

Human Behavior Model Interoperability For Training In Virtual Worlds

David McDonald, Richard Lazarus, William Ferguson,
Alice Leung, Talib Hussain
BBN Technologies
Cambridge, MA
{dmcdonald, rlazarus, wfergu, aleung, thussain}@bbn.com

Barry Silverman
University of Pennsylvania
Philadelphia, PA
barryg@seas.upenn.edu

ABSTRACT

It is commonly recognized that a key challenge of urban military operations is the complex terrain. However, the large number and diversity of people present another less recognized but critical challenge, especially for non-combat operations. Warfighters must understand the socio-cultural effects of their actions, handle direct interaction with non-combatants (and potential combatants), and understand the immediate and long-term consequences of their interactions: every soldier must be a skilled ambassador. Given that realistic training requires many participants, the high cost and limited availability of highly-qualified participants restricts access to quality training opportunities. Synthetic characters can provide the numbers, but present-day simulated characters have limited capabilities for interacting with each other and with humans in the rich, meaningful way that is required. Thus, we must populate these virtual training worlds with characters whose behaviors are generated by human behavior models (HBMs) rather than humans. While significant achievements have been made in developing HBMs that are able to control a single simulated character (or a single group of simulated characters), a serious limiting factor has been the inability of heterogeneous HBMs to interact with each other or share knowledge about their environment. We present an architecture and multi-level message framework for enabling HBMs to communicate with each other about their actions and their intentions. Also, we describe our conception of the “Mandscape,” which will facilitate the use of massive numbers of synthetic characters by representing the options and state of social relationships and interactions as shared affordances in the environment. Finally, we describe an application of our prototype training system in which three heterogeneous HBMs developed by various organizations interoperate within a single training-oriented scenario. We believe that this approach will encourage the development of standards for interoperability among HBMs that will lead to the development of richer training and analysis simulations.

ABOUT THE AUTHORS

Dr. David McDonald a Senior Scientist at BBN Technologies. He is a computational linguist with interests in conceptual modeling and knowledge representation. He is applying his skills in projects for understanding notices to airmen (NOTAMS) and Integrated Learning.

Richard Lazarus is a Director of Research at BBN Technologies. He manages a group pursuing research in scalable architectures, distributed and experiential learning, and knowledge representation and visualization. Also, he is the program manager for DARWARS – a program focused on bringing pedagogical and simulation interoperability to game-based experiential training.

William Ferguson is a Division Scientist at BBN Technologies. His research areas include artificial intelligence, game-based training, and human perception and learning. Currently, he is the co-Principal Investigator on the DARWARS Training Superiority program and co-Principal Investigator on the DMSO/AFRL/ARL SABRE (Situation Authorable Behavior Research Environment) project.

Dr. Alice Leung is a Scientist at BBN Technologies. Her current interest is in harnessing games for behavior research and training. Dr. Leung was the technical lead for the SCALE-UP project (Social and Cultural Analysis and

Learning Environment for Urban Pre- and Post-Conflict Operations). She is also a co-Principal Investigator on the DMSO/AFRL/ARL SABRE (Situation Authorable Behavior Research Environment) project. SABRE is designed to allow researchers to study the effects of culture on team communication and decision making within a controlled, game-based scenario.

Dr. Talib Hussain is a Scientist at BBN Technologies. He works on game-based training, evolutionary algorithms for robust tactical behaviors, and machine learning. He is currently a technical lead on the DARPA-sponsored POIROT project, which is developing the capability for integrated learning of tasks based on a single observation of a human performing the task.

Dr. Barry Silverman is a Professor at the University of Pennsylvania, where he is the Director of the Ackoff Center for Advancement of the Systems Approach. His research focuses on aesthetic and cognitive engineering of embedded game-theoretic agents that can help humans improve their learning, performance, and systems thinking in task-environments.

Human Behavior Model Interoperability For Training In Virtual Worlds

**David McDonald, Richard Lazarus, William Ferguson,
Alice Leung, Talib Hussain**
**BBN Technologies
Cambridge, MA**
{dmcdonald, rlazarus, wferguso, aleung, thussain}@bbn.com

Barry Silverman
**University of Pennsylvania
Philadelphia, PA**
barryg@seas.upenn.edu

INTRODUCTION

U.S. military personnel are being trained through a variety of techniques to prepare them for both the combat and non-combat missions for which they will be tasked. For non-combat operations in urban environments, these missions hinge on being able to interact with diverse and foreign cultures where warfighters are expected to influence local residents, seek their cooperation, and rely on them for information. Troops must understand the cultural and social effects of their actions, handle any direct interaction with non-combatants (and potential combatants), and understand the immediate and long-term consequences of their interactions. These skills seem more appropriate to ambassadors than to combat troops, but these are the skills required of many warfighters throughout every echelon of the deployed force.

