DEFINING SIMULATED TEAM MEMBERS: GUIDELINES BASED ON AN EMPIRICAL APPROACH

Alma M. Schaafstal, Ph.D.[1]
TNO Human Factors
3769 ZG SOESTERBERG, The Netherlands

Denise M. Lyons, Ph.D.
Naval Air Warfare Center Training Systems Division, AIR 4962
Orlando, Florida

ABSTRACT

Team training increasingly takes place in synthetic environments. However, team training in synthetic environments is often modeled after live team training without removing some of the disadvantages that occur in live training, such as instructor-intense performance monitoring, and the fact that all appropriate teammates have to be available. Simulated teammates are a promising alternative to human teammates, because they are always available, may be modeled after experienced training personnel, and may be more cost effective.

The Netherlands Organization for Applied Scientific Research (TNO) Human Factors and the Naval Air Warfare Center Training Systems Division are working jointly towards defining the requirements for synthetic teammates (SYNTHERs). The goals of this research effort are twofold: (1) to define the requirements for SYNTHERS and (2) to develop validated guidelines for the use of SYNTHERS in team training. In our approach to empirical validation of requirements a set of psychological experiments will be carried out, utilizing scripted humans as simulated teammates in a well-controlled simulation of a military command-and-control task using a modified version of the Dynamic Distributed Decision-Making (DDD), while taking a variety of measurements. Two experiments have been conducted so far. This paper relates the results of those experiments to an empirical validation of requirements, and provides guidelines for the design and use of SYTNHERS for team training.

ABOUT THE AUTHORS

Alma Schaafstal studied Cognitive Psychology and Philosophy at the University of Groningen, The Netherlands, and the University of Manchester, UK. She wrote her dissertation (1991) on expert-novice differences in diagnostic skill in papermaking. In 1985-1986 she spent a year as a visiting scientist at Carnegie Mellon University, Pittsburgh, PA. She has been with TNO, the Dutch Organization for Applied Scientific Research since 1986, and has been affiliated with TNO Human Factors since 1990, specializing in training complex skills in military environments with a strong focus on team training. From March 2000 until March 2001, as part of an exchange program between the US and The Netherlands, she worked at the Naval Air Warfare Center Training Systems Division in Orlando, FL, focusing on training teams with simulated teammates.

Denise Lyons is a Senior Research Engineer in the Science and Technology Division of the Naval Air Warfare Center Training Systems Division. Her current focus includes modeling and simulation, and training technology for programs in Improved Behavioral Representation of Computer Generated Forces (CGF), and Advanced Embedded Training Systems (AETS). She received her Ph.D. in Optical Sciences from the University of Arizona, and has over 13 years of experience in research and development for military systems.

---

[1] A large part of this research was carried out while Alma Schaafstal was a visiting exchange scientist, sponsored by DoD, at the Naval Air Warfare Center Training Systems Division in Orlando, FL.

DEFINING SIMULATED TEAM MEMBERS: GUIDELINES BASED ON AN EMPIRICAL APPROACH

Alma M. Schaafstal, Ph.D.[2]
TNO Human Factors
3769 ZG SOESTERBERG, The Netherlands

Denise M. Lyons, Ph.D.
Naval Air Warfare Center Training Systems Division, AIR 4962
Orlando, Florida

INTRODUCTION

Team training increasingly takes place in synthetic environments. However, team training in synthetic environments is often modeled after live team training without removing some of the disadvantages that occur in live training, such as instructor-intense performance monitoring, and the fact that all appropriate teammates have to be available. Simulated teammates are a promising alternative to human teammates, because they are always available, may be modeled after experienced training personnel, and may be more cost effective.

The Netherlands Organization for Applied Scientific Research (TNO) Human Factors and the Naval Air Warfare Center Training Systems Division (NAWCTSD) are jointly working towards defining the requirement of these synthetic teammates (referred to here as SYNTHERS) (Schaafstal, Lyons, & Stroomer, 2000; Schaafstal & Lyons, 2001). The goals of this research effort are twofold: (1) to define the requirements for SYNTHERS, and (2) to develop validated guidelines for the use of SYNTHERS in team training. Without a thorough validation process, requirements can be put forward which do not contribute substantially to the learning process. Thus, validation of requirements is essential to the definition of cost-effective SYNTHERS. This paper will report the progress that we have made so far, both in terms of defining requirements, and in terms of the validation process.

