:	Let's start by talking about the red forces and their situation.
Intro:	
Hinting:	
Leading:	
Success:	
Failure:	

Figure 7. EOTB Dialog Node Dial_RedPicture.

Dial_RedBattalion	
Entry:	
Target:	(AtLocation (subject (& EnemyUnit Battalion)) (object =EastOfRiver))
Setup:	
Intro:	What sort of force do you think you're facing in your area east of the river?
Hinting:	
Leading:	With reports of at least a company of mech, and most of a company of tanks, would you say you might be facing a battalion?
Success:	Yes. Given that you've seen at least a company of enemy mech in what was supposed to be your assembly area, and given reports of at least 10 T-62s, plus the fact they've been moving over the bridge for at least 2 hours, it seems reasonable to assume that there's something on the order of a red battalion already east of the river.
Failure:	Well, given that you've seen at least a company of enemy mech in what was supposed to be your assembly area, and given reports of at least 10 T-62s, plus the fact they've been moving over the bridge for at least 2 hours, it seems reasonable to assume that there's something on the order of a red battalion already east of the river.

Figure 8. EOTB Dialog Node Dial_RedBattalion.

Dial_RedOffensive	
Entry:	
Target:	((! Attack Guard Occupy) (who EnemyUnit) (where =EastOfRiver))
Setup:	
Intro:	What do you think the enemy is up to?
Hinting:	What task might a battalion-sized force be carrying out in this push across the river?
Leading:	So would you accept that a battalion-sized force could be the advance guard for a larger offensive operation moving into our territory east of the river?
Success:	Yes, since you've got most of a battalion on your side of the river, it's not unreasonable to guess that the enemy has beaten us to the punch, launching an offensive to the east, just as we were planning to launch one to the west.
Failure:	Well I would say that given you've got most of a battalion on your side of the river, it's not unreasonable to guess that the enemy has beaten us to the punch, launching an offensive to the east, just as we were planning to launch one to the west..

Figure 9. EOTB Dialog Node Dial_RedOffensive.

Dial_RedBridgehead	
Entry:	
Target:	(MilitaryOperation (who EnemyUnit) (why Secure) (where Bridgehead))
Setup:	
Intro:	If a red battalion has just crossed over the river in advance of a major red offensive, what do you think that battalion is now trying to do?
Hinting:	What would you do if you had just pushed a battalion into enemy territory across a bridge?
Leading:	Would it be reasonable for the red forces now east of the bridge to be trying to secure their bridgehead?
Success:	Absolutely. If we had just managed to push a battalion across a river into enemy territory, it's safe to say that we would be thinking about how to secure the bridgehead for follow-on forces.
Failure:	Actually I think if we had just managed to push a battalion across a river into enemy territory, it's safe to say that we would be thinking about how to secure the bridgehead for follow-on forces.

Figure 10. EOTB Dialog Node Dial_RedBridgehead.

Dial_RedBattalion_MechCoInAA	
Entry:	
Target:	(AtLocation (subject (& Company (! Mechanized BMP))) (object =EastOfRiver))
Setup:	
Intro:	How much mech do you think you're facing?
Hinting:	What did your scouts tell you was sitting in your assembly area?
Leading:	
Success:	Right. The scouts said there was at least a company of enemy mech in your assembly area, possibly more moving into the area.
Failure:	What the scouts actually said was there was at least a company of enemy mech in your assembly area, and possibly more moving into the area.

Figure 11. EOTB Dialog Node Dial_RedBattalion_MechCoInAA.

The tutor utterance, "Let's start by talking about the red forces and their situation." is actually the Setup Presentation associated with the Dial_RedPicture node. From the flowchart in Figure 3 we see that so long as a Dialog node's target pattern is not satisfied before we start processing, then the first thing that is done is to produce the Setup Presentation. Since this node has no target pattern, it cannot be satisfied, and thus this Setup is used. Dial_RedPicture has no Introductory, Hinting, or Leading questions, but it does have a set of recursive Dialogs. Thus the next item in the transcript is an utterance from the first child node beneath Dial_RedPicture, which is Dial_RedBattalion.

The tutor utterance, "What sort of force do you think you're facing in your area east of the river?" is the Introductory question from Dial_RedBattalion. This is because this node has no Setup.

The student input, "Enemy Mech is located at area east of the river" is a narrow answer to the tutor's question. The student simply pointed to the only red icon on the map (which the tutor understands to be an enemy mech company). The target pattern for this node specifies an enemy mech battalion, so this initial answer is not fully satisfactory. The node also has a set of disambiguation questions, one of which matches the student's answer and prompts for the missing echelon or unit size.

