Teaching Cross Cultural Social Competence in a Dynamic, Synthetic Environment

William Ferguson, Bruce Roberts, David Diller
Raytheon BBN Technologies
Cambridge, MA
wferguson@bbn.com, broberts@bbn.com, ddiller@bbn.com

Dan Shapiro, Michael Mateas
University of California Santa Cruz, Santa Cruz, CA
Santa Cruz, CA
dgshapiro@soe.ucsc.edu, michaelm@soe.ucsc.edu

ABSTRACT

To prevail in modern, asymmetric conflicts, most warfighters must socially engage with people of diverse cultures to accomplish a variety of military missions. In spite of this pervasive need, no scalable solution exists for training social skills. Live role-playing is prohibitively expensive and dynamic social skills cannot be learned using traditional, virtual environment based training architectures that rely on carefully scripted scenarios and statically animated, synthetic characters. This paper offers a possible solution for social training by describing an exploitation of Expressive Artificial Intelligence (AI) and an adaptation of cognitive apprenticeship to create a synthetic, mentored, social practice environment. Expressive AI views AI as an expressive medium, and aims at the algorithmic and architectural research necessary to create highly interactive and generative experiences. To allow for true social interaction, our team focused on creating combinable chunks of behavior that enable synthetic characters to participate in a wide variety of jointly meaningful social activities with each other and with a human learner. To meet the challenge of mentoring in this environment, our team borrowed from the deep teaching method of cognitive apprenticeship, exploiting techniques such modeling and scaffolding. To this mix was added real time coaching using the same social simulation mechanisms that create the synthetic characters in the simulated world. A demonstration version of the system was developed under the Defense Advanced Research Projects Agency (DARPA) Strategic Social Interaction Modules (SSIM) program, which in part is designed to illustrate training of general social competence in unfamiliar contexts rather than culture-specific knowledge and skills, using computer controlled characters and instruction in synthetic environments. This work on the design and engineering trade-offs and innovations in simulation control structure should spark interesting debate in the education and simulation communities as well as serving as the basis for others heading in this same direction.

ABOUT THE AUTHORS

Mr. William Ferguson is a Lead Scientist at Raytheon BBN Technologies with expertise in artificial intelligence, cognitive science, games and computer-based training. He is currently the Principal Investigator for the BBN IMMERSE project where the work described here was done. He also played a principal role in the DARWARS project, a DARPA funded program to revolutionize computer-based training for the military by exploiting ideas and technologies developed in the commercial gaming world. He was co-principle investigator of a project called Helical training, which exploited the experiences provided by Alternate Reality Games for pedagogical use.

Mr. Bruce Roberts is a Lead Scientist at Raytheon BBN Technologies. Mr. Roberts is responsible for the creation of tutoring and mentoring capabilities within the system. Mr. Roberts has over 30 years experience developing simulation-based intelligent tutoring systems for the DoD. He was the Principal investigator for DARWARS architecture and integration (part of DARPA's Training Superiority program), and led the rapid development and successful deployment of DARWARS Ambush!, a widely used, multi-learner, game-based training system.

Dr. Daniel Shapiro is the Director and technical lead of the IMMERSE project at the University of California Santa Cruz. He is responsible for system architecture, language design, and project management at UCSC. Dr. Shapiro has extensive experience in cognitive systems and machine learning research within AI. He created the Icarus architecture, with Dr. Pat Langley, which is one of the few extant general theories of cognition. He is currently applying ideas from cognitive systems to enhance character AI at the Center for Games and Playable Media.

Dr. David Diller is a Senior Scientist and Group Lead of the Immersive Training Technologies group at Raytheon BBN Technologies. Dr. Diller is the project manager for the BBN IMMERSE project. He has contributed to research and technology development in the areas of human behavior and cognitive modeling, simulation- and game-based training, artificial intelligence, mixed-initiative agent based systems, and simulation-based training applications. Dr. Diller has led a variety of efforts developing training and other applications for the U.S. Military, with systems deployed in the U.S. and overseas.

Dr. Michael Mateas is the director of the Center for Games and Playable Media, and Professor of Computer Science at University of California, Santa Cruz. On this project, Dr. Mateas is responsible for the design and development of Character AI, as well as leading the design of the social simulation engine. Dr. Mateas, who holds the MacArthur Endowed Chair, runs the Expressive Intelligence Studio (EIS), one of the largest technical game research groups in the world. His work explores artificial intelligence-based art and entertainment, forging a new research discipline called Expressive AI. With Andrew Stern, he created Façade, an award-winning interactive drama that uses AI techniques to combine rich autonomous characters with interactive plot control, creating the world's first, fully produced, real-time, interactive story.

Teaching Cross Cultural Social Competence in a Dynamic, Synthetic Environment

William Ferguson, Bruce Roberts, David Diller
Raytheon BBN Technologies
Cambridge, MA
wferguson@bbn.com, broberts@bbn.com
ddiller@bbn.com

Dan Shapiro, Michael Mateas
University of California Santa Cruz, Santa Cruz, CA
Santa Cruz, CA
dgshapiro@soe.ucsc.edu, michaelm@soe.ucsc.edu

INTRODUCTION

The Defense Advanced Research Projects Agency (DARPA) Strategic Social Interaction Modules (SSIM) program is designed to address the challenge of training our warfighters to successfully engage with people from other cultures in order to accomplish missions where success hinges on social competence. As part of this program, the IMMERSE project was conceived to demonstrate the possibility of conducting that training using an immersive, simulation-based training system to create a mentored, social-interaction practice environment. Technical challenges have included developing interactive synthetic social situations populated by robust, believable agents and providing a synthetic coach to deliver feedback and guidance. In IMMERSE, trainees can expect to interact with simulated characters the same way they interact with other people, through speech, gesture, comportment, and other forms of nonverbal communication.

