International Journal of Artificial Intelligence in Education 19(2009) 289-308
IOS Press

BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context

Julia M. Kim, Randall W. Hill, Jr., *Institute for Creative Technologies, University of Southern California, Marina del Rey, CA, USA*
{kim, hill}@ict.usc.edu

Paula J. Durlach, *U. S. Army Research Institute for the Behavioral and Social Sciences, Orlando, FL, USA*
paula.durlach@us.army.mil

H. Chad Lane, **Eric Forbell, Mark Core, Stacy Marsella, David Pynadath**, *Institute for Creative Technologies, University of Southern California, Marina del Rey, CA, USA*
{lane, forbell, core, marsella, pynadath}@ict.usc.edu

John Hart, *U.S. Army Research, Development, and Engineering Command Simulation and Training Technology Center, Orlando, FL, USA*
John.HartIII@us.army.mil

Abstract. Negotiation skills are essential in everyday life, whether in a professional or personal context. Negotiation enables two parties to address misunderstandings and avoid conflicts through an exchange that depends as much on the interpersonal skills of the negotiators as the tactics employed. Acquiring these skills requires not only sound conceptual knowledge but also practice and mentoring. This paper describes the BiLAT game-based simulation and tutoring system developed to provide students, initially United States Army soldiers, with an environment to practice preparing for and conducting bilateral negotiations. We describe the models that were created to implement BiLAT, with a particular focus on the challenge of designing for and tutoring in the ill-defined domain of negotiation. An initial assessment of the training effectiveness of the system indicates significant situation-judgment gains by novices.

Keywords. negotiation, intercultural communication, ill-defined domains, intelligent tutoring systems, evaluation, narrative-based learning environments, pedagogical agents

INTRODUCTION

In today's globally connected world where relationships and networks span continents, the ability to effectively negotiate with people of different cultures is increasingly important. Yet, there are a number of challenges to helping novices become proficient negotiators. These challenges are due in part to the fact that cross-cultural negotiation is an ill-defined domain.

1560-4292/09/$17.00 © 2009 – IOS Press and the authors. All rights reserved

First, many books and courses are available to train negotiation skills, but these materials are limited in their ability to provide practical instruction. As with other ill-defined domains, negotiation requires a task structure that is highly subjective to context (e.g., time, counterpart, desired outcomes) (Lynch, Ashley, Aleven & Pinkwart, 2006). Negotiation tends to require strong interpersonal skills (the "art") to complement the specific analytical techniques (the "science") of negotiation (Raiffa, 1982). These interpersonal skills and the understanding of negotiation context are types of practical knowledge learned primarily through experience but teachable through the use of case scenarios and live role-play exercises (Sternberg et al., 2000). When this instruction is available, it tends to be limited by its instructional capability, cost, or both. The negotiation training that exists, typically found in business or law schools and corporate professional development programs, tends to be generic with scenarios and feedback that are difficult to tailor. A practice exercise usually requires a human role-player. This limits the opportunities for repeat and follow-on training as well as throughput—the number of people who can receive the training at any given time. The exercise is also subject to the role player's skill. In cross-cultural training, the actor needs to be good enough to reflect culturally appropriate behaviors and manipulate the situation as the confederate to support learning.

Second, even if a practice environment were to exist, it is highly complex for a tutor (human or computer) to provide feedback. As with other ill-defined domains, negotiation principles and techniques are both difficult to verify (did this person agree with me because I applied specific principles or because he was in a good mood) and are usually made up of overlapping sub-problems (e.g., the need for culturally-aware interpersonal behavior except in some cases where breaking cultural norms is appropriate in order to apply other negotiation techniques such as threats) (Lynch et al., 2006). As the ITS community is already aware, unguided experience, even in a practice environment, is not the best or more efficient means of learning (Mayer, 2004; Kirschner, Sweller & Clark, 2006). Coaching and tutoring can accelerate the speed of learning by providing explicit instruction, hints and other forms of feedback that enable the student to learn successful strategies (Bloom, 1984; Anderson, et al., 1995). However, an intelligent tutor for negotiation must address the aforementioned complexities of an ill-defined domain.

The USC Institute for Creative Technologies (ICT), in collaboration with organizations within the U.S. Army (Research, Development, and Engineering Command Simulation and Training Technology Center; U.S. Army Research Institute for the Behavioral and Social Sciences; U.S. Army Research Lab Human Research and Engineering Directorate), has attempted to address these challenges through the development of the Enhanced Learning Environments with Creative Technologies (ELECT) Bi-Lateral Negotiation (BiLAT) prototype application. BiLAT provides negotiation training, initially for U.S. Army soldiers, but with the intent of providing a general training tool for negotiation.

BiLAT addresses the challenges listed above by providing a game-like simulation where students can practice their negotiation skills in a safe environment on real world scenarios. The issue of limited resources is addressed through the use of synthetic role players—these are animated characters that respond to the negotiation moves of the student. Unlike human role-players, they are always available, and the system can be run on a laptop computer. Coaching is provided in situ and, at the end of each session, a reflective tutor takes the student back through the high and low points of a meeting. Both the coach and tutor provide feedback and instruction on the finer points of the negotiation driven by the learning objectives and negotiation principles drawn from domain experts.

One kind of ITS identified for use with ill-defined domains is a model-exploration system using discovery learning. The implementation and success of BiLAT suggests, however, that there should be a hybrid category for ill-defined domains that combines a model-exploration system – a category

sometimes known as simulation – with guided feedback that leverages model-based tutoring techniques. The work with BiLAT further suggests that game-based instructional systems – sometimes known as serious games – could be well-suited as a platform for these kinds of solutions.

The paper is organized as follows. First, the principles of the negotiation domain that underlie the design of the practice environment are discussed. In the second section the student experience is described, illustrating the fundamental flow of the experience as well as giving a specific example. In the third section the agent models that drive the animated negotiation partner are described. The fourth section describes the tutor. The final section provides the results of the initial assessment.