REQUIREMENTS FOR SYNTHERS

Based upon a conceptual framework applied to modeling in the Anti Air Warfare (AAW) domain, we came up with a number of requirements for (ideal) SYNTHERS (Schaafstal, Lyons, & Stroomer, 2000). These include:

a) The ability to perform the task at the taskwork level, from the perspective of the individual team member. What this implies will vary from task to task.

b) The ability to interact and communicate with other members of the team: receiving and interpreting visual and auditory information and acting appropriately on the basis of this information. This implies that the SYNTHERS have the capability of communication.

c) The ability to flexibly and intelligently adapt to a changing scenario.

d) The ability to interact and communicate with both humans and other simulated team members.

e) The ability to behave as a believable, worthy teammate, instructor, coach or combination thereof.

f) The ability to demonstrate both taskwork and teamwork skills.

g) The ability to analyze trainee performance and provide appropriate feedback, possibly coupled with instructional strategies. This is important in those instances in which the role of the SYNTHER is that of the instructor/coach.

---

[2] A large part of this research was carried out while Alma Schaafstal was a visiting exchange scientist, sponsored by DoD, at the Naval Air Warfare Center Training Systems Division in Orlando, FL.

h) The ability to diagnose trainee performance. This is important, for example, in those cases where SYNTHERS will provide opportunities for practicing teamwork skills such as error correction and providing/requesting backup. That is, providing opportunities for practicing those skills relies on an extensive diagnosis of the trainees behavior (i.e., knowing and deciding when and how to interrupt the ongoing behavior). However, this argument does not apply to training all of the teamwork skills through SYNTHERS, e.g., using proper communication styles, situational updates, etc., relies on a much more superficial level of performance diagnosis.

An important implication of this framework is that instead of being experts in the task, SYNTHERS should provide opportunities for training. They should provide opportunities for practice, opportunities that may be hard to obtain in the real world.

Finally, the requirements for SYNTHERS depend upon the training goals and upon their roles in a scenario. If, for example, the SYNTHER acts as a fellow teammate for practicing teamwork skills without any coaching tasks, then its requirements will be different compared to a SYNTHER that is acting as a coach and is providing opportunities for error correction and/or commenting on the outcome of an event.

EMPIRICAL VALIDATION OF REQUIREMENTS

In our approach to empirically validating requirements for SYNTHERS, a set of psychological experiments will be carried out, utilizing scripted humans as SYNTHERS.

Two experiments have been conducted so far in which teams that are trained on teamwork skills, with or without confederates, are subsequently tested on their teamwork skills while working with different confederates. The task of the confederates is to systematically provide opportunities for demonstrating the acquired teamwork skills. Since the confederates are scripted to provide opportunities for practicing and demonstrating teamwork skills, and since they exhibit standardized behavior, the resulting variance in the data is mainly due to the variance in teamwork skills of the trainees as a result of the experimental manipulations. In the first experiment, which used two-person teams working with the Tactical Navy Decision Making System (TANDEM) testbed (see Johnston, Poirier, & Smith-Jentsch, 1998), SYNTHERS (confederates) provided opportunities for practicing teamwork skills, as well as guidance and feedback to the participant on both performance and underlying processes. The results showed that it is possible to script teammates in such a manner that their behavior will remain constant across teams The results also indicated that training with SYNTHERS leads to an improvement in teamwork skills (Stroomer, Timmer & Schaafstal, in 2001, in preparation). In a second experiment, we manipulated the behavior of the SYNTHERS. In this case, we focused on providing opportunities for practicing teamwork skills only, attempting to separate the effects of guidance and feedback from the effects of providing opportunities for practice.