This paper describes the work done to date from a pedagogical perspective, and describes how technology could be designed to meet the requirement for practicing social competence. IMMERSE, as of this writing, is a partially completed system capable of providing demonstrations and limited experiences to untrained users. Nonetheless, there have been valuable lessons and insights gained that can aid others who are trying to move into truly socially engaging simulation and reap the training value to be harnessed there.

Objectives

The original vision for DARPA's SSIM program was to understand, measure and teach the ability to be a "good stranger" – to interact effectively with people from other cultures without trying to be accepted as a member of that culture[1]. Because the expected SSIM learners were to come from the military and from law enforcement, SSIM focused on interactions with the potential for kinetic escalation (violence breaking out) and where the learner was often trying to "gain compliance" – get members of another culture to do something or provide information. In keeping with the good stranger vision, IMMERSE does not teach culture-specific skills, knowledge and behavior; rather, it aims to sensitize the learner to anticipate and perceive culture-specific behavior and adapt to it in the field.

Conceptually, the design of IMMERSE began with the following notion: if social competence is a skill, then it can be trained through relevant practice with feedback. Practice of real-time, embodied, cognitively rich skills is made more comprehensive by embodied interaction, real-time dynamism and socio-cognitive richness. Thus the virtual characters (VCs) that inhabit IMMERSE must be able to detect, interpret and react to the full body behaviors of the learner in real time. They must then behave as if they are part of the many overlapping social interactions that make up multi-person, social settings.

Besides this basic assumption, the design was guided by two secondary principles:

1. Performance Hypothesis: High fidelity simulation is not enough. Social action must be *performed*, as in the kind of thing actors and actresses do, in order to enable pedagogically effective social interaction.
2. Social Coaching Hypothesis: Coaching social competence is especially sensitive to the social competence of the coach.

[1] http://www.darpa.mil/Our_Work/BTO/Programs/Strategic_Social_Interaction_Modules_SSIM.aspx

The IMMERSE effort under SSIM is an engineering effort aimed at building a prototype rather than a validation-based experimental effort. Therefore, the guiding principles will not be validated (or disproven) by this work. Rather one hopes to show how a system can be designed and built which effectively exploits these hypotheses (and which could enable experiments to test them). That said, there is immense benefit to be gained if systems based on these conjectures and the IMMERSE prototype can be built and fielded – the simplest being the existence of a scalable technology for delivering social training, given that the current best practice—human role-play based training—is expensive, logistically cumbersome and prone to certain systematic weaknesses such as role player fatigue.

SYSTEM OVERVIEW

Interacting with the system is meant to be a natural experience for the learner. The learner stands in front of a large display, such as a 70" screen, equipped with a microphone and a Kinect 2 sensor. (Figure 1 illustrates use of the system). The learner is free to move around in front of the display within a 7 ft. deep by 11 ft. wide region, as constrained by Kinect 2 sensor. A noise cancelling, wireless microphone is used to overcome background noise, but the Kinect 2 sensor also offers a built-in microphone array, which can be utilized to reduce hardware requirements. Depending on the scenario, the learner may employ a simulated weapon that contains an embedded game controller.

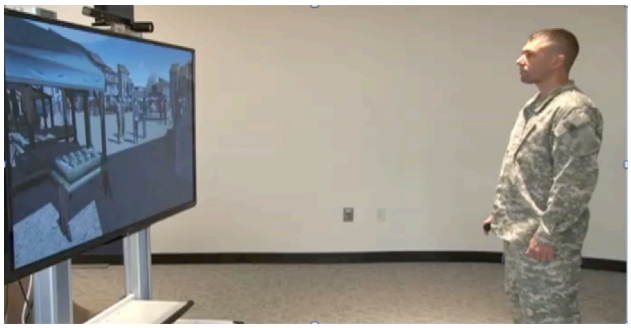

Figure 1. Illustration of use of the IMMERSE system

A typical session with a learner involves a short mission briefing to set the stage for the scenario, followed by a short system configuration phase, involving capturing a neutral facial expression of the learner, after which the learner moves into the scenario, which lasts on the order of two to ten minutes. During the scenario, the learner interacts with social virtual characters, including a virtual coach, which provides guidance and feedback as appropriate. After the scenario, the system provides an after action review (AAR) describing correct and incorrect actions and what the learner might do to perform better in the scenario.