DEFINING THE NEGOTIATION DOMAIN

For the BiLAT project, a number of techniques were combined to access the negotiation domain. First, the negotiation literature was surveyed, with particular focus on the principles found in Fisher and Ury's *Getting to Yes* (1991). Second, since the military was the target audience for this application, a cognitive task analysis (CTA) was conducted with Army officers considered expert negotiators and appropriate models for novices. The results of the CTA were compared to the principles drawn from literature and found to be consistent. The literature review and CTA were then broken down into two parts: (1) principles that were important for exercising good judgment and decision-making, and (2) procedural steps that supported the principles and increased the chances of success.

One key principle identified was negotiating based on one's larger interests (what one really wants) rather than specific positions (how one can get what one wants) in order to achieve a "win-win" outcome, also known as an integrative negotiation strategy. A common mistake in negotiation is focusing on a specific position when there are other positions that could further one's larger interests and be more acceptable to one's counterpart. For example, one's interest might be to increase a town's security. One position for achieving this interest is to have one's counterpart conduct security patrols. Another position that would achieve the same interest is to conduct joint patrols. If one's counterpart is unwilling to agree to the former position but would consider the latter position, the negotiator must consider whether this is an acceptable position since the interest will be achieved.

Experts were also found to use a set of procedures that increased the likelihood of (though did not guarantee) success (Fisher & Ury, 1991; Wunderle, 2007). These procedures fell into three general stages: preparation prior to a negotiation, negotiation, and follow-up after a negotiation. The process is broken down further into twelve steps: (1) review current mission objectives, (2) determine with whom to meet, (3) identify your intended outcome for the meeting, (4) conduct research on your meeting partner, (5) organize your research and develop meeting strategies in the prep worksheet, (6) obtain prior approval for resources or other negotiation offers, (7) rehearse, (8) conduct social opening, (9) engage in business, (10) conduct social close, (11) conduct an after-action review, and (12) follow-up on your promises.

Because the system was initially designed for U.S. Army officers preparing to go to the Middle East, cultural considerations were also investigated. The intent was not to reduce culture to a set of rules but rather to consider the sensitivities of culture when taking communicative actions. The result was additional "culturally-informed" negotiation principles:

1. Allow people to maintain face and personal honor
2. Develop personal relationships (over time)
3. Generate trust in all you do – translate words into visible actions

4. Take risks to be perceived as having "good intentions" but don't "break rules"
5. Know what you must accomplish, when you should escalate and when you should simply walk away (Clark, 2005)

The above described principles and procedures are themselves declarative and procedural knowledge that has been codified into an easily expressible form. However, as with other ill-defined domains, this declarative and procedural knowledge is highly contextualized. This reflects the reality that a domain can have both well-defined and ill-defined sub-domains (Lynch et al., 2006; Schön, 1983) that are often interrelated. Within the field of artificial intelligence, the ability to elicit and represent expert knowledge has long been a challenge. Experts tend to only be able to articulate a small percentage of what they actually know. While interview techniques have been developed for eliciting knowledge from experts, there continues to be a general class of knowledge that is considered difficult to codify. The intelligent tutoring community has identified this kind of knowledge as part of an *ill-defined domain*. Outside of the AI community, Michael Polanyi (1983/1966) described it as *tacit knowledg*e and observed that "we can know more than we can tell," meaning that not only is this knowledge nearly impossible to articulate, it is very difficult to state in propositional or formal terms. It is personal in nature and includes one's own commitments and passions when making an educated guess to reach a truth.

The BiLAT system supports instruction in the well-defined and ill-defined knowledge described above in two ways. First, the well-defined and ill-defined elements of the principles and procedures are woven into the design of the scenarios and agent behaviors in the system. Students are expected to apply the principles and procedures in order to achieve success. The learning experience, however, was also designed to accommodate increasing levels of novelty and expertise through controlled levels of complexity in the scenarios and agent behaviors, such that the student experiences the outcome of the principles and procedures in a variety of contexts. For example, making judgments about the accuracy of the information received during the preparation process, analyzing the counterpart's responses to understand intent, and deciding the cost and benefits of following through on promises are examples of context-dependent lessons. The specific implementation of this technique is described further in the student experience and agent behavior sections below.

Second, the principles and procedures were distilled into a hierarchical list of "Learning Objectives" (LO) that form the basis of guidance provided by the intelligent tutoring system. These LOs are then mapped to each specific scenario in order to provide context-specific feedback that helps guide the student's learning from the experience. This technique is described in more detail later.

STUDENT EXPERIENCE AND SCENARIO MODELS

BiLAT is designed to address the learning objectives described above by allowing students to practice conducting meetings and negotiations in a cross-cultural context. BiLAT was designed and developed through a multi-disciplinary collaboration of system designers and developers, game designers, educational researchers, instructors, and subject matter experts. BiLAT incorporates domain knowledge as well as stories from real situations. Among the goals of the collaboration were instructionally sound design, real world relevance, and engagement and motivation of the student (Hill, et al., 2006). The resulting BiLAT system is a virtual game-based environment that runs on a single computer.

The student experience is structured around a scenario that progresses as the student successfully meets with characters. Success is tied to the student's ability to apply the principles and best practices of negotiation. Initially, the system was developed with ten meeting partners living in an unspecified Iraqi town. The student, a U.S. Army officer, must work with these people in the town to resolve the town's issues.

For example, there are security problems at a local Iraqi hospital. The student is told that the doctors feel unsafe, the local police are unwilling or unable to help, and patient care is threatened. The student begins by deciding whether to meet with Aziz, the head of hospital administration, or Na'eema, a doctor who works for Aziz and, as it happens, is a woman. In a successful meeting with Aziz, he reveals that part of the problem lies with the local police chief, Saleem. If the student expresses a willingness to influence the police chief to provide better protection for the hospital, Aziz agrees to keep the hospital open. Aziz also reveals that supplies seem to go missing before they reach the hospital. The student must then meet with Saleem, and the scenario progresses from there.

Functionally, the BiLAT experience, as shown in Figure 1, mirrors the twelve-step negotiation process described earlier. The student is expected to prepare prior to the meeting, conduct a meeting, and go through an after-action review (AAR). The design reinforces a key learning objective – the importance of preparation prior to entering into any meeting or negotiation. In fact, BiLAT ties a student's ability to successfully negotiate with the quality of the student's preparation prior to the meeting.