THE DDD-TESTBED

The testbed that we used in the second experiment was a modified version of the Dynamic Distributed Decision-Making (DDD) simulation. This computerized task (see Figure 1) simulates a

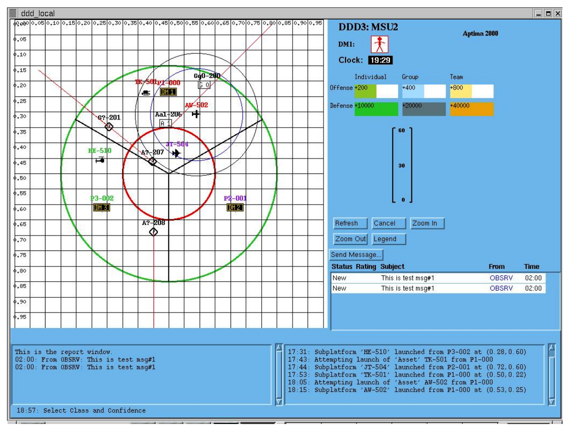

Figure 1. The DDD testbed

military command and control task where a team of decision-makers, each of whom controls specialized sub-platforms, needs to enforce a "demilitarized zone." This demilitarized zone is threatened by an enemy, but is also traversed by friendly ground vehicles and aircraft. The goal of the team is to first identify the nature of all vehicles and aircraft (i.e., friendly versus enemy) that encroach or enter the demilitarized zone, and then attack enemy targets that enter this zone, while not harming any friendly vehicles or aircraft. The DDD simulation was initially developed by Serfaty and Kleinman (1986), and has been used in a variety of contexts, examining a variety of research questions, including optimizing team organization, optimizing team decision making, and team training.

Through the DDD, the experimenter can manipulate the number of team members involved, the layout of the area, and the allocation of sensors and weapons. The DDD offers maximal control over the timing, heading, path and number of targets approaching the demilitarized zone. It also offers numerous opportunities for the experimenter to automatically measure team performance on a number of dimensions.

In this experiment, we used a three-person team adaptation of the DDD for the experimental group. The team consisted of two confederates and one subject, with the subject trained on teamwork skills. The subject was not aware of the fact that the two other team members were confederates. The confederates acted according to a tight time, and event-based, script. A control group, in which three naïve subjects were trained, was used as a comparison group. Although the DDD can be used with military personnel, we used university undergraduates as participants. This allowed us to carry out the experimentation in a relatively limited amount of time.

A variety of measures were taken and analyzed in this experiment. First, performance measures were recorded automatically through the DDD simulation, e.g., number of targets correctly identified and individual and team scores. Measures of coordination were also recorded, including number of sub-platforms transferred to other team members, how often identifications of targets were shared, and the efficiency of attacks. Second, analysis of video and audio materials collected during the experiments, relative to the teamwork skills exhibited, was also conducted. Third, data from several questionnaires was collected and analyzed. These questionnaires were administered before, during, and after the experiments, and related to a variety of topics such as:

- Collective efficacy (the confidence the subject has that this team is able to do the task);
- Self-efficacy (the confidence the subject has in doing any task);
- Task-self efficacy (the confidence the subject has in doing this particular task);
- Familiarity with working in teams (how familiar the subject is already with working in teams);
- Collective orientation (whether someone is oriented towards teamwork or would rather work on his/her own);
- Collaboration with teammates (how the collaboration with the other teammates worked out, which is of particular importance for evaluating the requirements of synthetic teammates).

Using confederates in combination with the DDD testbed, we were able to rapidly define what the requirements of the SYNTHERS should be, and what the relative contributions of certain components of the SYNTHER add to the training outcome, both quantitatively and qualitatively. The use of a number of different measures adds to the strength of the approach.

A total of 34 teams participated in this research (16 in the control condition, and 18 in the experimental condition). The results indicate that people that have been trained with SYNTHERS have better teamwork skills than people that have been trained with naïve teammates (Schaafstal, Lyons, & Smith-Jentsch, 2001, in preparation). The results also show that the variance in teams is reduced when working with SYNTHERS, which indicates that training with SYNTHERS results in more structured and standardized training than would otherwise occur.