THE PLAYABLE EXPERIENCE

The screen comes to life showing a kitchen scene where two people stand. On the left is a local national solder (LNS) whom the learner is supervising in a house search for contraband medical supplies. The man of the left is the head of this household (HoH); almost before the learner can react he picks up a bowl of fruit, turns toward the learner, takes a few steps forward and, smiling holds out the bowl. The learner decides he is being offered fruit and reaches a hand tentatively forward. One of the green orbs in bowl vanishes and the HoH nods. He then turns back toward the LNS, perhaps to offer him fruit as well; but the LNS has picked a slice of bread up off the kitchen table and is examining it. The HoH shouts at him and takes a step toward the LNS. The LNS looks up from the bread and leers at the HoH, moving the bread back a forth slowly. The HoH glances back at the learner, who does nothing, and then turns and begins to gesticulate wildly. The LNS extravagantly take a large bite out of the bread and the HoH begins shouting. The learner says "Calm down!" and waves his hand vertically. The HoH turns and gives the learner a puzzled, agree look. The LNS laughs and the learner shouts, "Stop that!" at him. The LNS looks uncertainly at the learner. The HoH picks up an empty plate and waves it at the LNS beginning to shout again. The learner takes a step forward and reaches out to him, but the HoH throws the plate to the floor shattering it. Startled, the LNS raises his weapon slightly (to a "low ready"). Just then another character enters the kitchen from a side door: an unknown young man wielding a shop hammer. He sees the LNS and raises the hammer. . .

What a mess! Can the learner keep violence from breaking out? More important, could he or she have acted differently and never gotten into this situation? This is what interacting with the IMMERSE simulator is like. The action can be fast or slow, dangerous or humorous, mission focused or exploratory but it is always centered around human interaction and it always puts the learner in a position where he or she can and must figure out, what is going on, what can be done and how to move things in a desirable direction while minimizing social damage.

SIMULATING A DYNAMIC SOCIAL CONTEXT

The goal of the IMMERSE system is to provide engaging social experiences between people and virtual characters. This is a significant challenge since it calls for imbuing virtual characters with a social presence that they have almost entirely lacked to date[2]. That challenge can be decomposed into three component issues:

1. Virtual characters need social intentions that they can employ to generate behavior.
2. Virtual characters need to interpret the actions of learners in social terms.
3. Virtual characters need to communicate social intentions, actions, and state to learners in a poignant form.

The first two issues concern behavior generation and recognition. These are addressed them in tandem by embedding a model of *social games* into the IMMERSE system. A social game contains a representation of social relations, motivations and moves, and a means of employing them to choose character behavior. Social games also contain a mechanism for sensor interpretation that extracts social signals from observed learner actions. The third issue concerns communication, where the goal is to send the right kinds of social signals to the learner. It is an issue of expressive performance. This is addressed it by writing character behaviors in a language designed for Expressive Artificial Intelligence, ABL, that facilitates nuanced control over the timing and content of character actions, plus run-time prioritization and interleaving of performance requests. The social game and expressive performance technologies within IMMERSE allow non-scripted interactions, where social experiences emerge from the interplay between the characters' and the learner's social moves. Social games and ABL are discussed below. Shapiro et al. (2013) provides additional detail on the IMMERSE technology, utilized in a different scenario.

Social Games

A social game defines a family of practices surrounding specific social state. A concrete example is the *Authority Game*, which was one of several active in the "Breaking Bread" scenario described above. The authority game tracks permission levels to use objects (to observe, touch, own), relations among people, and whether authority norms are being violated. It represents motivations to increase or express authority, and it relates those motivations to a variety of social moves, e.g., to manipulate someone's belongings, to order them to take action (or stop taking an action), even to give a dirty look or issue a derisive chuckle. Executing those moves impacts social state. For example, when the LNS ate the bread belonging to the HoH, the alliance game recorded a violation of an object norm.

The architecture for social games contains a mechanism for selecting motivations and choosing social moves. This process is called *intent formation*, as shown in Figure 2. It occurs periodically, and is implemented by ranking rules. For example, the motivation selection rule

$$Arrogant(X) \wedge Authority(Y,X,<50) \wedge Authority(X,Y,<80) \Rightarrow AuthorityUp(X,Y)+6$$

increases the desirability, to character *X*, of upping its authority over character *Y*. The rule matches in a situation where *X* has the Arrogant trait, *Y* has comparatively little authority over *X*, and *X* has anything less than a dominating authority over *Y*. Multiple rules of this kind interact to produce a final score for *AuthorityUp(X,Y)*. For example,

$$Permission(X,Obj,Owns) \wedge Permission(Y,Obj,LookAt) \wedge Touch(Y,Obj) \Rightarrow AuthorityUp(X,Y)+8$$

adds further volition to *AuthorityUp(X,Y)* when *Y* has touched something belonging to *X*. In Breaking Bread, these rules motivate the HoH (as *X*) to gain authority over the LNS (as *Y*). More broadly, the intention formation process runs all such rules for all possible character combinations, resulting in a score for all situation-relevant motivations. It selects all motivations above a certain threshold.

[2] For example, a video game character typically has the social presence of a menu item – it produces a social response (e.g., the bartender reports a rumor) given a specific cue (ordering a drink). The character remains inactive otherwise, even if surrounded by combat.

Intention formation employs a similar calculation to rank and choose social moves that pursue social motivations. For example, the rule

$$Volition(AuthorityUp,X,Y,Val) \wedge Permission(Y,Obj,Owns) \wedge Permission(X,Obj,<Touch) \Rightarrow ViolateObjNorm(X,Obj,Val)$$

retrieves the volition associated with *AuthorityUp(X,Y)*, and assigns it to a move that will cause *X* to violate *Y*'s sensibilities with respect to an object that *Y* owns. That move is expressed generically at the level of social games, and will be resolved into a situation specific behavior at the execution stage (such as walking to the object and picking it up). Intention formation ends by passing all moves above threshold to execution.

