Generalizing Automated Assessment of Small Unit Tactical Decision Making

Randy Jensen, Bart Presnell
Stottler Henke Associates, Inc.
San Mateo, CA
jensen@stottlerhenke.com,
presnell@stottlerhenke.com

Marshell G. Cobb
US Army Research Institute
Ft. Benning, GA
marshell.g.cobb.civ@mail.mil

James Lunsford
Decisive-Point, LLC
Kansas City, MO
jim@decisive-point.com

LTC Daniel Kidd
US Military Academy
West Point, NY
daniel.kidd@usma.edu

ABSTRACT

Intelligent adaptive training technologies augment or emulate the role of human instructors to support self-training through experiential practice, with individualized guidance and feedback. They can be powerful training tools, especially as budget constraints trigger reductions in instructional manpower and in the opportunities for live training. But the complexity and diversity of learning objectives and practice environments can make the authoring task time-intensive, frequently requiring a highly customized development process. There are no universal authoring solutions, although concepts have been proposed to reduce the burden by isolating reusable components. For example, an abstracted simulation integration layer can standardize performance data collection, so that instructional components and even logic can be reused across applications. In practice, an integrated trainer draws a virtual dividing line where instructional logic is either more or less abstract from the state and event data flow in a simulation environment.

This paper describes a self-training capability integrated with an existing tactical decision making game, as a case study for the tradeoffs in abstracting a set of automated assessment measures and techniques for complex real-world training objectives. The game environment is *Follow Me*, a small unit leader tactical training game used by instructors and cadets at the United States Military Academy at West Point. This training audience presents a classic use case for self-directed training; due to limited classroom time, cadets previously received little direct feedback on their own exercise performance. However with the addition of automated evaluation and feedback capabilities, the reach of instructors is extended and replicated. While this trainer operates with a two dimensional game, some of the same performance measures could be applied to a three dimensional game or other simulation platform. Examples from the *Follow Me* trainer are explored in terms of generalization and the entailed authoring implications for both instructional components and experiential environments.

ABOUT THE AUTHORS

Randy Jensen is a group manager at Stottler Henke Associates, Inc., working in training systems since 1993. His research areas include adaptive training, distributed learning, game-based training, behavior modeling, and natural language processing. He has led projects to develop Intelligent Tutoring Systems and automated after action review tools for the Army, Air Force, and Marines. Mr. Jensen is also currently leading an effort for the Army to integrate real-time automated performance assessment and feedback with a tactical decision making game. He holds a B.S. with honors in symbolic systems from Stanford University.

James Lunsford is the President of Decisive-Point, and an expert in the design, development and use of "serious games." He has developed and delivered serious games for the Army, Air Force, and Marines, and was a key participant in the DARPA DARWARS program. Mr. Lunsford graduated from the Virginia Military Institute in 1980 with a Bachelor of Science degree in Civil Engineering, and from the U.S. Army Command and General Staff College in 1993 with a Masters of Military Arts and Sciences. He is a retired Army officer with twenty years of active service. His military experience includes command and staff positions in numerous airborne infantry units, a two year tour with the British Parachute Regiment, and an assignment as an associate professor of tactics at the U.S. Army Command and General Staff College.

Bart Presnell is a software engineer at Stottler Henke Associates, Inc. He holds an M.S. in Computer Science from the Georgia Institute of Technology. Prior to joining Stottler Henke, Mr. Presnell worked for seven years as a software engineer for Electronic Arts and Stormfront Studios.

Dr. Marshell G. Cobb. In his current position with the US Army Research Institute (ARI), Dr. Cobb leads a field research team focused on developing effective tailored training strategies and tools to enhance institutional training across the Army. He has authored or co-authored numerous technical reports and articles on a wide range of topics, including skill acquisition and retention, drill and platoon sergeant training, trainee socialization, and digital training apps development.

LTC Daniel Kidd is the Director of the West Point Simulation Center, and a Functional Area 57 Simulations Officer. He was commissioned as an Infantry Officer from the Morehead State University ROTC. He has commanded from platoon through company and has served in staff positions at the battalion and brigade level. He has served in mechanized, air assault infantry units, as well as in ROTC instructor positions, USAR training units and as Team Chief for an Iraqi National Police Transition Team. He holds an M.S. in Adult, Continuing and Occupational Education from Kansas State University.

Generalizing Automated Assessment of Small Unit Tactical Decision Making

Randy Jensen, Bart Presnell
Stottler Henke Associates, Inc.
San Mateo, CA
jensen@stottlerhenke.com,
presnell@stottlerhenke.com

Marshell G. Cobb
US Army Research Institute
Ft. Benning, GA
marshell.g.cobb.civ@mail.mil

James Lunsford
Decisive-Point, LLC
Kansas City, MO
jim@decisive-point.com

LTC Daniel Kidd
US Military Academy
West Point, NY
daniel.kidd@usma.edu

INTRODUCTION

Intelligent adaptive training technologies augment or emulate the role of human instructors to support self-training through experiential practice, with individualized guidance and feedback. They can be powerful training tools, especially as budget constraints trigger reductions in instructional manpower and in the opportunities for live training. But the complexity and diversity of learning objectives and practice environments can make the authoring task time-intensive, frequently requiring a highly customized development process. This is often cited as a factor that has limited the adoption of intelligent training systems in military instruction, despite a track record of successfully demonstrated learning results. The consequence of custom implementation is that prospective future applications using similar methods for other domains face comparable development costs due to limited reuse of components and minimal economy of scale.