The data collected in this experiment can be contrasted with data that we collected from a subsequent experiment in which experts were used to perform the same task. These experts each had more than 70 hours of experience in playing this version of the DDD, but did not have experience in working with each other (they were confederates in the previous experiment). Their task was to maximize their team score. The results of this experiment show that the experts, on average, scored about 25% higher on each scenario then the best teams in the previous experiment. This suggests that in order to be good training partners, SYNTHERS do not have to perform at an expert-level. In fact the confederates, when playing by themselves, obtained a much higher score then when playing as confederates.

RESULTING GUIDELINES FOR THE DESIGN AND USE OF SYNTHERS

In recent years, it has been shown that a good approach to training teams with complex training technology is linking training goals to events in training scenarios in a controlled fashion. This is called the Event-Based Approach to Training (EBAT) (Dwyer, Oser, Salas, & Fowlkes (1999)). The primary goal of EBAT is to systematically provide opportunities for a training audience to develop critical competencies by receiving practice in simulated environments that are representative of actual operational conditions, and receiving feedback linked to specific events that occur during training. EBAT tightly links trainee needs, critical tasks, learning objectives, scenario design, performance measurement, and feedback. The general assumption of EBAT is that without a systematic linkage among these components there is no way of knowing or ensuring, with any degree of certainty, that the training exercise will have its intended effect. Figure 2 shows how the EBAT framework supports the design, development, execution, and evaluation of exercises.

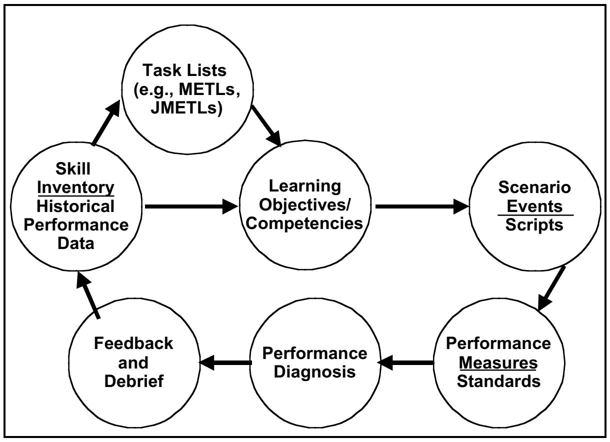

Figure 2. Components of an EBAT Framework. Adapted from Zachary, Cannon-Bowers, Burns, Bilazarian, & Krecker (1998).

The EBAT-approach provides a good starting point for the development of a conceptual framework for training teams with synthetic teammates. It shows us where synthetic teammates could play a role in training teams: they could act as fellow teammates, as collateral forces, or as enemy forces. Second, it shows that, as much as scenario events are scripted, SYNTHERS acting as teammates should behave in a scripted manner, too. This does not mean that their behavior has to be strictly 'canned' and predictable for the trainee, but it does imply that their behavior is predictable for the scenario designer.

Our two studies, in combination with previous research, indicate that at least three design components can be distinguished for SYNTHERS acting as teammates:

1. A task component: the level of task proficiency the SYNTHER should have. Analysis of our data indicates that SYNTHERS do not necessarily have to be experts in the task, but should master the task to the degree that they are believable and useful teammates in the context of a team training environment. This will probably involve some level of both taskwork (the knowledge and skills involved in carrying out the individual task) and teamwork (the knowledge and skills involved in cooperating with others) (Zachary & Le Mentec, 2000).

   An analogy from sports will help to clarify this point. Tennis instructors, practicing or playing with you from the other side of the net, do not necessarily have to be expert players to provide you with the right shot to practice your skills. Along similar lines, SYNTHERS do not necessarily have to be at an expert-level of proficiency in the task to provide valuable instruction and practice opportunities. In fact, our data show that our confederates were not playing at an expert-level of proficiency. When they were allowed to play the DDD in a free-play situation, they performed better than when they were playing the confederate role.