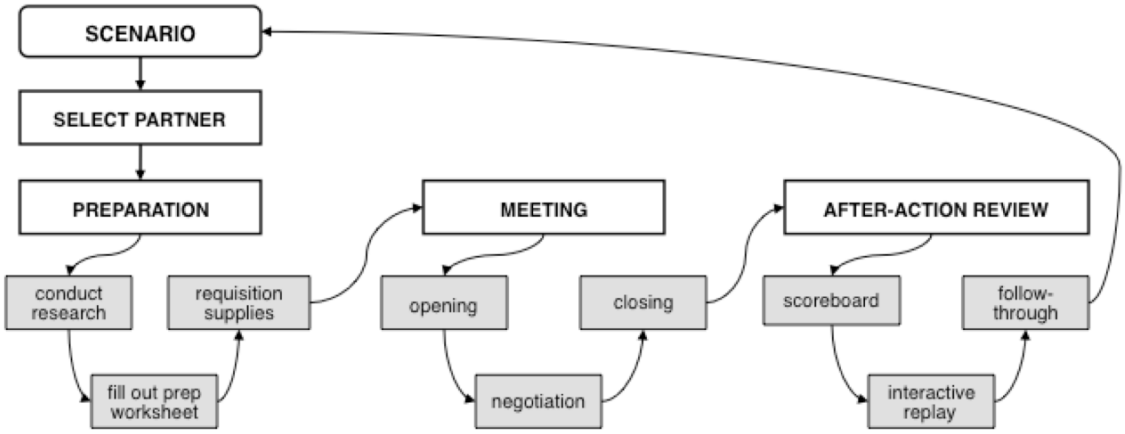

Fig. 1. Organization of BiLAT from the user's point of view.

During preparation, the student can obtain background information about the meeting partner and organize his or her thoughts using a leader preparation worksheet. The worksheet was originally developed as a job aid for soldiers preparing to conduct a meeting engagement (Wunderle, 2007). It is organized into macro issues, such as themes and messages to convey throughout the meeting, and micro issues, such as possible impasses and strategies for overcoming impasses. The worksheet covers the student's issues as well as the meeting partner's issues. The sheet is populated using research resources similar to those available in the real world such as computer databases, local newspapers, and intelligence assets, though the resources and processes are abstracted to reduce cognitive load. As items are placed onto the sheet, the student gains actions for use during the meeting, reinforcing the

notion that preparation increases the likelihood of success. For example, if the student learns that a meeting partner enjoys fine tea, the student gains the ability to requisition and offer tea as a gift.

Fig. 2. Screenshot from BiLAT – meeting with Aziz.

After preparation, the student meets with the virtual character. A screenshot of the meeting interface is shown in Figure 2. The student interacts using actions chosen from a menu. The menu system provides scaffolding for novice users who may not be able to generate actions, good or bad, on their own. The actions are organized into four categories: say, ask, give, and do. A basic set of actions is provided for each meeting and is augmented by the actions gained during preparation. Per meeting, there are on average thirty good and bad basic actions and an additional thirty good and bad actions after thorough preparation. Depending on the complexity of the character, anywhere from ten to twenty-five actions and even repeat meetings are required to successfully negotiate with a character. This includes the formal negotiation handled through a separate interface that manages offers, requests and counteroffers that are in play.

The action list was authored to allow students to pursue a variety of meeting and negotiation strategies good and bad. For example, the student can socialize by choosing the action “Flatter host,” or “Compliment locale,” or “Talk about religion.” If the student chooses “Talk about religion,” it will

garner a different response depending on the character. One of the secular characters will offer his opinion freely while one of the devout characters provides a troubled response. Among the ill-defined skills tested by the system is the student's ability to predict the effect of the action prior to taking the action, and the student's ability to interpret the meeting partner's responses and improve future predictions.

Table 1 is a partial transcript from a meeting with the hospital administrator, Aziz, after having first met with one of the female doctors, Na'eema. The student made a cultural error by talking to a subordinate of Aziz first. The exchange shows the second-order effect of the action (meeting with someone else first). Ideally, the student considered this potential repercussion prior to choosing to meet with Na'eema and re-evaluated the decision when seeing Aziz's criticism of the action.

Table 1
Excerpt from meeting with hospital administrator, Aziz, after first meeting with a female subordinate doctor named Na'eema.

Turn	**Actor**	**Action / Utterance**
9	Student	[Ask Aziz how he became a doctor]
10	Aziz	
11	Student	
	Aziz	"Good, it is better for us to talk in private."
		"What is a good way to address the problem?"
		"I would ask what you were here to see me about, but I am sure you are responding to the hysterics of one of my doctors. It is strange. Do you often go rushing to the screams of a woman before coming to see the man whose responsibility it is to oversee such things?"

After the meeting, the student is provided the option to follow-through on the promises made to the meeting counterpart. The student also finds out if the counterpart followed-through. The student then returns to the beginning of the next meeting cycle or, if they have completed their mission objectives for the entire scenario, they are introduced to the next scenario.

The BiLAT user experience and scenarios include both the well-defined and ill-defined elements of the negotiation domain. The procedural steps of the negotiation domain – for example, the twelve steps of negotiation and the decision-tree for conducting a meeting – are primarily implemented in the system as an opportunity for part- and whole-task practice. The students are expected to apply the procedures and are penalized by the system if they do not do so. The affordances of the game medium are leveraged to provide this feedback (e.g., penalties came in the form not having "won" available actions, feedback from the character). The "Across the Divide" scenarios are also deliberately written

such that students would be successful in the early meetings simply by applying the procedures identified (e.g., conducting a thorough preparation, engaging in socializing at the beginning of the meeting, waiting for the counterpart's lead to discuss business, and engaging in social close). As the student progresses through the scenarios, the novelty is increased such that a student must apply more of the ill-defined knowledge and skills (e.g., greater judgment –to conduct a successful meeting). For example, in one case, the counterpart will never agree to anything, even if Fischer & Ury's (1991) win/win (integrative) strategy is effectively applied, until the student returns for a second meeting. In fact, a significant amount of ill-defined knowledge is required during the meeting when the student must utilize interpersonal skills. The next section will describe the non-player character agents, including the integration of ill-defined elements to inform their behaviors.