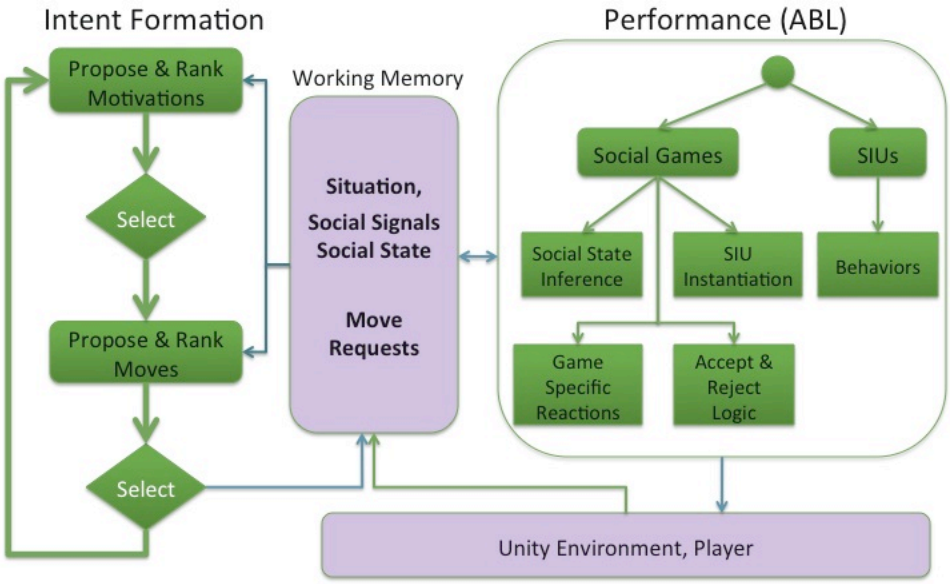

Figure 2. Social Game Architecture

Three social games were implemented for Breaking Bread: *Alliance*, *Authority* (described above), and *Threat*. The alliance game tracks the degree to which a learner and a virtual character (or two VCs) are on the same side. It contains motivations to increase or express alliance, while its moves concern hospitality and rendering assistance. The threat game tracks *danger*, *might*, and *face*. Its motivations are to increase, decrease or express a threat, and its moves concern acknowledging, producing, and responding to threats. Social games interact through social state. For example, some Alliance, Authority, and Threat moves change character alliance levels, while that level impacts volition for certain motivations and moves in all three games. Because of this interaction, social games can be combined to produce emergent behavior. *Alliance* + *Authority* + *Threat* generates a different social trajectory through interaction with the learner than employing the Alliance game alone.

Every social game and move has an executable representation that generates character performance. Social moves map onto programs, written in ABL, that run character animations and react to learner input. They are called Social Interaction Units (SIUs) in Figure 2. Social games also have persistent elements that give games the aspect of mindsets. For example, as long as the Threat game is active for a given character, it will have a strong startle response primed and ready to run in the event of a threat signal. The character will also have a sensor interpretation process that looks to infer instances of threats, possibly by aggregating signals over time. Outside of the threat context, these considerations are absent.

The final durative component of a social game is the accept/reject logic mentioned in Figure 2. This logic consists of ranking rules that choose game-specific responses to learner moves by prioritizing local options. For example, within the Authority game, the HoH experiences a choice between the options to obey the learner's "Stop that" command, comply after a fuss, or to actively refuse to comply.

The ABL language provides a variety of facilities for implementing SIUs (Mateas & Stern 2002; 2005). It is an agent-oriented language that distinguishes operations for sensing, acting, and pattern matching onto a local memory. It supports hierarchical expression of executable behaviors in terms of goals and subgoals, with parallel or sequential structure. In the IMMERSE system, behaviors bottom out into sensor signals inferred from learner gestures or obtained from the Unity environment, and into actions that run animations in Unity.

Because the social game architecture generates moves while SIUs are running, the performance system must respond to multiple, asynchronous, unscripted SIU execution requests. This creates a need for run-time control over parallel and interleaved performances, as well as a demand to author interruptible and composable SIUs. A number of coding abstractions and mechanisms were developed to manage this task, including monitors for learner actions, symmetric SIUs (where the learner or a character can perform any role), priority based behavior selection, run-time allocation of resources to behaviors, multiple resource layers (e.g., for head, neck, torso), and wrap-ons that apply attitudes to action (enabling them to be coded separately). Shapiro et al. (2013) provides detail on several of these mechanisms.

In summary, the architecture for social games enables social experiences with virtual characters, while social games combine to produce more complex behaviors. These mechanisms were employed to produce Breaking Bread, which is a playable social experience.

TRAINING IN A DYNAMIC SOCIAL CONTEXT

In IMMERSE, social skills are fostered through guided practice within a socially rich virtual environment that presents varied, authentic social encounters. A social coaching component supplements the experiences available in the social simulation to produce more effective training as well as more efficient training, an important consideration for the eventual integration of social competency training into military and law enforcement training regimens.

Designing Social Encounters

The learning experience is not the province of any single system component, but arises from the interactions of game play; the social simulation's ability to generate believable and engaging characters, behaviors and situations; and the pedagogical component's ability to deliver timely, apt instructional guidance. IMMERSE creates social contexts that go well beyond any structured scenario approach to experience design. VCs are simultaneously engaged in multiple behaviors, possess different social traits and able to initiate interactions with the learner.