   It is impossible to script the behavior of the SYNTHERS for every target in the scenario. This type of scripting would result in very unrealistic behavior with the same "brittleness" that is criticized in current Computer Generated Forces (CGFs). Therefore, we decided to define a set of "standard behaviors." For the DDD testbed, we defined the task component as:

   - The SYNTHERS staying in a predefined area of the screen, unless called upon by the trainee or by the script.
   - The SYNTHER playing the game in a normal fashion: detecting, identifying, and engaging targets that would appear in their area.
   - The SYNTHERS not communicating unless they had to (when there was a target in their area for which they lacked the power to attack), or when they were asked a specific question.
   - The SYNTHERS obeying the task requests as much as possible from the trainee.
   - The SYNTHERS following a time-based script, in which they would take certain actions (verbal or game-based) upon certain targets or their sensor and weapon systems. The events in the scripts are related to the instructional/practice component described in the next paragraph.

2. An instructional-practice component: what are appropriate instructional and practice strategies for the SYNTHER to exhibit? One of the most important features and potential benefits of training with SYNTHERS is that they can demonstrate certain behaviors in training and provide systematic and structured opportunities for practice. For example, providing backup to other team members in case they become overloaded is regarded as an important teamwork skill (Smith-Jentsch, Johnson, & Payne, 1998). In situations where one is being trained in an expert team, however, team members are not easily overloaded and the opportunities for practicing this teamwork skill are rare. However in training with a SYNTHER these opportunities can be presented, because SYNTHERS can be scripted to become overloaded through a

combination of external events and the action of the SYNTHER to the overload situation. Therefore, training with SYNTHERS provides the trainee with excellent opportunities for practice that might not otherwise exist. This is supported by the results of our experiments.

The events in the scripts that would result in opportunities for practice were based on a cognitive task analysis of supporting behavior in the DDD. In our experiments, we focused on the training of supporting behavior, which consists of three components:

- Providing backup to other teammates when they become overloaded.
- Requesting backup in situations in which you become overloaded.
- Catching and correcting errors before they have negative consequences on the teams' performance.

A cognitive task analysis was used to examine when and how these three components would show up in the DDD. For example, by carefully observing both expert and novice players we observed that the following errors are "natural errors" in the DDD:

- Attacking an enemy in the neutral zone.
- Relying on incorrect identification data and therefore mistakenly attacking targets.
- Prioritization errors, i.e., giving priority to targets which are less threatening than other targets, since they are further removed from a predefined no-fly zone.
- Attacking with a weapon system which is out of ammunition.

The scripting of these "natural errors," in combination with the "standardized behavior" as described in the previous paragraph, resulted in believable SYNTHER behavior. The data indicate that no participants questioned the realism of the SYNTHERS. There is also informal evidence from the sessions in the experiment themselves that participants were not aware of the fact that they were playing with SYNTHERS, and not with other "naïve" participants. This can be taken as evidence that SYNTHERS in this context do not have to perform at an expert level, and can be scripted without being predictable or believable for the trainee.

3. A performance measurement/feedback component: Our research has led us to question whether it will be possible to measure teamwork skills on-line and automatically in dynamic environments. The results show that this is possible to some extent, at least in the environment that we are currently working in, i.e., the DDD simulation. The generalizability of these results to real-life military environments is an issue that still needs to be addressed.

IMPLICATIONS

Based on our research results, we propose a layered model of scenario events for defining the behavior of synthetic teammates in team training. The first layer consists of the 'normal' scenario events that occur, i.e., the behavior of entities external to the team, such as a target approaching with a certain speed at a certain altitude.

A second layer consists of the behavior of the synthetic teammate(s) and attempts to answer the question "How should synthetic teammates behave during these events?" For example, if the training goal is to learn about providing backup to another team member, the training exercise must create an opportunity to provide backup. In the DDD this can be done by sending a wave of targets to one simulated team member, too many to act upon by him/herself, and have this teammate ask for help. However, this opportunity for providing backup might not be noticed by the trainee if he/she is overloaded as well. Therefore, we have to manipulate the scenario to ensure that our trainee is not overloaded. Finally, we can make the cue for providing backup even stronger if it is clear to the trainee that the third teammate on the team cannot help out; for example, because most of his/her weapon or sensor systems have run out of ammunition or fuel.