BEHAVIOR MODELS

BiLAT character behaviors are described by two separate models, one governing turn-by-turn dialogue behavior and another responsible for driving the formal negotiation of offers and counter-offers. The dialogue model is rule-based with preconditions closely tied to intrinsic character state (e.g., trust towards counterpart) and the state of the dialogue, such as whether it is the greeting, social, and post-business closing interactions. The negotiation model is largely based on a simulation tool called PsychSim, to be discussed later (Marsella, Pynadath & Read, 2004; Pynadath & Marsella, 2005). Unifying these models are shared character state and access to the scenario history (e.g., who has been met with, what objectives have been completed, etc.).

Dialogue model

The dialogue models define actions, such as complimenting the host's culture, gift-giving, and making basic inquiries, that affect dialogue state. When the student takes an action, the dialogue state is updated and the character's immediate response is determined from a rule-based matching procedure. The character state variables include trust (towards the student), power (influence), wealth, religious belief, and security. Each action is also tied to one or more learning objectives through associative links that are used by the intelligent tutor (discussed later). For the Middle Eastern culture model, trust is the primary variable being managed – actions taken in accordance with cultural expectations at appropriate phases of the meeting tend to increase trust while cultural missteps decrease it. During meetings, a character's degree of trust of the student is displayed in a "trust meter," providing some explicit feedback as to the state of the relationship. Trust and other dialogue mechanics are simple but effective game abstractions that facilitate the rapid development of a rich set of scenarios by savvy, yet non-technical authors.

The meeting phase itself is governed by a state machine largely influenced by trust and time in accordance with typical Middle Eastern business meeting expectations: begin with proper greetings; follow with a social and relationship-building period; conduct actual business when a level of trust has been reached; conclude with more socializing. The trust "payoff" schedule for a given action is rooted in cultural and meeting topic appropriateness modified by meeting readiness, character state (power, wealth, belief, security), prior trust, meeting phase appropriateness (greeting, social, business), and a small random element. Meeting readiness is governed by the student's action in the preparation phase that precedes the meeting engagement. During this phase they have the opportunity to learn about the

individual's personality, the locale and the culture – variables that are manipulated in the preparation phase and referenced in the action specifications for the meeting.

In addition to reacting to the student's actions, a BiLAT character can act on its own initiative. Every dialogue action contributes to the meeting's history and can pertain to a meeting topic being discussed (e.g., hospital theft problem), represent a cultural success or misstep, build or harm the relationship, or in some other way be tracked for later consideration by the agent. The character has access to past meeting histories and could be helpful or uncooperative dependent upon past successes or failures. When specified conditions are met, the character can, for example, take a more active voice and comment about the meeting's progress, or start a conversation thread that begins with an inquiry to the student. With these devices, scenario developers have access to a simple, yet powerful set of operators and states to provide characters with meeting- and scenario-wide "memory."

For the Middle Eastern characters, this context-tracking mechanism is also used to signal expectations to the student as to what is appropriate during transitions in the meeting by offering tea, or broaching a business topic. Table 1, Line 14 shows an example where the character, after a satisfactory social phase, notifies the student that he is ready to discuss business. The character remarks on the student meeting first with a subordinate doctor, an action perceived by the character as an error. This initially places the student in a disadvantaged position, but, thanks to good interpersonal and cultural skills by the student during the meeting, it is a recoverable bargaining position.

Negotiation Model

To model the negotiation behavior, we used the PsychSim social simulation tool. PsychSim can model an entire social scenario, where a diverse set of entities, either groups or individuals, interact and communicate among themselves. Each entity has its own goals, relationships with other entities (e.g., friendship, hostility, authority), private beliefs, and mental models about other entities. PsychSim generates the behavior for these entities and provides explanations of the result in terms of each entity's goals and beliefs. A critical aspect of the PsychSim design is that the entities are autonomous agents that have fully specified models of others. This gives PsychSim a powerful mechanism to model a range of factors in a principled way. For instance, it exploits this recursive modeling to allow agents to form complex attributions about others, including others' beliefs and goals, and to take these into account when determining their own behavior.

During the negotiation phase of a meeting in BiLAT, the student negotiates with a PsychSim agent that has its own goals and various negotiation moves such as making, agreeing to or rejecting various offers, accepting or rejecting the overall agreement and terminating the negotiation. The agent also has a model of the student that it uses to reason about what negotiation moves to make. Specifically the agent can reason about what countermoves the other negotiator (the student) can make, whether the other negotiator will likely deliver on a promised offer, and so on. It also can have a model of third parties and how the negotiation outcomes will impact those third parties. To meet the development and system requirements for BiLAT, all third parties to a negotiation were disabled from acting and served only as a model reference to the negotiating agent. The main goal of the acting agent is to increase its own negotiation satisfaction, which is based on the perceived payoff (including costs) to the agent of the various deals. This is modeled with a nonlinearity - as the perceived payoff of a negotiation continues, new, more favorable offers will impact negotiation satisfaction less. The agent may also be concerned about the negotiation satisfaction of the other agent. These two goals, his own negotiation satisfaction and the others negotiation satisfaction, can be weighted differently. It is

possible to create a range of negotiation strategies by giving an agent different weights on these goals and different kinds of negotiation moves. For example, a hardball negotiator can be modeled by giving a greater weight to the agent's own negotiation strategy. The agent will make and accept offers that benefit him at first but the nonlinearity will begin to give diminishing returns to making more demands and the agent will begin to make/accept offers that benefit the student, unless the student caves in first.

When negotiating with a human, there are many places where ill-defined knowledge and skills may be required. The challenge in designing the BiLAT agent behaviors was to be able to tune the predictability of the characters without sacrificing realism. As previously described, some of the characters were crafted such that students could use well-defined procedures to be successful while some were crafted to be more difficult and less amenable to well-defined procedures. The next section will describe how the information from the user experience and agent behaviors are used to provide feedback to the student in order to make the learning more effective.

INTELLIGENT TUTORING IN BILAT

As the learner selects conversational actions during the meeting and negotiation phases of BiLAT, implicit feedback is given through the characters' responses. An inappropriate action typically results in a negative response and loss of trust, while an appropriate response usually garners a neutral to positive response and added trust. To support deeper learning, an intelligent tutoring system (ITS) was developed for BiLAT to provide explicit feedback and guidance. In this section, we describe how the ITS interacts with learners and seeks to promote learning.

Table 2
Example of Learning Objective

TRAINING OBJECTIVE	Start a meeting with a social period
SHORT NAME	Small talk
CONDITION	While meeting with a partner with whom you are trying to build trust
STANDARDS	(REQ) **expose yourself** (do not overly protect) (USL) remove **protection** (e.g., weapons) (REQ) remove **coverings** such as sunglasses and helmet (REQ) begin with **greetings** and introductions (REQ) greet in **Arabic** (REQ) conduct **formal introductions** **…**

Learning Objectives

Learning objectives (LO) in BiLAT identify the knowledge components being taught and are a key resource for the ITS. The LO structure is similar to their traditional use in the Army, but with a greater

degree of specificity, which is needed by the ITS. A simple example is shown in Table 2. The substantive part of the LOs is described in the "standards" section and resembles a recipe that, if followed, will generally lead to success in BiLAT. Some elements of the recipe are required (labelled REQ) while others may not always be needed or appropriate (labelled USL for usually necessary). Some of the LOs, based on the cognitive task analysis and literature review, however, do not translate cleanly into a procedural representation. For this reason, LOs can be tagged with ROT (rule of thumb) and AVD (avoid) to capture some of these non-procedural aspects of the domain. For example, two LOs for interpersonal skills are:

- (AVD) **save face:** avoid offending partner and help him/her save and maintain face
- (ROT) **patient:** be patient

It is difficult to identify a specific action at a specific time that the learner should take to accomplish being patient or avoiding offence, so LOs of this type act more like constraints on all learner actions that should continuously be taken into consideration.

Coaching

The coach interacts with the learner during meetings. One example appears in Table 1. All actions are assessed as *correct*, *incorrect* or *mixed*. This is accomplished by determining whether the action is phase appropriate (e.g., small talk, business, etc.) and by identifying relevant LOs. Out-of-phase actions are labeled as incorrect by the coach. For example, an LO that one should not to rush into business (Nydall, 2006) implies small talk actions are needed early in meetings. When a business-phase action is taken in an opening phase, then it is recorded as incorrect. Actions are also assessed based on authored links between LOs and meeting actions. These include a *polarity* indicating if an action represents positive or negative evidence of a learner understanding an LO. If all links are of positive polarity, then the action is classified as correct. Similarly, if all are negative, the action is considered incorrect. Any combination that includes links of both polarities means the action is classified as mixed. A rudimentary student model is maintained that tracks the student's history of positive and negative applications of LOs.

Feedback from the coach is unsolicited and is at one of two levels: a more conceptual level (e.g., "it would be good to begin with a sign of respect") and a more concrete level (e.g., "you should take off your sunglasses"). There are four categories of coach messages. *Hints* are forward-looking suggestions on what action might be appropriate next. *Warnings* are forward-looking suggestions to avoid a certain action or class of actions. *Negative feedback* is backward referring, and describes a problem with the student's last action. Finally, *positive feedback* is backward referring and praises the student's last action. Thus, there are up to four messages per game element (i.e., LO or player action) that are written as part of the authoring process. However, some combinations turn out to be incoherent – if an action is almost always a good idea, negative feedback may not make sense.

Hints are selected by consulting an expert model that inspects the current game state and runs a search algorithm that (1) identifies all available actions (some may be unavailable to the student if preparation was inadequate), (2) filters out actions that are not phase-appropriate, (3) filters out previously performed actions, and (4) identifies those actions that are positively linked to learning objectives. The default hinting strategy is to first give a hint at the abstract level on one of the LOs

associated with the action list that has not been hinted on previously. If all relevant conceptual hints have been used, the coach will hint at a more concrete level.

The decision to deliver a coaching message is based on time and the number of errors the student has made in the meeting. The current implementation has the potential to support a variety of options such as giving feedback on a fixed schedule, in a probabilistic way, or by topic. The default strategy follows a model-scaffold-fade algorithm, inspired by cognitive apprenticeship (Collins, Brown & Newman, 1989). Here, the coach provides forward looking guidance and feedback very frequently at first (the modeling and scaffolding) then pulls the support away gradually over time. This algorithm is configurable in terms of how quickly fading occurs and the level of abstraction in the content.

Reflective Tutoring

The role of the reflective tutor is to engage learners in an interactive review of their meeting that includes reviewing specific actions, reasons for the character responses and meeting outcome, and ways the learner can sustain or improve performance in the future. These techniques were adopted from instructors in live exercise after action reviews or AARs (Morrison & Meliza, 2000), and share similarities to expert human tutors that leverage reflective tactics (Katz, Albritton & Connelly, 2003). The first element of the review is the student's scoreboard, a single-page overview of his or her performance. The scoreboard consists of three components: a textual summary, scores by LO category, and a list of all actions taken during the meeting. After the learner finishes inspecting the scoreboard, the reflective tutor conducts an interactive review of the meeting (see Figure 3). The default configuration of the reflective tutor reviews the meeting in chronological order and replays parts of the meeting in the playback window (on the right). The video replay window has a progress bar across the top with hash marks corresponding to student actions. The hash marks are color coded (green=correct, yellow=mixed, and red=incorrect) and light up when the tutor is discussing them.

The review is planned in advance based on the action history, relevant LOs, and action assessments. This is primarily a grouping task, which produces an agenda for the AAR. The game action that had the most negative impact is selected and forms the "seed" of the first agenda item. The "nucleus" of the agenda item is formed from the seed and actions adjacent in time to the seed that share links to the same LO. The "satellite" of the agenda item is formed from actions sharing links to the same LO as the seed but that are not temporally adjacent. Note, the terms nucleus and satellite are taken from Rhetorical Structure Theory (http://www.sfu.ca/rst). This process continues until all actions are grouped. In the current system the agenda is then sorted chronologically by the seeds although other orders are possible (e.g., worst or best actions first). The resulting agenda emphasizes the worst mistakes, organizes the AAR around LOs and proceeds in chronological order.

The reflective tutor conducts the AAR using tutoring tactics and natural language generation knowledge encoded in a rule-base (a derivative of our earlier work, Core et al., 2006). For each agenda item, the reflective tutor selects a tutoring tactic; for example, one tactic reiterates feedback given by the coach, such as why certain actions were good or bad, but with lengthier explanations. Other tactics present the learner with multiple choice questions, such as selecting a reason why an action may have been offensive to the virtual character, or choosing an alternate action that would have been better. These questions are contextualized by the video playback and timeline, which are intended to situate the learner back in the moment from their practice session (Peters, Bratt, Clark, Bon-Perry & Schultz, 2004).

Each tutoring tactic is implemented as a rule whose right-hand side contains templates that the system's natural language generator translates into English text. There are two major types of templates. "Say templates" produce utterances that are simply read by students. The natural language generator has some flexibility in how it describes people and actions (e.g., good job apologizing about the inconvenience, you calmed Farid's concerns about the inconvenience, and Farid appreciated the apology). The templates are generic so the name of the counterpart (e.g., Farid) is inserted automatically and not hard coded. "Question templates" require the student to answer a deep reasoning multiple-choice question. If students select an incorrect answer, they get a second try. If they fail a second time, they are told the correct answer.

Figure 3. BiLAT reflective tutoring screen.

EVALUATION OF THE TRAINING EFFECTIVENESS OF BILAT

Evaluation of training effectiveness is particularly challenging for an ill-defined domain because a precise external standard of correct and incorrect does not exist. The method proposed for evaluating

the BiLAT system involved the use of a Situational Judgment Test (SJT) administered to students pre- and post-training. SJTs are often used for personnel selection and prediction of job performance (O'Connell, Hartman, McDaniel, Grubb & Lawrence, 2007). A SJT question typically begins with a scenario. The scenario is followed by several interpretations, responses, or actions. The respondent must evaluate the appropriateness of the items. Thus, items on an SJT do not necessarily have correct or incorrect answers; rather responses indicate a pattern of judgment. An individual's pattern can then be compared to normative patterns of groups with known characteristics (e.g., experts vs. novices), and a respondent can be classified by whether his or her pattern conforms to the normative pattern from the "desirable group." In the case of BiLAT, the desirable group pattern was that of subject matter experts (SMEs) on bilateral negotiation in the Middle East. The BiLAT training was decided to be effective if a student's SJT responses after training with BiLAT agreed more with the SMEs than the student's responses before training with BiLAT.

The U.S. Army Research Institute and ICT collaborated on the creation of an SJT for testing knowledge of bilateral negotiation in an Iraqi cultural context. More than twenty scenarios with associated statements to be evaluated were written. One of the scenarios (not included in the final version) is shown in Table 3. The items were pruned down to a smaller set after review by expert instructors. This set was then given to four SMEs. The items were pruned down further to maximize SME agreement while ensuring that each top-level LO had at least one item associated with it. The final SJT had nine scenarios with three or four evaluation items per scenario. Agreement among the SMEs (as measured by correlation) on the items chosen was better than .95. Participants rated each item on a scale from zero (Poor) to ten (Good). Participants' scores were standardized and correlated with the average of the standardized SME scores. This produced one number, ranging between -1 and 1, for each participant. The number represents the extent to which a participant's answers agreed with those of the SMEs.

Table 3
Example of an SJT item. This particular item was not included in the final SJT.

After about a month of relative calm in a particular town, an insurgent attack at a local market occurred where 2 Iraqi civilians died and 5 others were injured. You have been tasked to meet with town's mayor to get some information about the attackers. The mayor appears to be neutral to the U.S., hoping that peace will prevail in his town. Prior to meeting with the town leader, what type of information should you seek out?	
Determine whether he lost any family or friends in the attack.	Poor Moderate Good 0 1 2 3 4 5 6 7 8 9 10
Identify the extent to which he is in a position to deliver on his promises to you	Poor Moderate Good 0 1 2 3 4 5 6 7 8 9 10
Determine whether or not his favorite shops in the market were destroyed.	Poor Moderate Good 0 1 2 3 4 5 6 7 8 9 10
Discover his likely motives in agreeing to meet with you.	Poor Moderate Good 0 1 2 3 4 5 6 7 8 9 10

Reported here are the results of an evaluation study conducted with soldiers of the 10th Mountain Division at Ft. Drum, N.Y. Thirty-one soldiers completed the study, 16 commissioned and 15 non-commissioned officers. The study was conducted over two days, with each soldier participating on one day only. Each day was divided into three sessions, a morning session and two afternoon sessions. All soldiers for the day participated in the morning and then were assigned a time to return that afternoon (13:00 or 15:00), with half of the people assigned to each time.

During the morning session, the soldiers completed a demographics questionnaire and the SJT prior to any other events. Study personnel provided an overview briefing that described BiLAT, the BiLAT learning objectives, and the two BiLAT scenarios that would be used during the day's training (Market and Power). Study personnel then demonstrated BiLAT, introducing the user interface and the flow of a meeting. Subsequently, soldiers were paired and allowed to go through the BiLAT Market scenario for approximately 90 minutes. Study personnel circulated through the room and answered questions when asked. This practice session was intended to ensure that the students understood how to use the software and allow the soldiers to ask content relevant questions. The two afternoon sessions were identical except for the participants. Each soldier worked individually on the Power scenario.[1] In the first 60 minutes, study personnel provided help only if a soldier had software trouble (e.g., computer "freeze"). After approximately 60 minutes, the team also provided substantive assistance if asked or the soldiers appeared frustrated. After 90 minutes, soldiers were instructed to complete their current meeting if they were in the meeting or AAR phase, but not to start a new meeting. If they were in a research or preparation phase, they were asked simply to stop work. Each soldier was then asked to complete an SJT and a usability questionnaire.

We sought to detect a significant improvement in SJT scores after training, compared with before training. Although such an increase could not be attributed solely to the BiLAT system, it would suggest that the experience as a whole (briefing, game play, and study personnel assistance) contributed to an increased level of expertise in judging the appropriateness of the behaviors on the SJT. Figure 4 illustrates the pre- and post-training SJT scores according to whether participants were commissioned or had prior negotiation experience in a non-western culture. In addition to these variables, pre-training SJT responses were affected by self-rated knowledge of Middle Eastern culture (Spearman r = .49) and post-training SJT scores were affected by previous formal negotiation training (Spearman r = .39). To take these influences into account, an analysis of covariance was used to assess the change between pre- and post-training SJT scores. These scores were treated as repeated measures, and were analyzed with two two-level categorical factors (commission and previous negotiation experience), and two covariates (level of previous formal training and self-rated cultural knowledge). Neither of the covariates actually accounted for a significant amount of variance in the analysis of covariance; however, there were two significant interactions, one between time of testing (pre- vs. post) and commission, $F(1, 25) = 4.60$, $p < .05$, and one between time of testing and previous bilateral negotiation experienced, $F(1, 25) = 5.51$, $p < .05$. As can be seen in Figure 4, participants' pre-training SJT scores were higher if they were commissioned or if they had prior bilateral negotiation experience in another culture. According to a post-hoc HSD test for unequal N, the pre-training SJT scores for non-commissioned participants without prior negotiation experience were significantly lower than for all other participants; but the four subgroups formed by crossing these two factors failed to differ in their post-training SJT scores. The increase in SJT scores from pre- to post- training was significant

[1] Anecdotally, the Market scenario is believed to be easier than the Power scenario, but no formal assessment of difficulty has been conducted.

for participants without prior negotiation experience (whether commissioned or not). For those with prior negotiation experience, there failed to be a significant change.

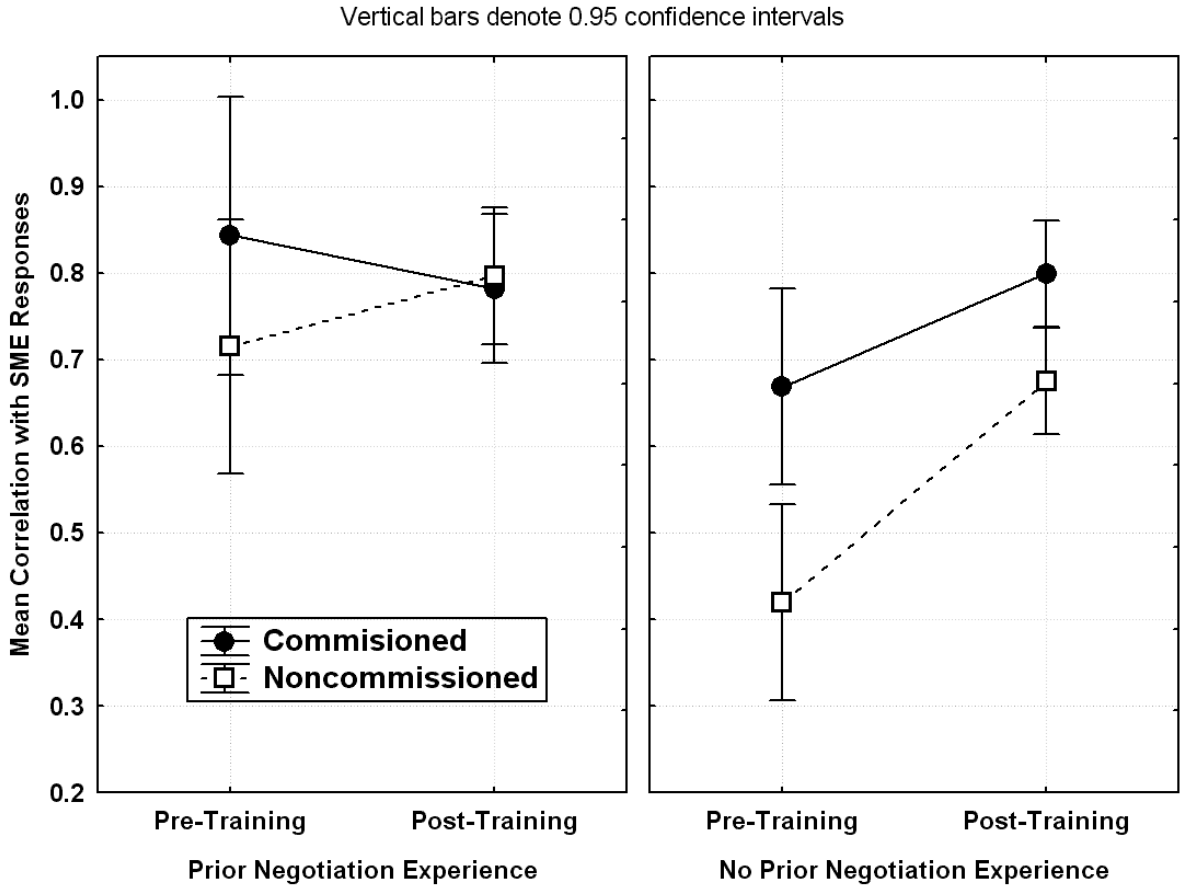

Fig. 4. Mean pre- and post-training SJT score as a function of whether participants were commissioned or noncommissioned, and whether participants had prior negotiation experience in a non-western culture (left panel) or not (right panel). SME = Subject Matter Expert.

Besides comparing pre- and post-training SJT scores, we also examined behavior during BiLAT, using data automatically logged by the application. The Power scenario tried by each soldier had three potential meeting characters and two objectives for each character (for a maximum of 6 objectives). A measure of success at the game was deemed to be the number of objectives obtained.[2] To the extent that this measure captures the same information about negotiating skill as the SJT, there should be a strong relation between this measure and SJT scores. Pre-training SJT scores correlated significantly

[2] After the study, we discovered through inspection of data logs that one particular interaction path which should have produced an objective failed to do so. There were alternative ways to obtain the same objective, so we did not attempt to adjust the data. The error highlights the difficulty of checking for logical errors when simulating the "spontaneity" of conversation and ill-defined behaviors and situations.

with the number of objectives obtained (Spearman r = .44). In turn, number of objectives obtained correlated significantly with post-Training SJT scores (Spearman r = .56). These results support the claim that the SJT scores and the number of objectives obtained in the game do overlap in the underlying capacity they measure; however, interpretation is complicated by a strong correlation between pre- and post-training scores (Spearman r = .52). A partial correlation analysis, controlling for number of objectives obtained weakened the pre- and post-training correlation but did not eliminate it. This pattern indicates that the number of objectives obtained partly accounts for variance in post-training SJT scores, but other variables were also in play. It is likely that these are variables that also contributed to the pre-training SJT scores (such as prior negotiation experience, familiarity with Iraqi culture, level of education, and IQ). According to O'Connell et al. (2007), SJT test results typically are affected by cognitive ability, conscientiousness, emotional stability, and agreeableness.