Writing robust behaviors characteristic of specific social settings (e.g., approaching strangers to enlist their aid in locating a missing person) has the beneficial side effect of creating natural forms of scaffolding for social interaction. This derives from the common ground inherent in all social encounters that presupposes both parties are genuinely involved and trying to communicate their intent. This can lead naturally to supportive behaviors being introduced, which invite the learner to participate; e.g., asking to see a picture of the missing person. These behaviors become pedagogical affordances made available to a coaching component, and play the role of guiding the learner to desired courses of action. Another characteristic of social encounters – prevalent in cross-cultural settings – is the unavoidable confusion or trouble that can arise within a discourse. Recognizing trouble and collectively seeking remedy is integral to social competence.

The Social Simulation as a Training Vehicle

The social simulation is more than just a practice environment, and includes the following features:

- "Sets up the problem" – confronts the learner with specific challenges.
- Provides scaffolding in the form of performance actions and modes that make affordances and consequences more clear and poignant than they might be in a natural setting.
- VCs can provide implicit modeling of efficacious (and counterproductive) behavior.
- Increases immersion and motivation to continue by using story and performance to draw the learner in.

Critical Junctures

Critical junctures are points and intervals in an unfolding social situation where the learner's diagnosis of the situation and course of action (or inaction) has a large impact of the outcome of the situation. Expressed as desired

social sate and co-occurrence of SIUs; not the same as dramatic peaks; brought about by the human designers of the scenario and supported by the pedagogical elements of the system. In the example from the introduction of this paper, the learner did nothing when the LNS first started to provoke the HoH? Perhaps without knowing it, he was at a critical juncture. He needed to:

- Understand what was happening well enough to intervene
- Realize some of the meaningful actions he could take
- Focus of balancing 1) persuading the LNS to desist in in his provocation and 2) soothing the HoH

Of course the learner may not even have realized he could or should do anything at all at that point: he may be unaware of the "social affordances" available in the situation. Social Affordances are social behaviors that will have specific effects at specific times. They can be thought of as options or opportunities that the situation makes available. In this case the learner's his authority over the LNS gives him the ability and obligation to intervene, while the HoH's offer of fruit suggests an option to pursue a connection with the HoH.)

The Social Coach

Instructional design in IMMERSE involves guiding the learner through situations where social skills matter and consequences are visible. The SSIM program posed this question to a cadre of social scientists in this form: What makes a warfighter a "good stranger", able to create positive social outcomes within the context of other mission objectives, even in the face of unfamiliar cultural contexts and conflicting social norms? Their findings are summarized by these cross-cutting skills: interpreting non-verbal cues, perspective-taking, managing the encounter, building rapport, maintaining self-control, maintaining self-awareness and recognizing social affordances. Social encounters in IMMERSE are designed to foster these skills; e.g., observing and adopting situationally appropriate rituals like greeting.

IMMERSE is a guided practice environment. The forms of guidance derive from the tenets of Cognitive Apprenticeship (Collins et al., 1989; Collins, 1991), which offers a compelling argument for modeling learning on traditional forms of apprenticeship – coaching and modeling behaviors to guide the learner experience in the virtual environment, providing feedback, scaffolding and fading, and a graduated exposure to increasingly difficult, complex and varied situations. However, the dynamic and dramatic qualities of the social simulation required extending traditional notions of guidance and feedback usually associated with coaching. If one theorizes that immersion in a social discourse is a prerequisite for learning, the coach must respect the dramatic flow of the interaction by not undermining learner engagement. To avoid being a distraction, the social coach must itself respect conversational conventions; e.g., through implicit turn-taking with the learner and other interlocutors. The earlier discussion of junctures also affects the role of feedback, and because learning social competency involves behavior change, exposure to extreme situations and failure may be an essential ingredient in the feedback process.

Coaching includes speaking to the learner as well as altering the visual scene by zooming the camera to focus attention on a particular VC behavior, or other forms of highlighting elements in the scene. These forms of explicit coaching are augmented by the implicit coaching inherent in the behaviors of the VCs. The rich social simulation and expressive performance capabilities makes possible natural forms of guidance and feedback. VC facial expressions, gestures and other embodied behavior can vividly reveal short and longer-term consequences to learner actions (or inaction). Moreover, VCs can model, actively solicit or suggest actions for the learner.

Despite concerns about in-game coaching interfering with immersion, there are situations that warrant such interventions. For example, when a learner reaches an impasse, the coach can suggest a course of action to keep the encounter moving forward. A recently added ability to pause the entire simulation will allow the use of such a "tactical pause" to reorient and scaffold a struggling learner. Unlike the experience described above, in which the interplay among characters is non-stop, other scenarios are more episodic – going from group to group in a village square to locate a missing person – and present natural gaps in the story that allow the coach to jump in.

Pedagogical Modes

IMMERSE currently supports two pedagogical modes: a free-play mode called the Sandbox and a story-driven mode exemplified by the "Breaking Bread" scenario described above.

Sandbox
The Sandbox experience places the learner in a village square confronted by two or three VCs whose main objective is to interact with the learner and to draw the learner into the experience. For example, VCs may engage in the following behaviors:

- Mirror some learner behaviors: smiles, frowns, bows, waves and head nods
- React to other learner behaviors in simple ways; e.g., look in the direction the learner points, approach when beckoned, or startle if the learner makes a loud vocalization
- If the learner is idle, they demonstrate a behavior and try to get the learner to repeat it back
- Engage in simple social exchanges; e.g., at the learner's request pick up small items in the scene give them to the learner. The VCs can get into mild disputes with each other about items, some wanting to control the items and objecting if others move them.

The purpose of the Sandbox is three-fold: 1) to invite learners to interact with the system, 2) to familiarize learners with behaviors that the virtual characters can recognize, and 3) to serve as a low-key practice environment for simple social interactions. The highly reactive and playful nature of the VC reactions, along with the gentle pace of the experience – nothing happens if the learner takes a long time to do things except for attempts to introduce them to new behavior – provides a very low engagement barrier as well as the gratification of nearly instant social feedback. A stealth goal for the Sandbox is to introduce a pallet of embodied communication acts that can be employed in later story-based training. It also rewards fundamental meta-cultural skills: observation, mimicry, engagement, and calmness in the face of trouble that inevitably arises in social interaction. The Sandbox experience progresses from simple mirroring of behaviors through simple decontextualized social games to practicing elemental social skills. It serves as a form of social part-task trainer and is able to introduce the learner to some of the dimensions along which social norms differ (e.g., proxemics).

Story-driven Experience
The story-driven experience situates the learner in a rich social context with a comprehensive backstory, conflicting goals, multiple outcomes, and high stakes. Experiences are designed to encourage the learner to display good stranger skills in order to succeed, to demonstrate competency under pressure, and to balance tact and tactics – concern for the social demands of the moment, short and long-term missions goals, team and individual safety. Doing so requires trade-offs and an ability to adapt (sometimes rapidly) to events and take advantage of available social affordances.

Because of the underlying social game representation, these experiences are inherently highly replayable. They invite repeated engagement and exhibit high levels of variability from the learner's perspective due the learner's own actions, initial conditions for the social states of the VCs and suggestions from a Pedagogical Director (described below). While visible consequences are the most potent form of feedback, the social coach contributes more additional focus both during and after each training experience.

The Breaking Bread scenario (See Figure 3) has been used to illustrate portions of this paper, however it is but the latest in a progression of scenarios devised to drive the technology development, the sensing and social modeling and the learner experience. Earlier scenarios involved searching for a missing person in a town square and manning a checkpoint at the entrance to a food distribution center.

Figure 3. Screenshot of Escalated Situation in IMMERSE

SYSTEM ARCHITECTURE

Figure 4 shows the top-level system architecture. The *learner*, shown at the far left in Figure 4, interacts with the system in the same way the learner would interact with people, through speech, gestures, comportment, and other forms of nonverbal communication. Learner input into the system is handled by the learner-sensing system, *MiBA (Multimodal Integrated Behavior Analysis)*, which takes input from a Kinect 2 sensor and microphone and analyses the learner's gaze, facial expressions, body pose, gestures, speech, speech vocalics, and multimodal affect and converts this input into a rich set of possible social interpretations. These signals are passed along to the *Social*

Pragmatics (SP) module, which performs the last stage of input interpretation, and thus takes its name by analogy to "pragmatics" which is the name given to the last stage of classical natural language interpretation. The SP module has three subcomponents:

1. Simple, rule-based conditioning and filtering of the signals
2. Routing the signals to appropriate VCs (via learner focus and the perception topology of the virtual space – i.e., who can see and hear whom.)
3. Agent-specific, rule-based interpretation of the social signals into affordances

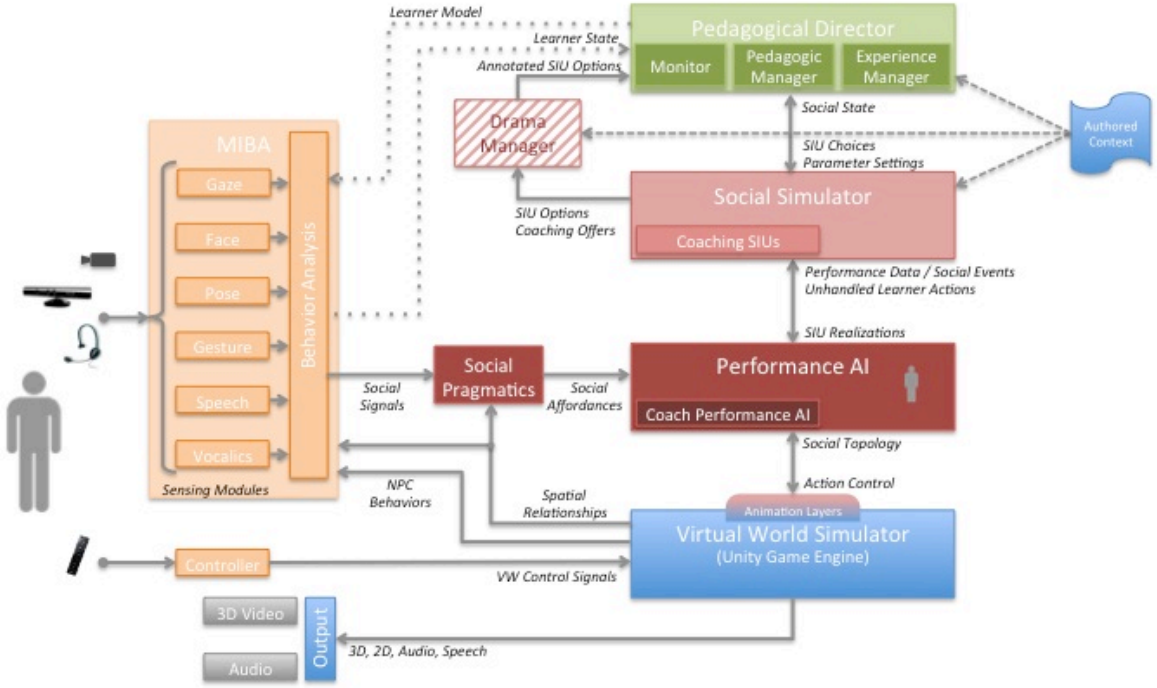

Figure 4. System Architecture

All of the system output seen and heard by the learner is presented by the *Virtual World Simulator*: the game engine rendering the VCs and the simulated environment. The *Social Simulator* (described, at length, above) is responsible for creating a rich social environment for learning, and models the high and intermediate level activities of the learner and the VCs. The *Pedagogical Director* (*PD*) is responsible for assessing the learner and influencing the learner's experience to provide an effective, individualized and varied training experience.

The PD receives SIU options annotated with a social volition score (from the Social Simulator) and a dramatic score (from the *Drama Manager*). The PD uses its learner model and a set of authored tailoring strategies to develop pedagogic goals against which the SIU options are evaluated. It then seeks to balance the sometimes aligned and sometimes competing goals of pedagogy, drama and social realism. The PD consists of three primary sub-components: the *Monitor*, the *Pedagogical Manager*, and the *Experience Manager*. The Monitor is responsible for understanding and assessing the learner's actions in the context of the current situation. The Pedagogic Manager is responsible for updating the learner model and selecting the pedagogical interventions, based on authoring tailoring strategies, based on the learner's current needs. The Experience Manager is responsible for selecting specific elements of the learner's experience (SIUs or other experience influencers) based on multiple factors.

The Performance AI module is responsible for creating the behaviors of the VCs and the coach. It receives fully annotated SIUs from the PD and works out the joint behaviors that the VCs need to perform to execute those activities. It handles conflicts over timing and the use of particular physical elements of the world (e.g., places to stand, body parts, face parts, objects in the world, etc.) It notifies the social simulator when actions result in changes to the social state. The module also serves as the gateway for learner actions passing into the simulator. If these actions are expected continuations of social activities already in progress, (e.g., a smile back at a smile made by a

VC) then they simply allow the activity to continue. If the learner actions are not part of an on-going activity, they are passed back up to the social simulator, which will try to interpret them as initiations of new activities.

Coaching, which is very much a social interaction itself, is implemented using the mechanisms of the Social Simulator and the Performance AI. Coaching options are generated and weighted (as are other intrinsic social activities) in the simulator. They are enacted in the performance engine so that they can be coordinated with other actions of the system (e.g., The coach will not speak over important VC dialog, etc.) Passing coaching actions through the main loop also allows coaching behavior to be gated and parameterized by the PD. The coach can express a concern by suggesting a high level coaching SIU to handle it. The PD can accept or reject this SIU and then, if it accepts modify it with its guidance to the coach.

RELATED WORK

For more than fifteen years, the military training community has been involved in the research and development of VCs designed to mimic human behavior – looking, acting, and responding through language, as might a real person (e.g., Rickel & Johnson, 1999; Kenny et al., 2007). While the social realism of virtual characters has at times been a necessary component of those efforts, a focus on training social skills using virtual social characters is relatively new. Building on work originally funded by DARPA, Johnson (2010, 2014) was one of the first to seriously explore creating compelling, culturally aware virtual characters for use in training. A number of researchers are conducting research into social skills training related to speaking/presentation skills. Hoque et al. (2013) developed a job interview training system, MACH (My Automated Conversation coacH), which senses and interprets the nonverbal behaviors (e.g., smiles, head nod/shake, and speech prosody) of a job interviewee and provided feedback after the simulated interview session. In similar work, the TARDIS project (Anderson et al., 2013), funded by the European Union FP7, is also focusing on job interview training. To date, much of that work (e.g., Chollet et al., 2014; Jones et al., 2014) has focused on developing virtual characters and audiences that dynamically respond nonverbally to provide appropriate and useful feedback to the learner. Similarly, Batrinca et al. (2013) has assessed whole-body behaviors, and vocal prosodics for use in creating dynamically reactive crowd behavior for public speaking training. Work in developing virtual patients for medical training (see Poulton & Balasubramaniam, 2011 for a review) has also touched on social skills development (e.g., Johnsen et al., 2007; Talbot et al., 2012), unsurprising since building rapport with patients is a critical skill, in addition to diagnostic acumen. Finally, a number of commercial computer games have been developed that have as a core gameplay element the social interactions between the player and VCs. Perhaps best known is The Sims franchise (Nutt, & Railton, 2003), but others include Supple[3] and Versu[4].

CONCLUSION

The Sandbox and the Story-driven Experience were demonstrated in May 2014 at a DARPA SSIM Principal Investigator's meeting. A small number of prototype deployments are planned for the Fall 2014. IMMERSE is intended to provide experiences with synthetic characters that are subjectively and qualitatively richer and more compelling than any created to date. If it continues to be successful, the IMMERSE project will advance the state of the art of synthetic training for social competence by demonstrating that correctly combining Expressive AI, dynamically combinable content, cognitive apprenticeship, and whole body sensing can enable new simulation capabilities. An important takeaway is that supporting real-time, embodied social interaction to create a subjective sense of social engagement requires considerably more powerful simulation mechanisms and richer content than is required by turn-based systems or branching narratives. Without this engagement—and the spontaneous learner behaviors it promotes—contemporary training systems fall short of providing critical practice of social skills. The emerging, profound need for training cross-cultural social competence requires a scalable solution, suggesting that efforts like the SSIM simulator deserve continued development.

ACKNOWLEDGEMENTS

This work is sponsored by the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred. This project was conceived by former DARPA Program Manager Brian Lande, who deserves special thanks for his inspiration

[3] http://supplegame.com/
[4] http://versu.com/

and support, and nurtured by DARPA Program Manager William Casebeer, who took up the mantle of social science research and this project at DARPA. The IMMERSE project team includes Raytheon BBN Technologies (integration, the virtual environment, and coaching), University of California, Santa Cruz (social simulation, behavior modeling, and performance), SRI Princeton (multi-modal learning sensing) and SoarTech (pedagogical guidance and learner modeling). The team gratefully acknowledges the contributions of other program co-contractors, without whom this project would not have achieved what it has.

REFERENCES

Anderson, K., André, E., Baur, T., Bernardini, S., Chollet, M., Chryssafidou, E., ... & Sabouret, N. (2013). The TARDIS framework: intelligent virtual agents for social coaching in job interviews. In *Advances in Computer Entertainment* (pp. 476-491). Springer International Publishing.

Batrinca, L., Stratou, G., Shapiro, A., Morency, L. P., & Scherer, S. (2013, January). Cicero-towards a multimodal virtual audience platform for public speaking training. In *Intelligent Virtual Agents* (pp. 116-128). Springer Berlin Heidelberg.

Chollet, M., Sratou, G., Shapiro, A., Morency, L. P., & Scherer, S. (2014, May). An interactive virtual audience platform for public speaking training. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (pp. 1657-1658). International Foundation for Autonomous Agents and Multiagent Systems.

Collins, A. (1991). Cognitive apprenticeship and instructional technology. In L. Idol & B. F. Jones (Eds.) *Educational values and cognitive instruction: Implications for reform (*pp. 119-136), Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hoque, M. E., & Picard, R. W. (2014). Rich Nonverbal Sensing Technology for Automated Social Skills Training. *Computer*, *47*(4), 28-35.

Hoque, M. E., Courgeon, M., Martin, J. C., Mutlu, B., & Picard, R. W. (2013, September). Mach: My automated conversation coach. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (pp. 697-706). ACM.

Johnsen, K., Raij, A., Stevens, A., Lind, D. S., & Lok, B. (2007, April). The validity of a virtual human experience for interpersonal skills education. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1049-1058). ACM.

Johnson, W.L. (2010). Using immersive simulations to develop intercultural competence. In T. Ishida (Ed.), Culture and Computing, LNCS 6259, 1-15. Berlin: Springer-Verlag.

Johnson, W.L. (2014). Using Virtual Role-Play to Prepare for Cross-Cultural Communication. Proceedings of the *5th International Conference on Applied Human Factors and Ergonomics (AHFE*), Kraków, Poland 19-23.

Jones, H., Chollet, M. Ochs, M. Pelachaud, C. Sabouret N. (2014, May). Expressing social attitudes in virtual agents for social coaching. In *Proceedings of Autonomous Agents and Multi-Agent Systems (AAMAS'14)*.

Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S., & Piepol, D. (2007, January). Building interactive virtual humans for training environments. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* (Vol. 2007, No. 1). National Training Systems Association.

Mateas, M., & Stern, A. (2002). A behavior language for story-based believable agents. *IEEE Intelligent Systems*, *17*(4), 39–47.

Mateas, M., & Stern, A. (2005). Structuring Content in the Façade Interactive Drama Architecture. *Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2005)* (Vol. 3). Marina del Rey, CA.

Nutt, D., & Railton, D. (2003). The Sims: Real life as genre. Information Communication & Society, 6(4), 577-592.

Poulton, T., & Balasubramaniam, C. (2011). Virtual patients: a year of change. *Medical teacher*, *33*(11), 933-937.

Rickel, J., & Johnson, W. L. (1999, July). Virtual humans for team training in virtual reality. In *Proceedings of the ninth international conference on artificial intelligence in education* (Vol. 578, p. 585).

Shapiro, D., McCoy, J., Grow, A., Samuel, B., Stern, A., Swanson, R., Treanor, M., and Mateas, M., *Creating Playable Social Experiences through Whole-body Interaction with Virtual Characters* (2013). *Artificial Intelligence and Interactive Digital Entertainment (AIIDE-13)*, Boston, MA.

Talbot, T. B., Sagae, K., John, B., & Rizzo, A. A. (2012, January). Designing Useful Virtual Standardized Patient Encounters. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* (Vol. 2012, No. 1). National Training Systems Association.