A third layer consists of training strategies linked to team performance measurements and providing on-line feedback.

All three layers are connected by their relationship to training objectives/competencies, but they will each contribute to achieving these goals in different ways. The first layer defines events external to the team. The second layer defines when and how the synthetic teammates will provide opportunities for training certain (teamwork) skills. The third layer defines how intelligent performance evaluation and on-line feedback could be provided by a synthetic teammate. The three layers can be linked in various ways: by time (when does a certain event at each of the three layers occur), by the contingency of events (e.g., the synthetic teammate will provide an opportunity for practicing error correction at specific targets), or by a combination of both.

By redefining SYNTHERS as entities that serve certain training goals, their role is not necessarily one of being a "normal", proficient teammate.

Therefore, the mission of synthetic teammates could be to provide opportunities for practicing certain skills, instead of being expert performers in the task itself. It is important to make the trainee the center of attention, and not the SYNTHER, since the latter's goal is to help train the trainee.

THE ROAD AHEAD

Our research focused upon empirically validated requirements for SYNTHERS in team training. The initial driving force behind this research was the concept of using SYNTHERS to fill the roles of teammates that could otherwise not be filled during a team training exercise. However, in the course of the research it became clear that SYNTHERS have more to offer. They can be assigned various roles: as teammates, as collateral forces, or as opposing forces. More interestingly, they can be assigned a role in the training process itself. They can act as coaches, they can provide opportunities for practicing certain skills, and they can help to diagnose, evaluate, and remediate trainees' knowledge and skill. Through the integration and fusion of SYNTHERS as Semi-Automated Forces (SAFs) and as Intelligent Tutoring Systems (ITS), we will be able to explore a new dimension of learning companions (Chan & Baskin, 2000) and pedagogical agents (Rickel & Johnson, 1999).

REFERENCES

Chan, T-W, & Baskin, A.B. (2000). Learning companion systems. In C. Frasson, & G.Gauthier (Eds.), *Intelligent tutoring systems: At the crossroads of artificial intelligence and education*.

Dwyer, D. J., Oser, R. L., Salas, E., & Fowlkes, J. E. (1999). Performance measurement in distributed environments: Initial results and implications for training. *Military Psychology, 11*, 189-215.

Johnston, J. H., Poirier, J., & Smith-Jentsch, K. A. (1998). Decision making under stress: Creating a research methodology. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 39-60). Washington, DC: American Psychological Association.

Rickel, J., & Johnson, W.L. (1999) Virtual humans for team training in virtual reality. *Proceedings of the Ninth International Conference on AI in Education, July 1999*, 578-585.

Schaafstal, A.M., & Lyons, D.M. (2001). Training teams with simulated team members: Defining requirements. *ITEC 2001, Lille, France*.

Schaafstal, A.M., Lyons, D.M., & Stroomer, S.M. (2000). Training teams with simulated teammates. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference 2000 (I/ITSEC 2000), Orlando, FL*, 1325-1331.

Schaafstal, A.M., Lyons, D.M., & Smith-Jentsch, K.A. (2001, in preparation). Providing opportunities to train teamwork skills: An empirical investigation.

Smith-Jentsch, K. A., Johnson, J. H., & Payne, S. (1998). Measuring team-related expertise in complex environments. In J. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training (pp. 61-88)*. Washington, DC: American Psychological Association.

Stroomer, S.M., Timmer, A., & Schaafstal, A.M. (2001, in preparation). Proving the concept of training teamwork skills with simulated teammates. TNO Human Factors, *Technical report*.

Zachary, W., Cannon-Bowers, J. A., Burns, J., Bilazarian, P., & Krecker, D. (1998). An advanced embedded training systems (AETS) for tactical team training. Paper presented at the *Fourth International Conference on Intelligent Tutoring Systems, San Antonio, TX*, *August 1998*.

Zachary, W. & Le Mentec, J.C. (2000). Modeling and simulating cooperation and teamwork. *Proceedings of the Advanced Simulation Technology Conference*.