Recently, these interpersonal and organizational skills have become a training focus throughout the military that includes language and cultural training. Training is provided through a variety of means, including didactic courseware, simulation- or game-based training, and live role-playing exercises. Each of these training genres has its place, but immersive, computer-based training has proven to provide a tremendous benefit/cost ratio, as well as providing ubiquitous access (Alexander et. al., 2005). However, such immersive, virtual environments are limited in their scalability. One of the great challenges of urban environments is the large number and diversity of people. Therefore, such training systems are limited by the availability of authentic and skilled role players. Another challenge is the time factor: successful human interactions are predicated on established relationships, and projecting influence or obtaining information depends on trust and other human factors that must be cultivated over extended periods.

Given the large numbers of participants that are required for realistic training, using highly-qualified people as the characters for experiential training is costly and severely limits the availability of quality training opportunities. Synthetic, computer-driven characters

can provide the numbers, but with a few exceptions, such as ISI's Tactical Language Trainer (Johnson et al., 2005), present-day simulated characters have limited capabilities for interacting with each other and with humans in the rich, meaningful way that is required. Currently, the primary technique employed is the hard-coded scripting of behaviors in a game-specific scripting language, which severely constrains their usefulness. However, some instances of more sophisticated behavior modeling are being employed with good results in constrained situations (Johnson et al., 2005). This increased use of behavior models is the direction that we need to take to reduce the reliance on large numbers of people to deliver meaningful training experiences. Recently, significant achievements have been made in developing human behavior models (HBMs) that are able to control a single simulated character (or a single group of simulated characters), but a serious limiting factor has been the inability of HBMs developed by different groups to interact with each other. We refer to the class of HBMs that are providing the behaviors for such avatars as *people engines* (PEs). People engines are analogous to game physics engines, which calculate the physical properties and motion of objects in a game or simulation, but PEs generate the motivations, emotions, desires, and actions for the virtual-world characters.

While the research and the development of more-sophisticated and less-scripted PEs is making significant progress, it is short-sighted to believe that any one PE or theory can provide the breadth and depth to populate virtual worlds in a sufficient manner to provide training across a broad set of skills. We believe that such rich virtual worlds should be constructed by being able to "plug and play" the best-of-breed PEs based on the training requirements. However, current PEs operate at different levels of abstraction and, typically, contain internal representations of the environment or context. In order to realize this vision of supporting multiple PEs cooperating in a common environment, we need to provide both an environment that contains an explicit, shared semantic representation of the world and communication mechanisms that support models at different levels of abstraction.

As we teach warfighters to navigate the physical aspects of the landscape, we need also to teach them to navigate the obstacles and pitfalls present in the mental world, i.e., “the Mindscape.” Moreover, if we can make this Mindscape explicit, we can provide a shared representation of the virtual space that will enable people engines to understand context and behave appropriately in a variety of situations, with minimal scripting. Another important aspect of our explicit Mindscape is that it allows us to provide a set of authoring and analysis tools by which one can control training experiences and understand decision-making processes and skills.

In this paper, we present our development of a synthetic-environment framework for training and for providing interoperability through common programming interfaces (APIs) and pedagogical tools. We describe our efforts at integrating three heterogeneous people engines with a massive multiplayer on-line game (for visualization and physical world modeling) as a prototype training system that we created as part of our SCALE-UP project (Social and Cultural Analysis and Learning Environment for Urban Pre- and Post-operations). To explore the benefits of our approach, we developed a specific training scenario in which a human player was given the goal of dispersing a crowd by using a sequence of dialog choices. Here, we review the results of this experiment. In closing, we describe our research plans and offer a future vision for generalized non-kinetic warfare (i.e., non combat) training.

THE MINDSCAPE

Overview

The social skills that we must impart to warfighters to prepare them for non-kinetic warfare operations revolve around the both long term and transitory relationships that they develop with indigenous populations. These relationships are not physical things in the world; you cannot touch them or actually see them, but they are very real nevertheless, and they exercise a great deal of control over human behavior. When someone greets you in the morning, you trivially greet them back. When you are asked a question, you know that you are expected to give an answer. Even if you cannot answer right away or do not want to, you still recognize that the act of posing a question has brought a questioner-questioned relationship into existence. Then, you will have to address it or else risk the possibility of endangering the larger relationship that you

are part of by being seen as rude or indifferent towards the other person.

Important cues that help us recognize and sustain relationships occur through our facial expressions, body stance, and actions. You recognize the store greeter because they walk towards you in a deliberate manner, wear a name-tag, and ask you a predictable question. Other relationship-driven actions can be quite subtle, as when a person in a conversation shifts their posture and facial expressions to indicate that they would like a turn to talk.