Researchers have made efforts to quantify development costs for intelligent tutors with varying combinations of feature sets and instructional strategies, using standard metrics such as the ratio of development hours to instructional hours (Folsom-Kovarik and Schatz, 2011). Still, definitive comparisons are difficult, considering the large number of variables defining different training requirements, applications, development processes, and instructional use cases. Further research may help to make the case for the levels of development cost that can be considered justifiable for different mixes of requirements, in terms of learning results and reductions in instructional manpower. In parallel with these worthwhile questions, there have been efforts underway to explore methods for generalization and reuse, with the goals to economize development and ultimately promote wider adoption of useful training technology (Chipman, Olney and Graesser, 2005; Sottilare, Goldberg, Brawner and Holden, 2012). One starting point for efforts to reduce authoring burdens is to isolate reusable components. For example, an abstracted simulation integration layer can standardize performance data collection, so that instructional components and even logic are independent from the experiential environment (Sottilare and Gilbert, 2011). The reward comes when assessment components can then be reused with another application; reducing the task to developing or integrating with a substitute integration layer, rather than re-implementing all components. In practice, an integrated trainer draws a virtual dividing line where instructional logic is either more or less abstract from the state and event data flow in a simulation environment. Even the act of conforming to a compartmentalized design architecture can help to enforce a component breakdown that will be better suited to potential reuse, for example by following a standard where simulation data protocols are implemented separately from assessment and feedback code.

This paper aims to contribute by offering observations from the context of a specific training application, using examples of individual components and methods that may be abstracted for potential reuse. In order to support self-directed training, constraint-based intelligent tutoring methods were integrated with an existing tactical decision making game, using automated assessment measures for complex real-world training objectives. The game environment is *Follow Me*, a small unit leader tactical training game used by instructors and cadets at the United States Military Academy at West Point. This training audience presents a classic use case for self-directed training. Due to limited classroom time, cadets previously received little direct feedback on their own exercise performance.

However with the addition of automated evaluation and feedback capabilities, the reach of instructors is extended and replicated. While this trainer operates with a two dimensional game, some of the same performance measures could be applied to a three dimensional game or other simulation platform. Examples from the *Follow Me* trainer are explored in terms of generalization and the entailed authoring implications for both instructional components and experiential environments.

TRAINING APPLICATION BACKGROUND

The goals for training tactical decision-making often center around giving experiential practice under dynamic conditions where there is no singular best answer. While there may be best practices and even doctrinal guidance for elements of how a tactical plan should be conducted, there also can be many viable courses of action with different tradeoffs in speed, risk, and achieving mission objectives. Thus training aims to combine the instruction of basic principles with the art of understanding the tactical situation, weighing alternatives, acting decisively, and remaining flexible. In broad terms, this can be true for tactical decision-making at different echelons and in different operational settings. This makes scenario-based training highly effective for tactical skills, but only if experiential events are accompanied with guided feedback. Practice alone is not sufficient for learning; it must be tied with feedback which is closely tied to performance assessment (Salas, Rosen, Burke, Nicholson and Howse, 2007; Ericsson and Ward, 2007). For complex decision-making domains, experienced human instructors are naturally expert at seeing performance patterns that call for feedback, either positive or negative. Therefore the objective of an automated intelligent tutor is to replicate this kind of assessment and feedback, in the unstructured free-play setting of a virtual environment. The following discussion provides background for the approach to implementing this with *Follow Me*, by giving an overview of the existing instructional setting and methods.

West Point currently uses *Follow Me* to facilitate the instruction of small unit leader tactical skills in the classroom. The traditional method for teaching these concepts to students includes lecture, written problems, the use of terrain boards or sand tables (example in Figure 1), tactical exercises without troops, and finally tactical exercises in a field environment.

Figure 1. Sand Table Exercise at West Point (image with USMA permission)

The classroom exercises, while relatively simple to conduct, provide only an elementary understanding of the basic concepts because the students are challenged to accurately conceptualize the dynamic effects of time, space, distance, terrain, and capabilities in the execution of a tactical plan. At the other end of the spectrum, the field exercises provide a greater understanding of dynamic tactical concepts and how they are applied, but are resource intensive and are limited to the available terrain types where training is conducted. The benefit of a game-based trainer is that it provides quick, fast paced, recorded exercises within a myriad of battle environments and against a wide variety of threats, and can be conducted by individuals or teams based on the complexity of the problem and the nature of the learning objectives.

In the existing classroom sequence where *Follow Me* has been used, cadets typically read the tactical concepts for homework, discuss the principles in class and then conduct an individual, single-player exercise within *Follow Me* that lasts about 15 to 20 minutes. Although some implicit learning occurs during scenario execution, most of the actual learning occurs afterwards with the instructor-led guidance provided in after action review. Following the completion of an exercise, the instructor selects several cadets to brief the class on their decisions and thought processes while the students' exercises are replayed using the captured after action review (AAR) log files on a large display (Figure 2).

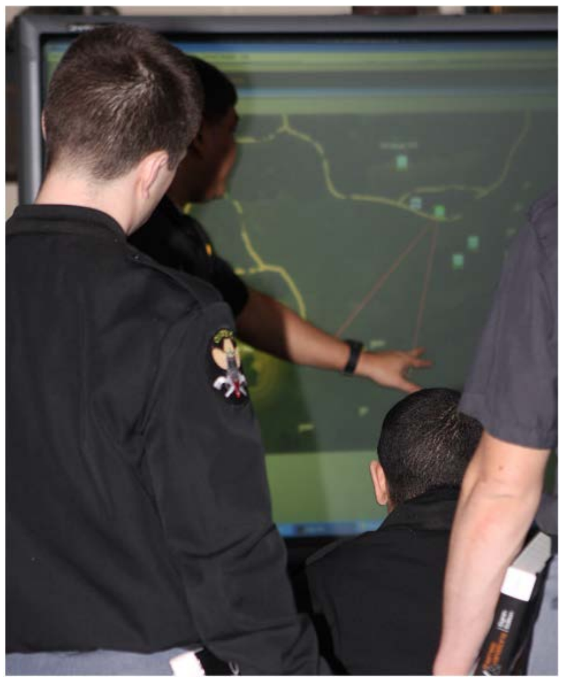

Figure 2. West Point Cadets Conducting an AAR after a *Follow Me* Exercise (image with USMA permission)

As the designated student briefs his/her solution, the instructor facilitates discussion while the remaining cadets ask questions or provide their opinions and insights. The instructor shapes learning by pausing the exercise playback at specific points where the most significant teaching points occurred within the exercise. Instructors ask questions or provide carefully worded comments to facilitate discussion and the students' understanding of what happened, why it happened, whether it was doctrinally correct or not, and how it affected the outcome of their mission. This is an extremely valuable process for learning, because of the individualized assessment and feedback. However, because of the time constraints on instructors, many cadets learn from a few specific examples chosen for discussion, but have limited opportunities for direct feedback on their own decisions and performance. This is one of the motivations for an intelligent tutoring capability in this instructional context, to create opportunities for each learner to go through multiple exercises and receive tailored feedback every time.

An example skill often practiced within *Follow Me* involves the proper employment of weapons systems during a tactical operation. Figure 3 is a screenshot from the AAR for an exercise conducted by a West Point cadet. During the AAR playback of the exercise, the instructor pauses the playback in order to highlight critical shortcomings in the student's choice of positions for the machine gun sections. In Figure 3 below, one of the machine gun sections is selected, with its sector of fire shown as a cone.

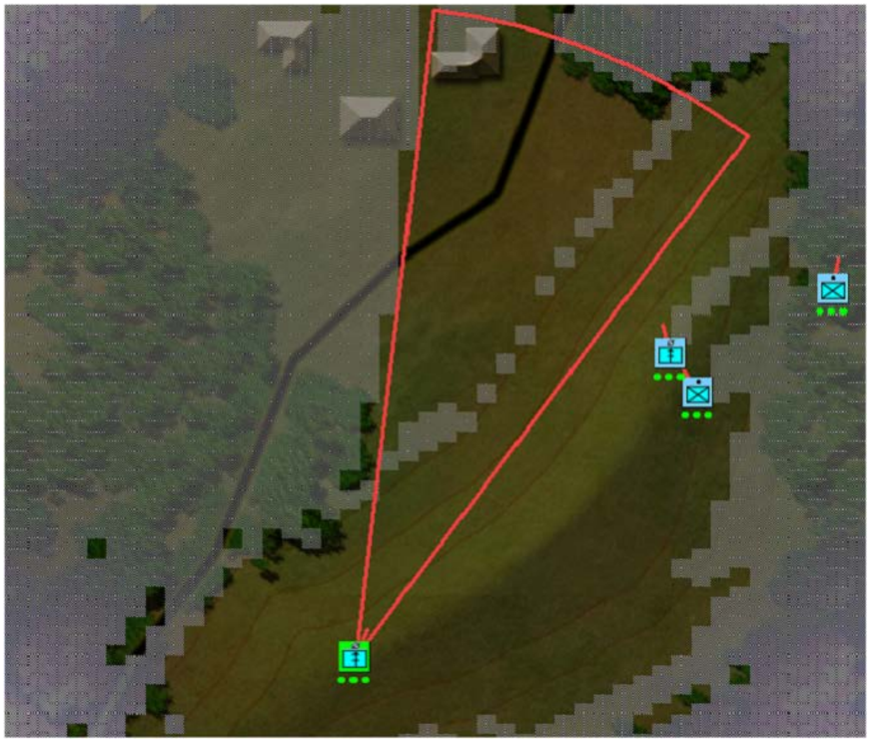

Figure 3. Machine Gun Section Visibilities and Sector of Fire Shown in *Follow Me* AAR

Each machine gun section within an Army light infantry platoon constitutes 20% of that platoon's combat power, and should be employed in areas where they can be properly controlled by the platoon leader. In this particular case, the selected machine gun section should have been located with the main body of the platoon for more effective command and control (C2). From the existing position it also has poor visibility on areas of likely contact, which also means a lack of mutual support with the other machine gun section, because they have minimally overlapping sectors of fire. In Figure 3, a significant portion of the village area is obscured from the viewpoint of the selected machine guns, while the second machine gun section has a direct view of the buildings.

While the instructor knows the terrain and distances in the scenario, it is still necessary to provide feedback to the cadet, effectively illustrating the concepts of positioning and mutual support. In this specific exercise AAR, the instructor uses the distance tool to show spacing between units, and the line of sight (LOS) function to graphically portray battlefield visibilities. The LOS tool produces an overlay that grays out areas that cannot be seen from a point on the battlefield because of elevation, terrain type, or objects.

The instructor-led AAR process involves execution playback, paused at key moments such as this one, with interactive discussion to review the basic principles (such as spacing from C2 and mutual support) using visual cues provided by the game environment. While it is true that there will always be nuanced observations by human instructors that go beyond the capabilities of automated training methods, still there are many conditions where these primary functions can be performed by intelligent software: what are the basic principles involved, how do they apply here, and what are the consequences?

AUTOMATED ASSESSMENT AND FEEDBACK IMPLEMENTATION

The tutoring system integrated with *Follow Me* is implemented as a plug-in component called Intelligent Game-based Evaluation and Review (InGEAR). Paired with the game, InGEAR is an example of a "situated tutor" (Schatz, Oakes, Folsom-Kovarik and Dolletski-Lazar, 2011), as the game offers a scenario-playing experiential environment with a wide range of decision points and free play, directly integrated with adaptive instruction. Forms of instructional adaptation are often described in terms of when instructional decisions are tailored to the learner, where macro-adaptation occurs between exercise events and micro-adaptation occurs within events (Shute, 1993). Micro-adaptation is a major component of InGEAR's pedagogical strategy, with feedback tailored to the individual's performance in a scenario with combinations of measures. InGEAR uses a constraint-based approach to implement automated assessment and feedback for the tactical decision-making principles exercised by *Follow Me*. While different student modeling strategies are suited for different forms of training, tactical decision-making concepts in a game-based setting present a unique challenge because branching is virtually unlimited. As Sottilare and Gilbert (2011) observe, the modeling task becomes complex with game environments because of the "broader range of granularity of events and stimuli" (p. 5), and the sheer quantity of additional variables involved. This is the practical motivation for a constraint-based approach, which essentially looks for noteworthy conditions that reflect significant negative or positive states that directly relate to training objectives, without trying to match every action against a certain model (Ohlsson, 1994). Although a constraint-based approach lacks traceability to cognitive states, it can offer cost benefits because the simpler student models do not need to represent all possible solutions and the sequences of cognitive states that produce them. Along these lines, Bratt (2007) similarly notes, "Making sure the student is aware of constraints on actions, and possible positive and negative effects, can be very useful, even if these constraints and causal links do not provide a ranking of optimal actions." (p. 349)

In practice, InGEAR's constraint-based implementation organizes automated assessment and feedback mechanisms in a hierarchy of individual measures including both basic tactical principles (such as spacing from C2 and mutual support) and higher level concepts associated with the overall tactical tasks involved in a scenario (such as destroy the enemy, or secure an area). The objective of this mixture of measures is to approximate the content of AAR discussion that human instructors lead with cadets. Each measure is designed based on input from subject matter experts and the instructional staff at West Point, attempting to be as general as possible about establishing the methods for assessing performance for different principles, without relying specifically on game attributes or scenario-specific assumptions. In cases where performance assessment involves the observation of thresholds (e.g., separation by a certain distance in meters, proportion of units with poor cover or concealment, times to accomplish higher level tactical tasks), the assessment mechanisms are designed for parameterized implementation without hard-coding the thresholds that define good or bad, so that these parameters can be adapted for different scenarios or instructional purposes. Each measure is specified with a unique set of instructions for the computational logic to interpret scenario execution for conditions reflecting either positive or negative performance, as well as a set of rules for the nature of training feedback. InGEAR structures assessment measures with three forms of feedback:

- Real-time feedback. Also known as immediate directive feedback, this involves prompts or cues delivered in the game environment during the execution of an exercise scenario. Several studies have demonstrated the benefits of immediate and directive feedback (Anderson, Corbett, Koedinger and Pelletier, 1995; Dihoff, Brosvic, Epstein and Cook, 2004; Jensen, Sanders, Marshall and Tasoluk, 2005; Kulik and Kulik, 1988), although each application faces the challenge to find the balance between instructional value and the potential for interrupting exercise flow. Just as an instructor overseeing a classroom exercise may withhold most feedback until the AAR, but still give comments or tips during execution when it would help, this is the purpose of the real-time feedback mechanism. In order to support a range of use cases, InGEAR divides real-time feedback into two categories for different modes of training or practice: coaching and exercising. In concept, coaching mode is for early practice of tactical skills and typically without results being used for official grading or assessment, while exercising mode requires the student to perform and results may be reported for grading. In coaching mode, real-time cues and feedback may be more frequent and explicit, with less concern for interrupting game flow or influencing player outcomes. In exercising mode, real-time feedback is scaled back for minimal intrusion, without impacting performance demands still to come. However, these modes can be used in different ways as instructors or scenario authors see fit, and the general distinction is provided to allow for two different levels of feedback for each assessment measure, again using parameters to establish the differences in timing or content of feedback.