Overall, the pattern of results suggests that the SJT and number of objectives obtained during a BiLAT scenario are both valid but imperfect measures of negotiation skill in an Iraqi cultural context. The two measures likely differ in their sensitivity to discriminate differences in the underlying skill and also may be differentially affected by other influences (such as cognitive ability). At this point, there is not enough data to test a model of latent contributors to negotiation skills. As measured by increases in SJT scores, the BiLAT experience provided effective training for participants who had no previous experience negotiating in a non-western culture; however, those with prior experience failed to manifest a benefit. This could be because BiLAT as applied here was effective at training only the basic level skills these soldiers already possessed. One reason this may have been the case was that soldiers were given a relatively short opportunity to play the game individually (only about 90 min). Termination was based on time, not standard (e.g., such as number of objectives obtained). Thus, it is possible that more opportunity to interact with BiLAT, especially training to a standard (such as obtaining all six objectives) would have benefited even those players with a higher initial skill level.

Alternatively, the lack of a detectable training benefit for the more experienced participants may have been due to a lack of sensitivity of the SJT itself. The test was deliberately designed to be as short as possible, in order to fit into the time constraints that existed with the soldier participants. Consequently, the test may be unable to discriminate changes at the higher end of the scale. A more elaborate SJT might have been able to detect such changes. Only additional research will disambiguate these issues.

CONCLUSIONS AND FUTURE DEVELOPMENT

Good negotiation skills are essential to everyday life, whether in a professional or personal context. Any social situation where the parties have conflicting goals or understanding will lead to a need for negotiation. BiLAT provides a holistic practice environment for learning these skills: it blends a compelling story and characters, an interactive experience and automated tutoring to guide the student toward higher levels of expertise without the risk of costly failure. While future versions of BiLAT will expand the technical capabilities of the system, our ultimate goal is to propagate the ability to negotiate in everyday life by making it easier to create new characters and scenarios in any domain.

The initial assessment of BiLAT indicates that a game-based practice environment is useful for providing part-task and whole-task practice as a part of negotiation training. Though the scenarios were initially designed for the U.S. Army, the basic negotiation skills recommended for soldiers are similar to the skills required by anyone who negotiates (Clark, 2005; Fisher & Ury, 1991; Nobel,

Wortinger & Hannah, 2007; Tressler, 2007; Wunderle, 2007), indicating that the BiLAT design has applicability to other negotiation contexts albeit with different scenario content.

The design of the BiLAT system assumed that instruction would occur prior to the use of the BiLAT system. In reality, little more than rudimentary instruction was provided for the evaluation. Perhaps novices could have experienced more improvement if greater instruction had been provided beforehand. Early evaluations of the ITS's effectiveness suggest that it does support more efficient learning of certain cultural elements (Lane et al., 2008), but further research is needed to determine optimal patterns of intervention. Furthermore, since a longitudinal study has not yet been conducted, BiLAT's mid- and far-transfer is still to be determined.

The use of spoken natural language understanding for inputting actions was considered, but the click-button approach was chosen due to the time constraints under which the system was initially built. A menu of actions is useful for providing scaffolding to novices who may not be able to generate actions on their own. The novice's transition to an expert's ability to generate actions on one's own should be further investigated. This could include a hybrid model wherein the student tries to generate an action and is then presented with similar choices from the system. A student model, including biographical information such as prior experience, could also inform how much help the student is given with generating actions. It would also be possible to implement scenarios where other negotiation strategies, such as a distributive approach, could be taken, with the scenarios as well as the tutoring helping students to see the effects of such approaches.

The use of scenarios in ill- and well-defined domains should continue to be investigated. There remains an on-going tension between creating scenarios that appropriately reflect both the well-defined procedural knowledge and ill-defined knowledge, as well as creating a sense of realism and engagement. Realism will support the need for real world problems in guided instruction, and engagement, while on its own not correlated to learning, could motivate students to engage in greater practice that would lead to greater learning.

ACKNOWLEDGMENTS

The project described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred. We are greatly indebted to the awesome ELECT BiLAT team past and present without whom this would not have been possible. We also gratefully acknowledge the support of the soldiers who provided input and feedback, with special thanks to LTC Bill Wunderle and LTC Ken Smith.

REFERENCES

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

Bloom, B. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16.

Clark, R. (2005). Summary of CTA interviews. Unpublished report. October 16, 2005.

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. 453-493). Hillsdale, NJ: Lawrence Erlbaum Associates.

Core, M.G., Traum, D., Lane, H.C., Swartout, W., Gratch, J., van Lent, M., & Marsella, S. (2006). Teaching negotiation skills through practice and reflection with virtual humans. *SIMULATION*, 82, 685-701.

Department of Defense. (November 2005). Directive 3000.05: Military Support for Stability, Security, Transition, and Reconstruction Operations.

Fisher, R. & Ury, W. (1991). *Getting to Yes: Negotiating Agreement Without Giving In* (2nd Ed.). New York: Penguin Books.

Fullerton, T., Swain, C., & Hoffman, S. (2004). *Game Design Workshop: Designing, Prototyping, and Playtesting Games.* Gilroy, CA: CMP Books.

Hill, R., Belanich, J., Core, M., Lane, H., Dixon, M., Forbell, E., Kim, J., & Hart, J. (2006). Pedagogically structured game-based training: Development of the ELECT BiLAT simulation. Poster presentation at the 25th Army Science Conference, November 2006, Orlando, FL.

Katz, S., Albritton, D., & Connelly, J. (2003). Going beyond the problem given: How human tutors use post-solution discussions to support transfer. *International Journal of Artificial Intelligence in Education*, 13(1), 79-116.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.

Lane, H.C., