THE USE OF MASSIVE MULTI-PLAYER GAMING TECHNOLOGY FOR MILITARY TRAINING: A PRELIMINARY EVALUATION

Shawn A. Weil[1], Talib Hussain[2], Tad Brunyé[1], Jason Sidman[1], Lisa Spahr[1]

[1]Aptima, Inc.,
Woburn, MA

[2]BBN Technologies
Cambridge, MA

Publicly available Massive Multi-Player Games (MMPG) allow multiple individuals to work together in simulated situations. To meet complex game objectives, users must exhibit high degrees of coordination. This is similar to the types of interactions required for effective coordination in military team environments, and this resemblance has not gone unnoticed. DARWARS is an initiative that aims to support a diverse array of distributed simulation-based military instruction, including those that allow large numbers of participants to interact in on-line virtual worlds. In effect, MMPGs are being considered as platforms for future training. In this paper, we describe a framework and evaluate a preliminary methodology for training teamwork skills (e.g., information exchange, teammate monitoring) in an MMPG environment. Data collected from a field exercise of 40 infantry soldiers suggests that MMPGs are capable of training teamwork skills in distributed environments. Based on the results of this exercise, we provide practice-oriented guidelines for using MMPGs as a training tool, and offer some suggestions for future research into effective performance measurement paradigms in this environment.

INTRODUCTION

Massive Multiplayer Games (MMPGs) enable hundreds or thousands of users – distributed geographically – to interact in complex virtual environments. Transported to imaginary realms, users engage in complex strategy, virtual combat, and intricate social/economic interaction (Castronova 2001) from the comfort of their computer desks. To achieve the in-game objectives, users must coordinate and collaborate in ways that seem analogous to the ways in which co-located teams interact. This exercise – Gorman's Gambit – was a preliminary step in verifying that behaviors indicative of teamwork exist within MMPGs, and can be observed and measured to support future distributed team training.

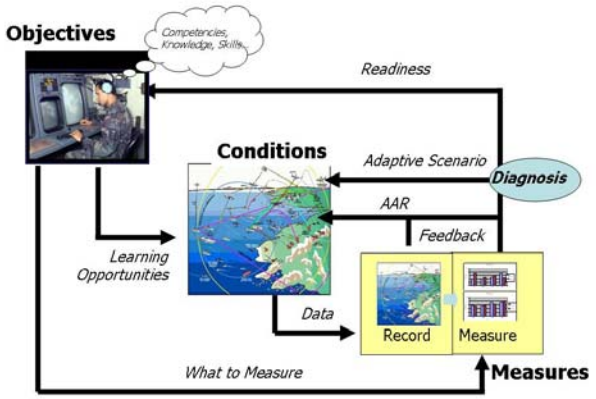
Background

The Gorman's Gambit project was inspired by Gen. Paul Gorman (US Army, Ret.), who asserted that effective teamwork training can be achieved using commercial MMPGs. The level of realism does not have to match modern military operations. For example, a game that requires team members to storm a medieval castle could be used to teach modern assault tactics. Therefore, many existing games have the potential to provide an experience that has pedagogical value, even though they were not initially designed with this in mind. This thesis has a firm ground in cognitive theory, in that learning by analogy is an effective means of obtaining skills that will be applied in unrelated contexts (Gentner, 1989). In short, although the level of technology and the obstacles faced are very different, the elements of teamwork may be similar for both the "Siege of Camelot" and the "Siege of Baghdad."

Due to the increasing cost of large-scale real-world exercises, the distributed nature of teams, and the availability of low-cost internet connectivity, many entities (military, law enforcement, corporate, academic) are interested in using MMPG technologies as a training platform. Because of this interest, Gorman's Gambit project was undertaken (Weil et al., 2004). Gorman's Gambit is part of a larger government program, DARWARS, concerned with facilitating effective experiential training. The exercise was designed to examine the technical, logistical, and pedagogical design issues involved in using an MMPG as a basis for providing a military training exercise in which players interact in distributed teams. For the purpose of the Gorman's Gambit exercise, we employed Neverwinter Nights (NWN; Trademark of Wizards of the Coast, Inc.) as our gaming environment. This paper examines the pedagogical issues surrounding the potential use of MMPGs for training – and the development of effective methodologies for assessing behavior in those environments.

A Framework of Objectives, Conditions, and Measures

The Objectives, Conditions, and Measures (OCM; Figure 1) framework serves as organizing principle for training in DARWARS. When designing training systems and content, we typically start with a general process that moves from Training Objectives to Conditions – as manifested in scenario design (Figure 1). This general approach has been used to develop, for instance, team performance assessment measures for Air Force F-16 pilots flying in a four ship simulation training environment (MacMillan et al., 2000). The process systematically decomposes Training Objectives into knowledge, skills, abilities (KSAs) that can be addressed via tasks that must be performed in the context of a particular set of training scenarios.



I

Figure 1. The Place of Objectives, Conditions and Measures in Simulation-Based Training.

In this framework, trainees have specific Training Objectives to gain competencies, knowledge, or skills in specific subject or domain areas. They are matched to the appropriate training experience based on these objectives. Trainers have the ability to craft a virtual world to help the trainee meet those objectives by specifying the characteristics or conditions of the virtual environment with more precision and control than is possible in the physical world. Trainee responses within the simulation can be recorded and contextualized in terms of the Training Objectives. Meaningful measures can be taken based on trainee performance, recorded, and then extracted to provide assessment and feedback to trainees both during a training session (in the form of coaching) and after a session (in the form of an After Action Review [AAR]). The measures give trainers insight into trainee progress with regards to the objectives. Diagnosis of these measures enables the next match between training experience and trainee. The detailed record of performance in the simulated world may allow predictions about trainee readiness beyond it.

Explicit instantiation of measures with regards to objectives also allows trainers to efficiently interpret performance, diagnose the trainees' needs, and to personalize the training curriculum—all the while updating a record of the trainee's remaining objectives. Similarly, the performance of teams and other groups of people can be evaluated based on performance within the simulator and assessed in relation to past performance and established criteria. Over time, information about the progress of many users through multiple training experiences may reveal trends that can be used for scenario and curriculum development.

Teamwork Objectives

As the primary training objective for a Gorman's Gambit, we focused on the instruction of teamwork skills in a platoon level infantry unit. According to Orasanu and Salas (1993), accepted characteristics of a team include multiple individuals, multiple information sources, interdependence among team members, defined roles, and common goals. Hence, teamwork skills are the behavioral processes that support effective team functioning, and these skills are critical for most military operations.

The research literature provides ample evidence that teamwork skills do exist, that these skills can be defined, trained, and assessed, and that teamwork skills training can improve performance (e.g., Salas & Cannon-Bowers, 2001). More specifically, several key skills have been identified that support team effectiveness (e.g., Freeman et al., 2003; Sims et al., 2004; Smith-Jentsch et al., 1998). These skills include, but are not limited to, such items as team member monitoring and back-up, information exchange, and leadership. These teamwork skills are the focus of our inquiry. We ask whether these teamwork skills can be observed in MMPG environments and how best to support the measure such skills in these environments.

METHOD

The Gorman's Gambit exercised was conducted over two days. The first day consisted of six hours of training and practice on how to play NWN, how to use the capabilities of an assigned avatar, and how to work together as a squad in NWN.

The second day consisted of the main Gorman's Gambit exercise. The exercise scenario was similar to a capture-the-flag event at a "Platoon vs. Platoon" level. The participants were separated into two competing platoons of size 20. Each platoon was comprised of 3 squads, each with a similar mix of player types. Each participant was assigned a specific avatar and a role (platoon leader, squad leader or squad member). The avatars varied in their individual capabilities in order to promote teamwork and collaboration between players as they had to cooperate to apply different combinations of capabilities in order to meet mission demands.

Participants

Forty members of a U.S. Army Infantry Platoon participated in the exercise. The soldiers were between 19 and 33 years of age (M= 23.6 years) with between 1.5 and 174 months of military experience (M= 51.2 months). Participants ranged from E-2 (Private) to O-1 (Second Lieutenant) (M= E-5, Sergeant). The soldiers averaged 51.2 months of service, and 25 of the 40 soldiers had been deployed within the last year. The soldiers averaged 3.7 hours per week of computer use in the last year, with 4 hours of game-playing (PC based or video game console systems [e.g., Playstation, Nintendo]) experience per week in the last year.

Platoon and Communication Organization

The character organization was designed to best support the objective of teamwork demonstration. We encouraged inter-squad teamwork by allocating different resources and abilities to different character types. Because of the heterogeneous composition of the squads, squadmates were dependent on one another to accomplish certain tasks. We encouraged intra-squad teamwork further by controlling communications, creating a hierarchical communication system in which we determined who could talk to whom. In

order to coordinate multi-squad tactics (or tactics involving single players from different squads) communication needed to be sent up the proper chain of command, involving the Platoon and Squad Leaders.

Each team consisted of 20 members. The characteristics of each character class were manipulated and balanced to encourage behaviors indicative of teamwork. The Platoon and Squad Leader positions were determined by the participant's real-world rank and experience.

Scenario

Using the built-in scenario editor, the Aurora toolset (Trademark of Bioware Corp, copyright 1997-2005), we created a scenario loosely based on "capture the flag." This design was chosen because it could accommodate a relatively large number of participants, it is an engaging and competitive task, and it required relatively modest development resources because it does not require extensive non-player character (NPC) development or the use of confederates. In each of two camps, a flag was placed which indicated possession or ownership of that territory. Adjacent to the flag was a lever that was used to indicate a change in possession of that flag and surrounding territory. Once a lever was pulled, it would remain in the possession of the puller's platoon until a member of the opposing platoon gained access to the lever, thereby claiming it as their own. In addition to the levers in the two camps, a third flag and lever were located in a "Hidden Camp." Players were told that the hidden lever could be found somewhere in the environment, but were not given its specific location. The hidden camp was protected by NPCs who could inflict damage to the avatars. The stated goals of the game were to defend your flag while capturing the flag of the opposing team or of the hidden camp.

The Gorman's Gambit exercise consisted of three successive game sessions. Each session involved running the game scenario and initiating the mission from the same starting condition (each platoon in its own camp). Each game session lasted 30 minutes and was preceded by a planning period of 15 minutes and followed by a debriefing period of 15 minutes. During planning, the platoons were placed in separate rooms for privacy. Platoon Leaders could relay their strategies for the upcoming session to the entire platoon.

During the game session, all participants sat at their computer stations to play the mission. All of Team A was located in one large room, where they were seated as pairs at computer cubicles. Due to facility constraints, Team B was distributed through four rooms, with squads largely seated together.

During debriefing, the two platoons were again divided into separate rooms. The debriefing periods were opportunities for each team to reflect on specific occurrences during the session: successes, failures, and surprises. Lessons learned during the debriefing were noted and applied to future sessions. These debriefings were conducted by the platoon leadership.

At the end of the third and final session and debriefing, an AAR was conducted jointly with both teams. While debriefings were opportunities to reflect on specific occurrences during the sessions, the AAR provided an opportunity to reflect on the potential for using MMPGs for training. Participants were asked questions by the experimenters, who also facilitated discussion. The AAR lasted approximately 30 minutes. A camera was used to videotape selected activities during each planning period, game session and debriefing, as well as to record the entire AAR.

Teamwork Measures

The measurement emphasis was on the careful observation of teamwork skills. Several approaches were used in concert to elicit instances of teamwork, and are described below. Our aim was not to capture every instance of teamwork, but instead to obtain a sample of skills to use as evidence of the benefits or disadvantages MMPGs bring as training tools.

The Observer In-Game Evaluation Form (Table 1) allowed observers to document event details (Event #, Time, Participants Involved), a detailed description of the event, the presence or absence and quality rating of several key teamwork skills, and comments on details of the event. Each observer carried several forms at any given time during the exercise. An example of a completed event capture can be seen in Figure 6.

Key teamwork skills assessed in this exercise include monitoring, back-up, team orientation, communication push and pull, and coordination. Observers used the operational definitions of each teamwork skill as described in the content dictionary (Table 1). Skills and definitions were adapted from current teamwork literature (e.g., Salas, Sims, & Burke 2004). Observers were also free to capture other aspects of teamwork as exhibited. Observations made and the quality of events captured varied. Specific observations sometimes included quotes directly from the participant or detailed descriptions of the event with names, individual actions, and overall outcomes. Broad observations included a summarized event with action or outcome as the main descriptor.
Observers were present with each team throughout the exercise. During game-play, observers moved throughout the team experiencing game-play over the participants' shoulders. In this minimally obtrusive method, observers could see where participants were in the game, orient themselves to the participants' perspective, hear voice communications from both the game (other players) and the participant in view, observe actions directly and remain in the gaming atmosphere without being disruptive to the participants. Every few minutes the observer would move to another participant for another perspective to observe. Between two and four observers were filling in the Observer In-Game Evaluation Forms during this time. As this was essentially a pilot study, additional observations were made from within the game using a specialized view (discussed below), after the game based on video footage, and directly from the users during a debriefing and questionnaire.

Table 1. An event captured on the Observer In-Game Evaluation Form. Quality scale raged from 1 to 5; 1 indicates a poor example of the behavior; 5 indicates an exemplary example.

#	Time	Participants Involved	Description of Event	Teamwork Skills Exhibited			Comment
1	08:00	Team A Archer and Team B Tank	Tank and Archer engage in battle. Archer calls up squad members for back up. Squad confirms message receipt, arrives and destroys Tank.	**Skill Shown**	**Y/N**	**Quality (1-5)**	*Formal comms used – low ambiguity.*
				Monitoring	Y	3	
				Back-up	Y	4	
				Coordination	Y	3	
				Comm – Push	N	n/a	
				Comm – Pull	Y	5	
				Other… Closed Loop Comms	Y	5	

RESULTS

More than 115 instances of specific teamwork skills were extracted and reviewed, based on the data sources Observer In-Game Evaluation Forms and user responses. It would be feasible to extract several hundred more examples of teamwork from these data sources. However, our intent in this discussion is not to document all of the instances of teamwork, but rather, to show a number of instances in order to demonstrate that teamwork does occur, and therefore, that MMPG environments may provide at least some of the conditions necessary for teamwork skills training. The teamwork behaviors seen in these instances are enumerated in Table 2.

Table 2. Instances of teamwork seen in the current exercise.

Behavior	Number of Observations
Monitoring	75
Coordination	88
Pushing Information	57
Pulling Information	35
Leadership	69
Orientation	66
Backing-Up	53
Adaptability	43
Closed Loop	6
Mental Models	58

Most events involved multiple teamwork skills (5-6 skills), as teamwork skills often work in conjunction with each other, such as leadership and monitoring. Leadership involves directing and coordinating the activities of others, assessing performance and/or assigning tasks (Salas, Sims, & Burke 2004). Monitoring involves understanding the team environment and applying appropriate strategies (Salas, Sims, & Burke 2004). The two skills have overlapping definitions that would make it probable, although not without exception, that where you have leadership you also have monitoring and vise versa.

DISCUSSION

The observers cataloged many instances of the teamwork skills and the participants indicated that they were consciously aware of the need to exhibit these behaviors to be successful. This indicates that the behaviors indicative of teamwork can be observed within the context of MMPGs – the primary question of interest for Gorman's Gambit. The sampling methodology used in this preliminary study was effective in capturing instances of teamwork. However, it is unclear how many observers were required to achieve a truly representative sample of behaviors in this dynamic, multi-person environment. It is also unclear if the fields included in the Observer In-Game Evaluation Form were the ideal fields, and if the scales used were sufficient to capture the complexity of interaction. These will need to be validated in future efforts.

The DARWARS OCM model relies on valid measures of performance and interaction to inform Training Objectives. This requires the ability to capture data about game-play directly from the game. Most commercial games, including NWN, are not instrumented to capture the aspects of interaction that are indicative of teamwork behaviors. This is an obstacle to using COTS technology for training purposes. To compensate for this deficiency, our primary method of assessment was human observation. However, as the number of simultaneous users increase, the sustainability of using human observers diminishes. It would have been advantageous to capture and log data on the activities of the players directly from the game in order to be able to automatically measure the types and degree of teamwork exhibited by each player. This would have allowed a larger sample of behaviors to be recorded, and would have supported the monitoring of improvement over time. NWN did provide the potential, through its scripting language, to measure and log some simple game-play interactions (e.g., who attacked whom and when, which team members were nearby at specific times, who healed whom). Future efforts should integrate these measures with human observation measures.

One requirement for observer based assessment is the ability to track interaction from a global perspective and feedback performance measures to the users. NWN did provide one method for unobtrusive in-game observation. Within NWN, a special viewing mode, represented by the "Dungeon Master" (DM) character, may be used to observe

game-play without being perceived or affected by the players. Multiple DM characters may be played simultaneously, allowing multiple observers. The DM character has the same limited graphical viewing capabilities as any avatar. Thus, it is capable of achieving a limited "bird's eye" view, in a small area around the DM. To compensate for this shortcoming, the DM has the capability to capture the general location of all avatars in the game through a text-based list. The DM can move instantly to any location in the virtual world, and may use this to focus in on a specific avatar. Our initial goal was to monitor game-play for instances of teamwork via a this DM perspective, allowing observers to view players and their communications. However, the perceptual limitations of this assessment tool did not sufficiently allow for the observation of more than one small region of the scenario space at any given time. Thus, without a very large number of observers (one or more per squad), the assessment/monitoring tool provided with the NWN software package did not allow for comprehensive monitoring of teamwork and coordination. 'Over the shoulder' techniques were subsequently used for the majority of observations.