- AAR playback. The system-guided AAR playback is directly analogous to how instructors currently use playback for feedback. In order to enhance the playback, bookmarks are created as a result of the real-time assessment results during execution, along with any post-facto synthesized knowledge about significant times in an exercise. Each bookmark is associated with one or more assessment measures, which are defined with templates for how the corresponding tactical principles can be illustrated. A learner engaged in self-directed training can either let the entire playback run with pauses at each bookmark, step through the bookmarks by iterating through the list, or go directly to the bookmarks that are of the most interest.

Figure 4 shows the playback controls and list of named bookmarks generated from an exercise. In AAR playback, events are paused at each bookmark, as an opportunity to give feedback that illustrates the tactical concept and how it applies in the current situation.

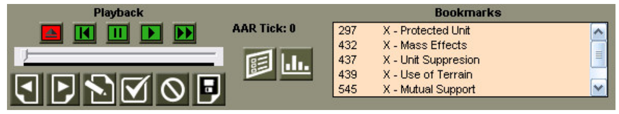

Figure 4. Playback Controls and Bookmarks

Based on the template for the associated assessment measure, feedback may include text, highlighting of one or more units, visual effects presented as overlays on the tactical view (such as the visibilities and sectors of fire shown in Figure 3), and even multimedia. Also the text may be declarative or interrogatory, where in the latter case the student may answer a multiple choice question or provide a free-text answer which will be saved into the AAR log. The explanatory information for a playback bookmark is only shown while initially paused, and then disappears when resuming playback or skipping to another bookmark.

- AAR summary and analysis. In addition to the AAR feedback provided directly via playback, a second category of AAR feedback involves an overall summary and analysis of exercise performance and various statistics. Information about unit status over time is presented visually in graph form (Figure 5), where the intention is to convey higher level observations about key points in the exercise that may not necessarily be tied to individual performance measures. For example, a graph of massed effects and lethality on the same timeline with unit status, ammunition levels, fatigue, and other related measures helps to provide insight on the tactical situation in a way that was not previously available to instructors and cadets.

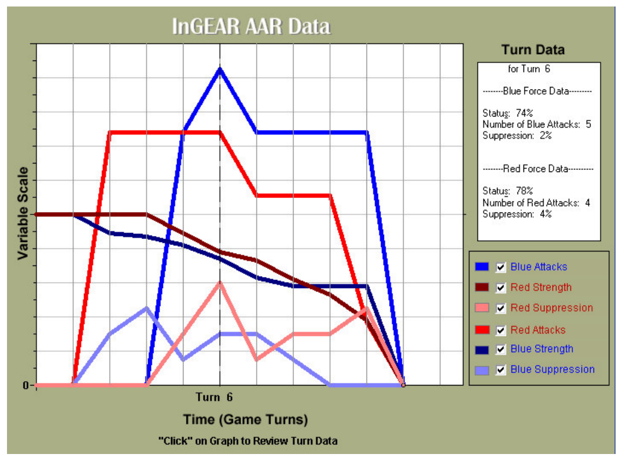

Figure 5. Unit Status Graph in AAR Summary

The design of the assessment measures resulted in a specification that defines the nature of the feedback to be delivered for each of the categories above. Referring back to the earlier examples of C2 spacing and mutual support described in the context of instructor-led training, these each appear in the InGEAR implementation as unique assessment measures. Real-time feedback for each of these measures is only provided in coaching mode, where it is useful to give immediate feedback about what is wrong with the relative positions of units. There is not feedback for these measures in exercising mode, because this feedback during execution would have the potential to significantly impact overall outcomes, where those consequences can also be an important lesson worth preserving. Both of these assessments generate corresponding playback bookmarks for AAR, illustrating concepts in context with the same in-game visual effects that instructors use (e.g., Figure 3). For the AAR summary report, an overall performance assessment is given for each of these measures, based on how often problem conditions occurred in the exercise.

A more complex training example involves friendly fire incidents and fratricide. Avoiding fratricide is one of the foremost tactical principles not only for platoon level operations, but also for nearly every echelon and operational setting of kinetic live-fire warfare. While outcomes of concern may range from firing in the area of friendlies, to firing on friendlies, to causing casualties or actual fratricide, the underlying concepts in any of these conditions involve situational awareness, understanding weapon effects, and coordination of maneuver and fires. Thus from a diagnostic perspective, there are several different conditions for the fratricide-related assessments to detect, potentially with more than one occurring in the same engagement event. For example in the simplest case, a fratricide incident occurring in the game environment will typically include separate states where fire on friendlies precedes the actual fratricide event. However, the converse is not true; fire on friendlies may sometimes occur without leading to fratricide. Similarly, the order of events can provide important information about the underlying decision-making failures that led to a fratricide.

Figure 6 and Figure 7 show a simplified set of scenario conditions where two different sequences of events lead to the same fratricide event outcome, but with different pre-conditions that are important to capture and reflect in training feedback.

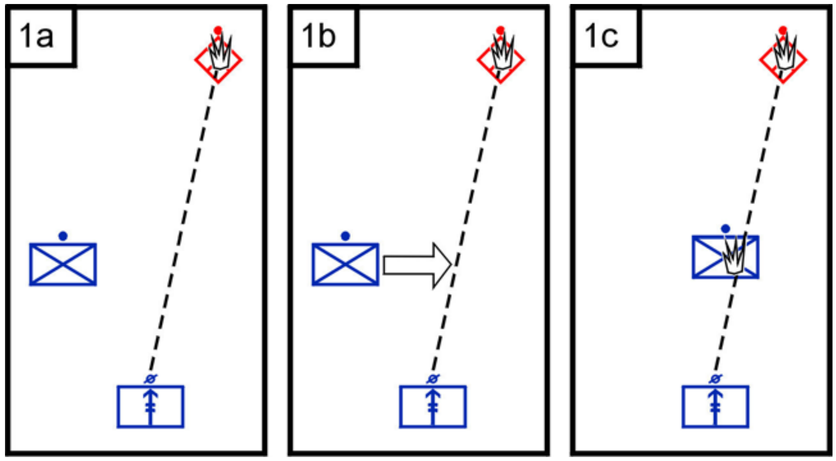

Figure 6. Fratricide Event Resulting from Friendly Unit Movement into Line of Fire

In the above sequence of events, the Blue machine gun team is actively engaged on an enemy target (time 1a), when another Blue unit moves into the active line of fire (1b), leading to a fratricide event (1c). In the game environment, there are two relevant operator commands: the engagement command on the machine gun team with the specified target, and the movement command on the second Blue unit. Also, implicitly, if the operator had an intentional tactical reason to move the second Blue unit, then the failure to stop the machine gun engagement is also significant.

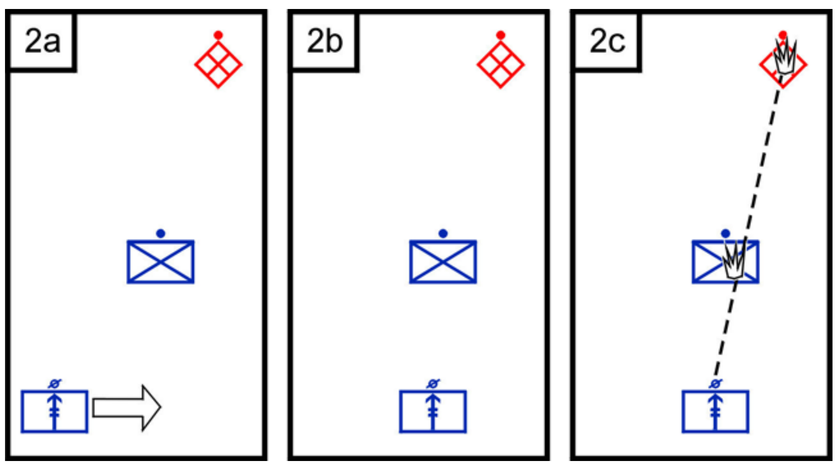

Figure 7. Fratricide Event Resulting from Line of Fire across Friendlies

In this second sequence of events shown in Figure 7, the main factor in the fratricide is the control of the machine gun team. Initially there are no fire events as the machine guns move (2a) and then take a position (2b). But then the machine guns engage an enemy target despite the fact that the line of fire crosses another Blue unit, leading to the identical fratricide event outcome (2c) as in the previous example (1c). In this case the significant commands are the movement of the machine guns, and more importantly the engagement command issued to the machine guns with a line of fire passing over friendlies to the target. Implicitly the failure to reposition either the machine guns or the other Blue unit before engagement may also be a factor.

Although the depictions above are intentionally simplified, these kinds of conditions do occur, especially in the more complex scenario environment where terrain features, buildings, and the endless variety of relative unit positions come into play. In both cases above, all of the significant commands are reflective of decisions that should be factored into the feedback. Even in a constraint-based approach where nominally the function of automated assessment is to detect states, it's important in a tactical decision-making domain to make sure that the notion of significant states is not limited entirely to outcome conditions like (1c) and (2c) above. By including prior actions and states in the assessment logic, the resulting feedback can effectively draw connections to decision-making.

In the InGEAR implementation, there is also some processing in both real-time and the preparation of AAR, to consolidate feedback for related or repeated conditions. For example, the engagement in which a fratricide occurs may involve many successive discrete fire events, but the objective for real-time feedback is to provide input exactly one time per incident, where an incident is defined by the fire command (using Engage or Assault in the game) and the unit receiving friendly fire. Likewise for AAR playback, the choice of the appropriate time for a bookmark is based on the nature of the incident. For example, in the events leading to fratricide in the first example above (Figure 6), it is ideal to create a playback bookmark at time (1a) or (1b), so that the viewer can see the situation and how the fratricide occurred. Some of this information would be lost, or at least require additional rewinding steps to see, if the bookmark were simply placed at time (1c) when the actual fratricide occurs. This also helps to make the causes apparent and reduce the dependence on textual feedback explaining what's happening. The visual cues automatically generated for a playback bookmark also help to supplement the textual prompt. Figure 8 below shows a close-up view from an AAR playback, where the friendly unit receiving fire is highlighted with impact graphics and a solid bounding square, and the active lines of fire are shown as well.

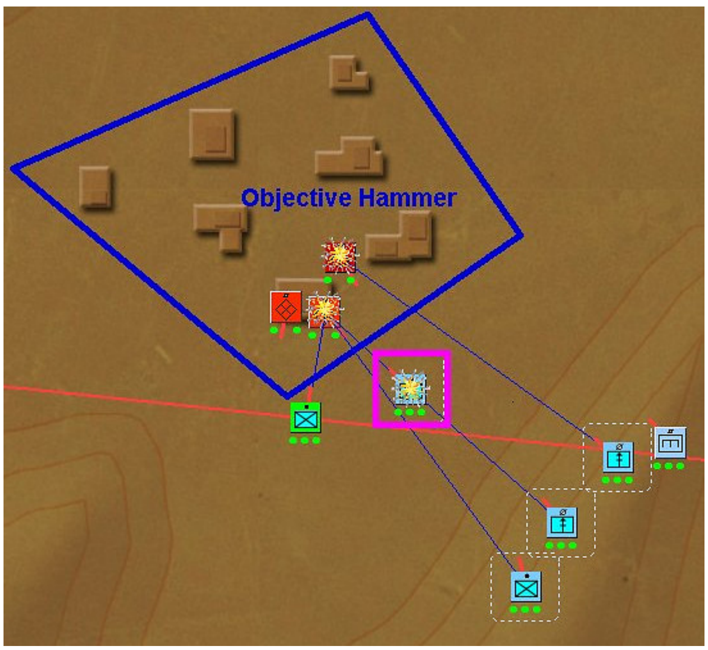

Figure 8. AAR Playback Bookmark for Fratricide Event Illustrated with Visual Effects

AUTHORABILITY AND GENERALIZATION

The InGEAR module integrated with Follow Me is an operational game-based situated tutor, instantiated with a set of small unit tactical decision-making assessment measures, but it is constructed as a mostly custom, tailored application. Echoing the earlier discussion, custom development needs are considered in some circles to be an obstacle to wider adoption of this kind of technology. Thus, toward the goal of exploring generalization and reuse, this section offers observations about parallels between InGEAR and current emerging approaches for simplifying development tasks. For example, the US Army Research Labs' Generalized Intelligent Framework for Tutoring (GIFT) (Sottilare, Goldberg, Brawner and Holden, 2012), is an emerging architecture offering a collection of tools, methods, and standards to facilitate a spectrum of computer-based tutoring systems. Although the GIFT framework targets a much wider space of instructional development than the specific use case for the InGEAR system, there are a number of noteworthy concepts that apply to this class of game-based trainer, such as the idea of introducing a standardized structure for instructional data exchange, using a common instructional data model.

Similar to how virtual environments and some "serious games" have been increasingly developed or adapted to support standardized event models for discrete virtual states and events, a standard protocol and/or model for instructional data could have a direct impact for modularity and reuse. In order to explore how this might be pursued in practice, consider the instructional goals related to the tactical concept of avoiding fratricide, as discussed earlier with the InGEAR implementation. The risk of fratricide applies in many military operational domains, certainly not limited to small unit ground missions, as described in the context of the Follow Me game and the InGEAR assessments. Suppose we go one step further in abstraction, attempting to outline general features of fratricide and the related tactical concepts, for the purposes of defining a model for the information needed for instructional assessment. The basic features of interest for a fratricide event, generalized for nearly any operational context or experiential scenario environment, might be proposed as the following list:

1. Who was the shooter?
2. Who was the victim?
3. Are the shooter and victim on the same side?
4. How severe was the incident? (This might range from the victim being inside of a surface danger zone or ricochet area without being hit, to experiencing a degree of weapons effects, to catastrophic effects.)
5. What commands were actually given and when?
6. Was the fratricide risk present when the fire command / authorization was given? Or did the situation change after the command was issued?
7. What was the timeline for the entire incident (i.e., a spontaneous accident or longer incident)?
8. How are the shooter and victim separated in the command structure? Are they controlled by different commanders?
9. Specifically what decision-makers were involved? Who should have known about the victim's position or movement (and by what chain of control)? Who should have given the fire command?

A generalized assessment model could then map combinations of results for these data features, to a delineated set of instructionally significant conditions. The theoretical value in this kind of abstraction is that each future application is essentially an instantiation with a specific set of training objectives associated with a domain and experiential environment. For the InGEAR implementation with *Follow Me,* features (1) – (7) are all applicable, and in fact the assessment model directly uses features like (6) to differentiate between feedback conditions for the two examples illustrated in Figure 6 and Figure 7. The idea is that this logic is entirely reusable. For InGEAR, (8) and (9) do not apply, at least not in the current implementation as an individual trainer where all decisions are assumed to be handled by the same operator. In a team training domain like combined arms, all 9 features may be significant, and important to tease apart in an adaptive trainer. The value is in the ability to reuse the underlying logic that defines how different conditions reflect different nuances of performance for the same tactical principle.

For application with a given virtual environment, the value of a standardized instructional data model is that it helps structure the development of the interface between components. Whether the task is for game developers to publish new data or messages, or for training system developers to create a custom layer to synthesize existing data into the designated features, the existence of the data model helps establish structure. Similarly for instructional designers, once there is a formalism that maps features to instructionally significant conditions, this provides a headstart for the next task of defining the pedagogical strategy that will be applied in these different conditions, including the nature of feedback, scenario interventions, AAR, and macro-adaptations in exercise selection. This also suggests the potential to reuse pedagogical strategies across domains, with reduced development work for adaptation.

Another benefit to a modularized architectural approach, with a separate component for the interface layer controlling data exchange with the virtual environment, has to do with reuse across related virtual environments. Many games are implemented as families based on a common engine and wrapped with specific adaptations or "mods" to fit different narratives, scenarios, levels, performance criteria, playing modes, etc. While these kinds of variations take many forms, one specific example directly related to InGEAR is that *Follow Me* is based on the same engine as the game *Crucible of Command*, which is used at the US Army Command General's Staff College to teach tactical decision-making concepts in more advanced scenarios than cadets see at West Point. So in this case, the interface layer implemented for *Follow Me* is almost directly reusable for a potential adaptation of the InGEAR capability to *Crucible of Command.*

CONCLUSION

The InGEAR automated assessment and feedback capabilities have been delivered to West Point, and the immediate next step is to extend scenario authoring mechanisms already provided with the *Follow Me* game, to allow instructors and scenario designers to control the instructional behavior in scenarios as well. Although no formal evaluation of learning outcomes has been conducted to evaluate learning in the self-directed training use case, the frequency and quantity of cadets using the system presents an attractive opportunity to gain further insight about practical results with this kind of technology.

ACKNOWLEDGEMENTS

The InGEAR effort described in this paper was sponsored by the US Army Research Institute. The views expressed in this paper are those of the authors and do not necessarily represent the official policy or position of the Department of Defense, the US Army, the US Army Research Institute, or the US Military Academy at West Point.

REFERENCES

Anderson, J., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4, 167-207.

Bratt, E. O. (2009). Intelligent Tutoring for Ill-Defined Domains in Military Simulation-Based Training. *International Journal of Artificial Intelligence in Education,* 19, 337-356.

Chipman, P., Olney, A., & Graesser, A. C. (2005). The AutoTutor 3 architecture: A software architecture for an expandable, high-availability ITS. *Proceedings of WEBIST 2005: First International Conference on Web Information Systems and Technologies.* Portugal: INSTICC Press, 466-473.

Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: learning is enhanced by immediate but not delayed feedback. *The Psychological Record* 54: 207-231.

Ericsson, K. A., & Ward, P. (2007). Capturing the Naturally Occurring Superior Performance of Experts in the Laboratory: Toward a Science of Expert and Exceptional Performance. *Current Directions in Psychological Science,* 16(6), 346-350.

Folsom-Kovarik, J. T., & Schatz, S. (2011). Return on investment: A practical review of ITS student modeling techniques. *M&S Journal, Winter Edition,* 22-37.

Jensen, R., Sanders, M., Marshall, H., & Tasoluk, C. (2005). FCS Intelligent Structured Training – Experimental Results and Future Applications. *Proceedings of the Interservice / Industry Training, Simulation, and Education Conference (I/ITSEC 2005).*

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research* 58(1): 79-97.

Ohlsson, S. (1994). Constraint-Based Student Modeling. In J. E. Greer & G. McCalla, eds. *Student Modelling: The Key to Individualized Knowledge-based Instruction.* Berlin / Heidelberg: Springer-Verlag, 167-189.

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine* 78(5): B77-B85 Suppl. S.

Schatz, S., Oakes, C., Folsom-Kovarik, J. T., & Dolletski-Lazar, R. (2012). ITS + SBT: A review of operational situated tutors. *Military Psychology* 24(2), 166-193.

Shute, V. J. (1993). A macroadaptive approach to tutoring. *Journal of Artificial Intelligence in Education,* 4(1), 61-93.

Sottilare, R. A., & Gilbert, S. (2011). Considerations for Adaptive Tutoring within Serious Games: Authoring Cognitive Models and Game Interfaces. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED) 2011*.

Sottilare, R. A., Goldberg, B. S., Brawner, K. W., & Holden, H. K. (2012). A Modular Framework to Support the Authoring and Assessment of Adaptive Computer-Based Tutoring System (CBTS). *Proceedings of the Interservice / Industry Training, Simulation, and Education Conference (I/ITSEC 2012).*