If PE-driven synthetic characters are going to interact effectively with human trainees, like humans they will have to recognize and maintain the social relationships that are expected of them given their role in the training scenario. The PE that is playing the role of a community leader negotiating with a human trainee on behalf of a crowd of synthetic villagers needs to know how to indicate through its avatar's body stance, movements, and choice of language whether it agrees with the trainee's proposals or not. It needs to know that one of its options in a negotiation is to threaten to stop talking or to walk away for a private consultation with other village leaders. *The PE not only needs these behaviors in its repertoire, but it has to understand their likely consequences and counter moves by the other party in the negotiation.*

Relationships as Externalized Affordances

To provide this social relationship information to the PEs, we developed the concept of the *Mandscape*. Just like a continually-updated description of the locations of the furniture and other people (avatars) in a room gives a PE a model of the landscape, allowing it to navigate without hitting anything, the Mindscape represents the social relationships that are active from moment to moment. It provides the PEs with a model of the relationships in which they participate and the roles that they play. It tells them which actions are available given their roles, and which would be inconsistent for those roles. (For example, you do not shoot at someone while you are negotiating with them.)

Social relationships by definition involve multiple people at the same time. People have internal models of these relationships, and it is our extraordinary skill at recognizing how other people's activity and manner affects their state that enables us to be social beings. The PEs that we have worked with possess none of these recognition skills, but we can provide their equivalent by externalizing the Mindscape as a shared resource within the game architecture. Then, PEs can

use it to guide their actions, just as an ordinary scripted synthetic character (also called a non-player character, or NPC) uses the game engine to tell it about enemy positions and whether its shots are on target.

We refer to objects that populate the Mindscape as *mentities* (mental entities). Using computer-programming techniques we can represent mentities similar to how we represent an explicit, externalized landscape: mentities are instances of standard classes and have expected properties that can be overridden for particular cases. A chair, for example, is a movable object that provides affordances such as sitting, moving out of the way, or even throwing. A particular chair, however, may have a weak seat and so, for it, the affordance of ‘standing on’ would be unavailable.

Gibson’s concept of affordance (1979) is central to our conception of the Mindscape. From a computational perspective, the use of affordances consists of moving the knowledge about what can be done with something, such as a chair, out of the agents’ internal code and into the object itself. This technique has been used to good effect in some applications of the Human Performance Moderator Function Server (PMFserv) to handle physical objects (Cornwell et al., 2003) as well as in the SIMS game (Cass 2002), and recently by McAlinden and Clevenger (2006) to “paint” the ground plane of the game environment with broad-brush cultural dispositions (e.g., conservative Christian church) that guide the actions of agents as they move over those regions. Our use of affordance is quite different than these in that we apply it to inter-PE (avatar) social relationships rather than to objects. The PsychSim project at ISI (Marsella et al., 2004) has a similar focus on social interactions but places this knowledge inside the agents in a decision-theoretic framework rather outside the agents as affordances.

In our approach, a social relationship is a Mindscape object (i.e., a mentity), relationships encompass roles, and roles afford actions in a given context. So, unlike objects in the landscape, Mindscape objects are intrinsically relational. Their existence depends on the participation of the PEs whose social activities they represent. A PE (via its avatar) participates in an existing relationship by adopting one of the roles intrinsic to that relationship. Consider a crowd that has gathered to demonstrate (see Figure 1). In this situation mentity, “being part of a crowd,” the roles could include crowd member, spokesman, bystander, or even falafel seller. Through its roles, a mentity affords certain actions to the PEs at certain times (“shout,” “murmur,” “hide”). The availability of these roles (or their unavailability)

is one of the affordances of a Mindscape object, and can change as the state of the relationship changes.

Mentities have their own logic and state. The state of a crowd, for example, is different from the sum of the states of its individual members. A crowd may be “growing” as more PE avatars come and join it (enter into the social relationship of being part of that crowd) or “shrinking” as individuals or clusters of people move away to become “bystanders” or to leave the area completely. The logic within the crowd mentity dictates how many people have to leave before the crowd per se – the situation instance within the Mindscape – ceases to exist except perhaps in the memory of its former members. Along different dimensions of its state, the crowd may be “loud” or “quiet,” “stationary” or “surging forward,” “happy” or “angry.”

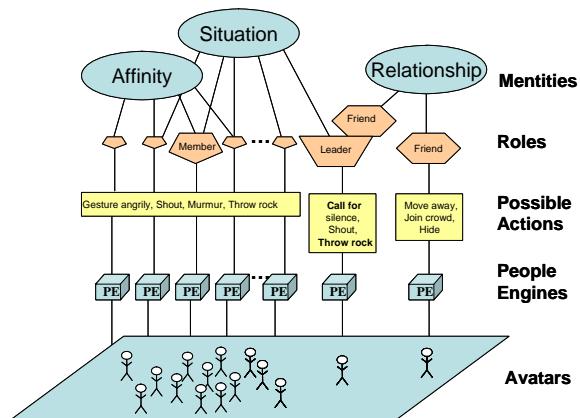


Figure 1. Mindscape Components

The state of a mentity in the Mindscape is visible both to the PEs that are members of it and those that are observing it (e.g., soldiers observing a crowd). Where mentities have a manifestation in the landscape (e.g., the avatars of the crowd members), the observers obviously have to be able to perceive the physical object (e.g., hear the noise of a crowd gathering outside their building) before they can observe the corresponding social object’s state. Many of the more interesting Mindscape relations, however, will involve synthetic characters that are functionally rather than physically related, such as the members of a clan or a family and the cultural dispositions that they share.

Mandscape objects (i.e., mentities) are active. They incorporate logic to change their state in reaction to the actions of the PEs that are part of the relationship that they represent, including coming in and out of existence. New mentities are instantiated by avatar actions that meet their initiation criteria. From the perspective of the Mindscape, for example, every person affords

the possibility of starting a conversation with them. If you shoot someone while negotiating, you left the “negotiation” relationship the moment you took out the gun and shifted into a “fight” relationship. This would be reflected in the Mindscape and seen by all the other PEs that could observe this negotiation.

Because the relationships in the Mindscape have been externalized, they are shared by every instance of a synthetic crowd or a pushcart vendor or queue of people at a checkpoint that an extended scenario may require. (Shared in the sense that each mentity is an instance of the same relationship class, though their state will be instance-specific.) This makes for economy of development: a large scenario may involve many instances but comparatively few classes. On the other hand, it is important to point out that crafting a rich, effective, and compelling relationship class that models the social interactions of a set of PEs can require as much work and background research as crafting the cognitive model or stress-reaction facilities inside a PE.

The mentity classes are also available as a repository of the knowledge of how an avatar should behave in a given role (e.g., to start talking to someone you move close, face them, and indicate to them that you want their attention). They can provide this knowledge to a PE (as an action script) and simultaneously to the animation engine (for low-level details of avatar behavior). Also, they can do this for any PEs that take on that role, relieving the PE developers from having to do it themselves, and ensuring that the behaviors will be consistent.

Since the Mindscape is externalized rather than implicit in each of the PEs (as it is implicit in people), we can simplify the creation of scenarios by using the Mindscape to mediate the (virtual) perceptions of the PEs. Consider a crowd made up of several affinity groups that we want to act as though they have different degrees of experience with Americans, and imagine that they are in a scenario where they have just seen an American give the “thumbs up” sign. If we had the time and development resources, we could give them rich back stories and cognitive models that would motivate their perception of the sign internally. But, we could equally well use the Mindscape to provide the affinity groups with already-interpreted, external perceptions of the gesture, reflecting different cultural awareness. A group that is to act as though it knows Americans well will see it as it was intended: a gesture that means there was a good outcome. A group that only knows the customary Iraqi interpretation will be told by the Mindscape that the gesture is an extremely rude insult.

HETEROGENEOUS MESSAGING THROUGH MULTIPLE LEVELS OF ABSTRACTION

In this section, we describe the five levels of abstraction in our messaging framework, from the most concrete to the most abstract. The messages can be constructed by the PEs, at the levels that they can handle, or can originate in the mentities, as just described in the previous section. In the following sections, we show how these levels are mapped to communications between the PEs, the game client, and the world model that mediates among them.

Level 1: Perceptual

At the lowest, raw *perceptual level*, the flow of activity in the game or simulator is “represented” by the audio and video that a rendering client can produce. Utterances are sound, gestures are sets of pixels, and actions are some combination of the two. At this level, messages contain raw information with little to no annotation. In order to process this information, the behavior model must apply perceptual mechanisms directly. For example, a statement may be provided as an audio clip, and speech recognition would need to be applied to process it.

Few game-behavior models currently have perceptual or manipulation mechanisms that would allow them to operate at this level, so we doubt that there will be much call for them in typical game domains. However, operating at this level would be important for some purposes, such as constructing robot testbeds.

Level 2: Literal

At the *literal level*, the flow of time and activity in the simulator is divided into discrete events. At this and subsequent levels, these events are annotated with computer-readable, symbolic descriptive information. A given event might have annotations from any or all of levels two through four.

At the literal level, utterances are represented as a single event annotated with the speaker and the sequence of words and (if possible) prosodic information, as well as descriptions of the coordinated non-iconic gestures and facial expressions that accompany the speech. Iconic gestures are represented as events, which include a physical description of the motions that occurred and who performed them. Actions are represented in world-centric terms, such as absolute coordinates for motion, and other primitives that are natural for the simulator. (In systems where the simulator represents the world at a more abstract level, the

literal and semantic annotations for action could be identical.)

Level 3: Semantic

At the *semantic level*, events are annotated with a symbolic representation of their content. We use the term *semantic* because when the event is an utterance, the annotation at this level resembles the interpretation that a good semantic parser would produce.

Utterances are annotated with their literal meaning. So “*I’m cold*” will be represented as a statement about temperature rather than an indirect request to make the speaker warmer, and “*Do you know what time it is?*” would be represented just as a query about a capacity to provide knowledge. Gestures are annotated by an unambiguous (though perhaps abstract) representation of their meaning as the agent making the gesture intended it. Actions are annotated in functional, scenario-relevant, terms such as, “move-to(door-1),” rather than with the spatial coordinates that appear at the literal level.

Level 4: Interpreted

Interpreted annotations include the intent of the performer of the event. They may also include suggested responses or intended consequences. Meanings of the responses to events can also be provided by other components besides the instigator of the event. Information at the interpreted level can be self-contradictory (and often will be if provided by different PEs). The social model (active mentities) may also annotate events at the interpreted level by providing cultural or context specific interpretations of events (or possible responses, etc.).

In the interpreted level, the message contains information about the intention of the act (i.e., saying “*You must leave the area!*” is imposing a demand), as well as an interpretation of the effect of the action on the people to which it is addressed (i.e., to show disrespect).

Level 5: Narrative

At the highest level, the purpose, role or function of events is included in annotations. This purpose comes from something external to the simulated world, even external to the representations held by the PEs in that world. Narrative annotations on events can indicate their dramatic function in a story (e.g., foreshadowing, building suspense, surprising the player, misleading him or her, etc.). If the virtual world is serving some

purpose other than to entertain, then the narrative annotations can relate to that purpose as well. A simulation that is meant to teach might annotate events as examples of pedagogically relevant objects, such as phenomena that were to be observed, apprehended as evidence, or appreciated as counter-examples.

The components that infer or rely on narrative annotations are not the participants in the world, but the shapers, measurers, authors, and other entities whose purpose is to ensure that the entire system (including people, agents, simulators, renders, etc.) carries out its purpose (entertains, teaches, extrapolates outcomes, etc.). Reidl and Stern (2006) discuss a “scenario director” that keeps a learner on track, and Ferguson and Leung (2005) discuss the power of user authorship and the dilemma of managing the learning experience. Our contribution with the Mindscape is a system with common APIs and representations that facilitate such manipulations rather than control the experience through interfaces to a variety of subsystems.

COUPLED-WORLDS ARCHITECTURE

To support communication among PEs and between the PEs and the interface to the human players, we developed the coupled-worlds architecture shown in Figure 2. The lines that connect the components are labeled with the abstraction level of the message being passed. The interactions among the user, game client, and physical world model are typical of “conventional” games. The additional components and interactions were developed to support the social interactions of synthetic entities and the “mental world,” i.e., the Mindscape.

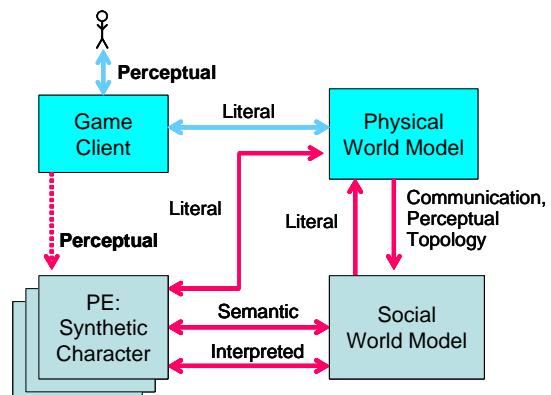


Figure 2. Levels Passed Among Components

On the top left, we have the game client. This is where human users access the scenario. The arrow connecting

it to a person is double headed because it will carry events initiated by the player back to the virtual world. The dotted line linking it to the synthetic characters is to remind us of the possibility that some PEs might want the raw data, and that we should not discount that option.

On the right is a pair of tightly-coupled blackboards that provide shared world models for the benefit of the PEs. We distinguish physical from social models first to emphasize that the semantic and interpreted levels convey interpersonal information that will not make sense outside of the cultural and social situation playing out in the scenario, and also to reflect the fact that additional work is being done by the combination of the two.

The primary task of the coupled components that sit between the synthetic characters and the game client is to transport the event descriptions from the characters or game clients that create them to the ones that need to know about them, dividing out levels according to what is appropriate for the receiving PE. The downward arrow from the physical to the social model indicates that the locations of the characters' and human player's avatars matter in all but the most trivial virtual worlds. What can be seen and heard is location dependent (though it may only be represented topologically), and the flow of events has to reflect this.

The second task, reflected by the Literal arrow pointing upward, is to provide translations for PEs that only communicate at the semantic or interpreted level. The PE driving a particular avatar may be very rich but operate at a level of granularity that is too coarse to provide animation instructions to its avatar (literal information). The social world model can be explicitly programmed to provide a mapping between that agent's interpreted output and animation that would reflect it. This is also a place to share a rich natural language capability that could take semantic-level information and render it as text or speech.

In summary, an agent's action is described by annotations at several different levels of abstraction simultaneously. There is no expectation that every agent will be able to understand or produce every level, and in some instances we arrange for our mediating components to fill in the missing information.

TRAINING ENVIRONMENT AND SCENARIO

To understand the interoperability issues and then demonstrate the feasibility of the messaging framework that we designed to address it, we applied our architecture to create a training scenario that integrated three distinct human behavior models as people engines, and used the Big World™ game environment to render the virtual world and portray the behaviors chosen by the people engines.

In our training scenario, a human user plays a *squad leader* who is trying to peacefully resolve a problem with a crowd that is populated with synthetic characters of a particular culture. Three different people engines were used, each playing a different character (or character type) in the scenario. One people engine, Edutaniacs' PMFserv (Silverman 2001), controlled the *members of a crowd* that were expecting a food distribution that had not arrived. Another PE, CHI System's iGen/VECTOR (Zachary et al., 2001), controlled the *community leader* who was the crowd's spokesman. The third PE, BBN's ENDER, which was developed as part of the project, modeled an *agitator* who tried to influence the crowd against the squad leader. Figure 3 shows the crowd scene.



Figure 3. Squad Leader Addressing the Crowd

The people engines were implemented using different software bases, and each had distinct needs and capacities for input and output. For the most part, each PE was designed to interact with a simulated world at different levels of abstraction. The interactions between the three PEs and the human player all used a three-level message format consisting of the literal, semantic, and interpreted levels. The perceptual level was pro-

vided by Big World™ for the benefit of the human player.

The possible actions of the person playing the role of the squad leader were limited to navigating a fixed dialog tree of utterances and accompanying actions. (This dialog tree also served to bound the range of actions that the developers of the PEs had to consider when adapting their systems to our scenario. However, with one exception, their actions were not scripted.) The simulation framework created a fully complete message (all abstraction levels) for each of the options that the person could choose, and issued it to all of the PEs at the moment the choice was made. (Everyone in the scene is assumed to be able to hear the entire conversation and to see all gestures made by any other character.)

The natural language responses made by CHI System's community leader were also scripted by the dialog tree. CHI's PE used the literal level of the messages from the squad leader in order to recognize which action was taken. The unscripted emotional and cognitive reactions of this PE led to specific literal-level behaviors in the messages that it sent out: for example *Idle-CrossArms*, to reflect the change in the PE's internal emotional state to becoming angry.

The semantic level was used by BBN's ENDER (the agitator character) when it needed to handle an action by the squad leader that required deeper domain knowledge in order to understand its cultural implications. The scenario allowed the squad leader to choose to distribute MREs to the hungry crowd. Because the semantic-level description of this action contained additional information about the type of food in the MRE, the BBN PE was able to examine this information and recognize that the MRE contained an ingredient that was religiously prohibited. Unlike the other two PEs, this PE had only a minimal mental and emotional state, but did have the domain knowledge needed to reason about halal food, and to interpret the distribution of non-halal food as disrespectful. The agitator's reaction upon recognizing this disrespectful act was to tell the crowd members about it at the interpreted level. At that point, it also used the literal level to change its "idle state" (the animation sequence that its avatar takes between explicit commands) to reflect its increased anger and to simulate the effect of talking to its neighbor avatars by moving its avatar's head from side to side and cupping its hands to its mouth.

The twenty avatars that made up the crowd were individually controlled by a single instance of the PMFserv PE. Silverman and his group at the University of Penn-

sylvania have had extensive experience in modeling crowds for computer simulation by PMFserv (see, e.g., Silverman et al., 2006a,b). In this crowd model, we drew on the work by Eidelson (2003) on "Dangerous Ideas." In brief, this model says that a person's affinity with a group and the possibilities of conflict within it or the propensity of individual members to stay or leave depends on how the member see themselves and the others in terms of five modeled beliefs: vulnerability, injustice, distrust, superiority, and helplessness. The model of these beliefs in PMFserv depended on the values that individual crowd members (usually clones of several archetype members) had in their goals, standards, and preference trees. (See Silverman & Bharathy 2005 for more on how these GSP trees are used. See Ortony, Clore, and Collins 1988 for the original development of many of the ideas embodied in the GSP trees of PMFserv.) The crowd members' perception of the events in the scenario and their reactions to it are all filtered through their present emotional and cognitive state and combined with the other performance modulator functions to arrive at the expected utility of the various actions they could take.

Since the crowd has been modeled at a very abstract level when compared with the other PEs, PMFserv (which generated the crowd) viewed the conversation between the squad leader and the community leader in appropriately abstract terms represented at the interpreted level. All of the literal and semantic content of the events in the conversation were projected to the interpreted level in terms of Maslow's hierarchy of needs (Maslow 1987) (life for a villager can be a day-to-day struggle) and rendered as different levels of respect, security, or food. Similarly, the output actions of individual crowd members were summarized by PMFserv as their "grievance state," given on a linear scale. This, in turn, was translated by the simulation framework (the coupled-world models) which filled in the literal level of messages from the individual members of the crowd.

This use of the literal level was primarily for directing the game client's visual rendering so that the human player could understand the state of the crowd in a natural way (its reaction to the action the person had just made and the collective effect of his actions so far). The simulation framework first mapped a crowd member's interpreted-level grievance value onto literal-level actions, and then used those actions to request specific a perceptual-level rendering. For example, a particular level of grievance in a crowd member's event message might lead the social world model to fill in the message's literal level with an action such

as *PoundFist*, and the physical world model would then request the corresponding game animation.

DEMONSTRATION RESULTS

From a training perspective, the vision of the SCALE-UP project is to enable experiential, game-based training systems that allow human learners to interact with a variety of computer controlled entities driven by different people engines. In order for students to learn from scenarios populated by people engines, the characters encountered must react appropriately to the student's actions or decisions. In particular, culturally insensitive actions should have a visible, negative impact on the student's ability to succeed in their mission. Additionally, since the student may need to learn about the differences among several populations, any scenario should support being populated with sets of characters that will react differently to the same series of student actions.

To test the efficacy of our heterogeneous-PE environment, we developed a demonstration of our SCALE-UP prototype training system where a student executing a series of "naïve" decisions would have a less successful outcome than a series of more culturally informed actions, where the outcome is essentially the attitude and reactions of the synthetic characters. We also wanted to show that either series of decisions would have a different impact on two crowd models: one of more hostile individuals and one of more moderate individuals. Table 1 shows the different outcomes that we were able to demonstrate with four demonstration trials.

Table 1: Crowd reactions to series of decisions

	Moderate crowd	Extreme crowd
Culturally naïve actions	Hostile, but not violent	Violent or nearly violent
Culturally aware actions	Cautious, cooperative	Hostile, but not violent

In our demonstration, we selected two distinct paths (sequences of dialog choices) through the scenario. One path was chosen to reflect a player who understands how to deal with an Iraqi crowd and community leader. The other path represented a player who was inexperienced in dealing with such a scenario and naïve about the crowd's culture. We then instantiated two versions of the playable scenario, each populated with

a different set of crowd members. One crowd was tuned to have moderate views towards Americans, while the other had more extreme (negative) views towards Americans and a higher propensity towards violence. Further, we placed a clock on the user's display and gave the human players a goal of completing the scenario in the minimum amount of time. This encouraged expedient actions over socio-culturally appropriate actions.

Our focus was on demonstrating differentiated outcomes and not on accurately modeling a specific culture or situation. Though limited to a relatively small set of gestures, the crowd avatars were able to visually show hostility through vigorous arm waving and throwing motions. Similarly, distrust was expressed through crossing arms, and agreement through nodding. Because the individuals in the crowd were parameterized differently (suggesting variations in personality and attitude), they did not execute these gestures all at the same time and produced a more realistic scene. In addition to the visibly differentiated crowd behavior that we observed through the game environment, we examined the internal state of the crowd members to verify the differentiated outcomes of each set of conditions. The average grievance state of the crowd members could be tracked throughout the short scenario. This measure of the crowd's mood could be used as a measure of the student's success in convincing the crowd to disperse.

CONCLUSIONS

Our experience with the SCALE-UP prototype showed that different, independently developed systems for generating computer-controlled behavior (i.e., PEs) could "play" in the same scenario, facilitated by a simulation framework using messages with multiple abstraction levels to communicate between characters (human and synthetic). Moreover, we were able to select people engines that were appropriate for the character's role in our scenario. While these people engines were originally developed for other scenarios and simulation systems, we integrated them into SCALE-UP with minimal customization.

As game- and simulation-based training systems become more complex, system developers, for better or worse, will be faced with the same problems that we encountered in our work. Achieving a realistic level of social behavior with convincing details, especially if it involves natural language, will inevitably lead to incorporating heterogeneous sets of PEs into such systems. These PEs will likely be developed by people

with different scientific and engineering backgrounds, and have different strengths and weaknesses. Limitations in time and resources will mean that the game framework will have to adapt to the interface limitations and requirements of these PEs, rather than the other way round.

Training systems such as the SCALE-UP prototype, which are made up of multiple people engines and a game client as the human interface, can benefit from using a physical world model outside of the game. Originally, we intended to model some aspects of the physical world that were not covered by the game itself, such as "off-stage" character movements and interactions, separately from the game's physical world simulation. However, it became clear that there were also benefits to using such separate, more abstract, physical world models to represent things that the game's world model already covered. For example, Big World™, like most game engines, represents physical locations using a coordinate system. Specific physical coordinates in the scenario map were associated with a list of more abstract notional locations, such as "front-of-crowd" or "home." The people engines selected movement actions based on the notional locations. Thus, when the absolute locations of the avatars was changed, the people engines still behaved correctly; only the mapping between absolute coordinates and notional locations needed to be updated. Off-stage interactions between people engines do not need to be visualized, but might need a sense of locality. Many of these interactions can occur through the Mindscape based on explicitly modeled relationships and associations with various affinity groups (such as family or religion). However, some are dependent on locality, for which an abstract representation provides more flexibility and less complexity than grid coordinates.

This notion of separating the social and physical world model from the game client should also allow the SCALE-UP training system to use different game clients as appropriate. For example, the Big World™ game client was suitable for the first-person scenario demonstrated, but a more strategic-level game client might be better for other types of scenarios. Since most game clients will not be designed to interface with external people engines, it is likely that each new game client will require a custom interface to the rest of the training system.

While our work focused on people engine interoperability, progress along a number of dimensions is needed to provide deep, meaningful, and diverse interactions for generalized non-kinetic-warfare training. Fortunately, there exists a lot of research being per-

formed along these dimensions. One such effort investigates the use of multimodal interactions techniques so that the learner can interact in more natural ways. Another develops deep human behavior models that exhibit recognizable cultural and personality traits. Another uses semantic representation of the world with appropriate interactions and relationships, i.e., our Mindscape.

One avenue to foster collaborative research in this area would be to define a set of challenge problems (scenarios) that could be addressed in virtual worlds. These challenge problems would help focus research in such a way that could lead to interchange standards and better comparison of various groundbreaking research in the research areas just noted. In our continued research, we plan to work towards establishing a robust communication framework and architecture that will lead to improved cooperation among PEs and, eventually, enhanced training and analysis applications.

ACKNOWLEDGEMENTS

This work was sponsored in part by DARPA contract NBCHC050067. The authors wish to thank the many project contributors from BBN Technologies, the University of Pennsylvania, CHI Systems, and Total Immersion Software.

REFERENCES

- Alexander, A., Brunyé, T., Sidman, J., & Weil, S., (2005). From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games. *The Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.
- Cass, S. (2002). Mind Games. *IEEE Spectrum*, 39(12), pp. 40-44.
- Cornwell, J., O'Brien, K., Silverman, B., & Toth, J., (2003). Affordance Theory for Improving the Rapid Generation, Composability, and Reusability of Synthetic Agents and Objects. *12th Conference on Behavior Representation in Modeling & Simulation*.
- Eidelson, R.J. & Eidelson, J.L. (2003) Dangerous Ideas: Five Beliefs That Propel Groups Toward Conflict. *American Psychologist*. 58, 182-192.
- Ferguson, W., & Leung, A., (2004). Simulation as Storytelling: Capturing Lessons Learned through User Authoring of Scenarios. *Proceedings of the European Simulation Interoperability Workshop*, June 28-July 1, 2004.

- Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Johnson, W., Vilhjalmsson, H., & Marsella, S. (2005). Serious games for language learning: How much game, how much AI? *12th International Conference on Artificial Intelligence in Education*.
- Marsella, S., Pynadath D., & Read, S. (2004). PychSim: Agent-based modeling of social interactions and influence. *Proceedings of the International Conference on Cognitive Modeling*, 243-248.
- Maslow, Abraham H., (1987). *Motivation and Personality*, 3rd ed. New York: Harper & Row.
- McAlinden, R., Clevenger, W., (2006). A Culturally-enhanced Environmental Framework for Virtual Environments. *Proceedings of Behavior Representation in Modeling and Simulation*, (BRIMS).
- McAlinden, R., Johnson, W., Vilhjalmsson, H., & Marsella, S. (2005) Serious games for language learning: How much game, how much AI. *12th Intl. Conf. AI in Education*.
- Ortony, A., Clore, G., & Collins, A., (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.
- McDonald, D., Leung, A., Ferguson, W., and Hussain, T., (2006). An Abstraction Framework for Cooperation Among Agents and People in a Virtual World. *Proceedings of Artificial Intelligence and Interactive Entertainment Conference: AIIDE-06*.
- Riedl, M., & Stern, A., (2006). Believable agents and intelligent scenario direction for social and cultural leadership training. *Conference on Behavior Representation In Modeling and Simulation*.
- Silverman, B., Johns, M., Cornwell, J., & O'Brien, K., (2006). Human Behavior Models for Agents in Simulators and Games: Part I – Enabling Science with PMFserv. *PRESENCE*, Vol. 15, No. 2, 139-162.
- Silverman, B., Bharathy, G., O'Brien, K., & Cornwell, J, (2006). Human Behavior Models for Agents in Simulators and Games: Part II – Gamebot Engineering with PMFserv. *PRESENCE*, Vol. 15, No. 2, 163-185.
- Silverman, B., & Bharathy, G., (2005). Modeling Cognition and Personality in Leaders. *Conference on Behavior Representation In Modeling and Simulation*, May 2005.
- Zachary, W., Santarelli, T., Ryder, J., Stokes, J., & Scolaro, D. (2001). Developing a multi-tasking cognitive agent using the COGNET/iGEN integrative architecture. *Proceedings of 10th Conference on Computer Generated Forces and Behavioral Representation*, (pp. 79-90).

Big World™ is a trademark of BigWorld Pty Ltd