The use of communications analysis to assess teamwork is a common practice in the Team Performance literature (e.g., Entin et al., 2003; Freeman et al., 2003). Such analysis may help exercise coordinators identify and timestamp instances of teamwork, and ultimately, such a capability is essential for student evaluation in a training tool. Hence, it will be advantageous for a MMPG designed for training to have the capability for communications capture and automated analysis. NWN provides the capability to log the in-game chat communication, and the third-party chat and verbal communication tools we used also provided this potential. However, we did not exploit these capabilities for the exercise due to logistical constraints. Future efforts should capitalize on this functionality for both research and training purposes.

Collectively, these findings indicate that MMPG environments may be functional models to emulate for distributed multi-user military training. The behaviors of interest do occur and can be observed. However, the findings also highlight several inherent challenges with using COTS gaming systems in assessment-intensive applications such as military teamwork training. Our results therefore support the potential utility of MMPGs for training with the caveat that additional training components would need to be developed. In circumstances in which teams are distributed or environments are inaccessible before deployment, MMPGs may be a strong tool. If the assessment and lessons learned challenges are met effectively, then the DARWARS vision of using existing technology to supplement military training has the potential to be fully realized.

ACKNOWLEDGMENTS

The authors of this report would like to thank several individuals for their impact to this project Ralph Chatham at the Department of Defense Advanced Research Projects Agency, General Paul Gorman (U.S. Army, retired), Harold O'Neill, Richard Wainess, Scott Hardcastle, Bruce Roberts, William Ferguson, Fred Diedrich, and Elliot Entin. Particular acknowledgements are given to the soldiers who participated in the final exercise at Ft. Benning, GA.

REFERENCES

Castronova, E. (2001). Virtual Worlds: A First-Hand Account of Market and Society on the Cyberian Frontier. *CESifo Working Paper No. 618.*

Dwyer, D.J., Fowlkes, J.E., Oser, R.L., Salas, E. & Lane, N.E. (1997). Team performance measurement in distributed environments: The TARGETS methodology. In M.T. Brannick, E. Salas, and C. Prince. *Team Performance Assessment and Measurement*. Mahwah, NJ: Erlbaum.

Entin, E.E., Diedrich, F.J. & Rubineau, B. (2003). Adaptive Communication Patterns in Different Organizational Structures. Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, Denver, CO.

Freeman, J., Diedrich, F. J., Haimson, C., Diller, D. E., & Roberts, B. (2003). Behavioral representations for training tactical communication skills. Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation, Scottsdale, AZ.

Gentner, D. (1989). Mechanisms of analogical learning. In S. Vosniadou & A. Ortony, (Eds), *Simularoty and analogical learning* (pp. 199-241). London: Cambridge University Press.

MacMillan, J., Littleton, E.B., Price, J., Miescher, S.C., Entin, E.E., Gentner, F.C., Tiller, T.C., & Cunningham, P. (2000). A Streamlined Scenario-Based Methodology For Evaluating Distributed Team Performance. Aptima Technical Report, Aptima, Inc., Woburn, MA.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. Annual Review of Psychology, 52, 471-499.

Sims, D.E., Salas, E., & Burke, C.S. (2004). Is there a "Big Five" in teamwork? 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Smith-Jentsch, K.A., Johnston, J.H., & Payne, S.C. (1998). Measuring team-related expertise in complex environments. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271-295). Washington, DC: American Psychological Association.

Smith-Jentsch, K.A., Zeisig, R.L., Acton, B., & McPherson, J.A. (1998). Team dimensional training: A strategy for guided team self-correction. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271-295). Washington, DC: American Psychological Association.

Weil, S.A., Hussain, T., Diedrich, F., Ferguson, W., & MacMillan, J. (2004). Assessing Distributed Team Performance in DARWARS Training: Challenges and Methods. Proceedings of the 2004 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).