The tutor utterance, "Can you characterize what sized unit you are seeing here?" results from triggering a Dial_RedBattalion disambiguation question. Such extra prompts can provide an opportunity for the student to think a bit more deeply about the situation and the question that was really being asked. Other disambiguation questions trigger if the student says there is a battalion east of the river (but not that it is an enemy battalion), or if they say there is an enemy battalion (but not that it is east of the river). The student input "company is located at area east of the river" is taken as fleshing out the original answer, and so is interpreted as an "*enemy company* is located at area east of the river."

In this case the student has stuck with their initial shallow interpretation of the first question so Dial_RedBattalion's pattern is not satisfied and processing of the node continues. Since Dial_RedBattalion has no Hinting question, processing moves on to its recursive Dialog nodes: Dial_RedBattalion_Mech-CoInAA, Dial_RedBattalion_10T62sSeen, and Dial_RedBattalion_FlowingFor2Hrs. The idea is that by reviewing these three points of the initial situation description (ideally, getting the student to generate them), the argument for believing there is an enemy battalion east of the river will become apparent.

When the tutor starts on these recursive nodes, its first utterance is, "How much do you think you are facing in terms of armored forces?" This is the Introductory question for the second sub-node, Dial_RedBattalion_10T62sSeen. The system effectively skips over the first sub-node Dial_RedBattalion_Mech-CoInAA (see Figure 11) because the student has already said that there is an enemy Mech company east of the river; the target pattern for the node is thus pre-satisfied and so, following the flowchart, the system goes directly to the success end-state, skipping over all Presentations.

The student could answer a number of ways to this question. They could base their answer on the report from the scenario introduction and say, "*10 T-62s* is located at area east of the river." The system would also accept, "*10 tanks* is located at area east of the river." Finally, the system will also accept the inference-based answer, "*company* is located at area east of the river" (interpreting that as meaning, "*armor company* is located at area east of the river").

Any of these answers would satisfy the node's pattern and cause the tutor to follow the flowchart path to use its Success Summary: "Yeah, the host nation recon patrol said they had counted 10 T62s cross the bridge in the last half hour." Other answers would lead the tutor to try again with its Hinting question, "What did the host nation recon unit tell you they had seen cross the bridge?" A correct answer here would lead to the same Success Summary. Since the node has no Leading question, an incorrect answer to this second question would lead to the Failure Summary: "The host nation recon patrol actually reported that they had counted 10 T62s cross the bridge in the last half hour."

Whichever way Dial_RedBattalion_10T62sSeen ends, the system will move on to the Dial_RedBattalion_FlowingFor2Hrs node and ask its Introductory question, "How long has the enemy been moving forces across the bridge?" Again, this is a relatively straightforward factual question, so we expect the student will get it right. Still, the system has a Hinting question and a Failure Summary as backups.

Here we assume the student recalls the relevant fact from the introductory briefing and correctly answers, "Move has been going on for 2.0 hour." This elicits the tutor's Success Summary: "Good. The host nation recon patrol reported they were forced back across the river under fire 2 hours ago. So enemy forces have been moving in for something like 2 hours." The tutor has now finished all three of its recursive nodes, and so continues through the flowchart for Dial_RedBattalion.

It finds the Leading question, "With reports of at least a company of mech, and most of a company of tanks, would you say you might be facing a battalion?" We assume the student agrees with this argument, answers "Yes." and gets the Success Summary: "Yes. Given that you've seen at least a company of enemy mech in what was supposed to be your assembly area, and given reports of at least 10 T-62s, plus the fact they've been moving over the bridge for at least 2 hours, it seems reasonable to assume that there's something on the order of a red battalion already east of the river."

Of course the student might not be convinced by the detailed situation analysis and so insist on answering, "No." This is probably the situation where the system most seriously diverges from the behavior of live expert tutors. A clever tutor would be able to explore what the student was thinking and why they were not accepting a seemingly logical conclusion. Though an author could have prepared more than one argument for this point, here, we assume the system would simply have to insist on its interpretation and move on.

Finishing our initial transcript segment, "What do you think the enemy is up to?" is the Introductory question from Dial_RedOffensive, which represents the tutor moving on to the next node under Dial_RedPicture.

TOPICS AND EVALUATION NODES

In the previous section we traced in detail part of a run through the Dialog tree under the universally relevant Evaluation Eval_RedPicture. The EOTB scenario has about twenty other Evaluations, each organizing discussion topics, many of which are only raised by the tutor when the student exhibits particular behaviors.

Beyond building a coherent Red story, including where their forces might be, and what they might be trying to do some of the other major themes covered by EOTB Evaluations include: (1) recognizing the key problem and the general nature of the necessary response (i.e. the enemy is flowing into our territory over the bridge we need for our mission, so somehow we have to take back control of the bridge to stop the flow and salvage our own offensive), (2) prioritizing actions based on that understanding of what is most important (i.e. don't spend all your resource fighting the enemy you happen to see in front of you rather than dealing with the bridge), (3) considering and evaluating a range of COAs for getting to the bridge, (4) understanding what it means to take control of the bridge, (5) thinking in some detail about the risks in the situation and the appropriate sequencing of activities, and (6) considering in some detail the uses of a reserve force in a fluid and uncertain situation such as this.

As an example consider the Evaluation Eval_Attacking-AAMultiCos. This node delves into the possible risk of fratricide between Blue attacking forces when a student tasks two different companies (including at least one tank company) to attack the assembly area. Triggering of this Evaluation, and execution of the accompanying Dialogs only happens when the student lays out a relevant family of COAs. If the student does not attack the assembly area, if they use only one company, or if they use two mechanized infantry companies for an attack, this discussion would not be entered.

There are several different types of Evaluations in the system. Eval_RedPicture is an example of a **run-always** Evaluation—it has no trigger pattern and so is run for all students. Eval_AttackingAA is an example of a **trigger-only** Evaluation—it has no Dialogs of its own, but controls a set of nested Evaluations, ensuring that more detailed Evaluations of *how* the student proposes to attack the assembly area are only checked once the tutor has established that the student is in fact attempting an attack there (e.g. Eval_AttackingAA-MultiCos mentioned above). Eval_HoldingBridge is an example of an **absence-triggered** Evaluation—it's pattern checks for things that the student has *not* said or done, such as a failure to place friendly forces on the far side of the river to hold the bridge (assuming some unit has been tasked with retaking the bridge).

There are also a range of possible relationships between Evaluations. Evaluation nesting (described for trigger-only Evaluations) is one such relationship. Nested evaluations often extend the trigger patterns of their parents. For instance Eval_AttackingBridge checks that some unit has been assigned to attack the bridge, and remembers the tasking and the unit; then the nested Eval_Envelopment checks that some unit has been assigned to attack the assembly area, *and that the two taskings and two units are distinct*. Another kind of relationship between Evaluations is one Evaluation's trigger pattern explicitly looking for student input that is solicited in the discussion of another Evaluation. For instance Eval_AttackingAANoEvac critiques the student for not moving their scouts out of the way of an oncoming assault, and its discussion ends with a final prompt to the student: "Why don't you give the Scouts some orders about what they should be doing—especially where they should go." The sibling Evaluation Eval_AttackingAAEvacMethod and its children are ready to respond to various orders the student may choose to give to the scouts.

COSTS, BENEFITS, AND FUTURE DIRECTIONS

The previous two sections presented examples giving some sense for how the system makes decisions at the Dialog and Evaluation levels, and what sorts of adaptive student experiences result.

Developing tutoring scripts that produce appropriate behavior is heavily dependent on the authoring process, which in turn relies on authoring tools to guide and speed Scenario development while helping to minimize mistakes. We devoted substantial effort to the initial authoring tool suite. Based on our initial experience developing two complete scenarios, we have generated a detailed list of ten authoring tasks and estimates for the costs associated with each task.

We estimate the core effort devoted to getting a new scenario into a state suitable for initial student use at 14-20 days—roughly 1 person-month of effort. An initial testing period with a set of sample students—say 10 students at 2 hours apiece—would likely reveal the majority of issues with the newly authored Scenario. If we allow 1-2 hours to review each sample transcript and up to 4 hours to address the issues it raises, we would allocate another 2 weeks for Scenario review, and refinement before release to the general population.

The estimates so far also leave out a significant set of non-core tasks aimed at refining representations used across all scenarios: the domain ontology, input-form templates, and the curriculum (with accompanying presentations). Effort devoted to authoring such cross-scenario resources ought to decline over time (as more scenarios have been authored and most common requirements have been met). Still it seems prudent to allocate 2 additional weeks to such cross-scenario enhancements. That suggests a final estimate around 2 person-months of authoring effort per scenario, which at billing rates typical for the highly skilled personnel involved could run up to $50,000/scenario.

To balance these costs, the benefit is a kind of one-on-one tutoring for battlefield command skills that is rarely available to typical Army officers. Alternate techniques for building automated Socratic tutors have rarely been applied to so complex or open-ended a problem. The leading lower-cost contender (Graesser, et. al., 2001)—a technique that relies heavily on light-weight text-processing techniques originally developed for information retrieval tasks—is not particularly well suited to building a tutor that is sensitive to the fine distinctions that may affect the appropriateness of tactical courses of action, or that may reveal important gaps in student understanding or rationale.

That said, we note complementary efforts in a related Phase II SBIR (Ryder, Graesser, McNamara, Karnavat, & Popp, 2002), as well as a computer-based program developed by ARI being using in the Armor Captains Course at Fort Knox's University of Mounted Warfare (Shadrick & Lussier, 2002). All of this work traces its roots to earlier analyses of the underlying training problem presented by Lussier, Ross, & Mayes (2002).

Future development for this line of work leads in three main directions: (1) driving down the costs of authoring by building better authoring tools that embed more knowledge about both (a) the domain and (b) the Socratic tutoring process, (2) improving the quality and naturalness of the interaction by addressing issues such as (a) incorporation of a stronger underlying ontology with better inference capabilities, and (b) revisiting the issue of language-based input, and (3) integrating this Socratic style of instruction with other approaches that form a natural complement including (a) case-method instruction, and (b) real-time tactical role-play.

As this list of remaining issues suggests, cost-effective high-quality Socratic tutoring for battlefield command is not yet an established capability. We cannot yet put a senior officer and talented mentor in a box. ARI's sponsorship of this line of work has been framed by recognition that military command is an extremely complex and open-ended problem. The SBIR program was chosen as a vehicle for this line of necessary high-risk high-reward research, in part on the understanding that even intermediate results and approximations to the final goal could be quite valuable. In fact, some of the capabilities established in our work on this system are already being applied in other domains such as emergency medical response. Meanwhile some of the outstanding issues raised by this work are beginning to be addressed in ongoing research.

Fortunately, there are many training needs that, while not as difficult as preparing commanders to face fluid tactical situations, can still benefit from incorporation of some degree of discussion-oriented scenario-based tutoring, multi-modal interaction for situation and plan visualization, and student modeling that tracks curriculum exposure and success in mastery to inform tutoring decisions. For such applications, our existing Socratic ITS technology provides a solid base.

ACKNOWLDGEMENTS

This work was supported by the U.S. Army Research Institute, Fort Knox, Kentucky. Funding was provided under the Cognitive Readiness SBIR program by the Office of the Secretary of Defense. The authors wish to acknowledge the tutoring modeled by our collaborators, Major (retired, USMC) John Schmitt and Brigadier General (retired, USMC) Keith Holcomb whose collective expertise in training tactical thinking forms the basis for our research and development.

REFERENCES

Domeshek, E. A., Holman, E., & Ross, K. G. (2002). Automated Socratic Tutors for High-Level Command Skills. *Proceedings of the 2002 Interservice/Industry Training, Simulation, and Education Conference,*

Graesser, A.C., Person, N.K, Harter, D., and the Tutoring Research Group (2001). Teaching Tactics and Dialog in Autotutor. *International Journal of Artificial Intelligence in Education*.

Lussier, J. W., Ross, K. G., & Mayes, R. (2000). Coaching Techniques for Adaptive Thinking, *Proceedings of the 2000 Interservice/Industry Training, Simulation, and Education Conference*.

Ong, J. & Ramanchandran, S. (2000) Intelligent Tutoring Systems: The What and the How. *ASTD Learning Circuits*. Feb, 2000. Available at http://www.learningcircuits.org/feb2000/ong.html

Ryder, J. M., Graesser, A. C., McNamara, J., Karnavat, A. & Popp, E. (2002). A dialog-based intelligent tutoring system for practicing command reasoning skills. *Proceedings of the 2002 Interservice/Industry Training, Simulation and Education Conference*.

Schmitt, J. F. (1994). *Mastering tactics*. Quantico, VA: Marine Corps Association.

Shadrick, S. B. & Lussier, J. W. (2002). Think Like a Commander at the Armor Captain's Career Course. *Proceedings of the 2002 Interservice/Industry Training, Simulation and Education Conference*.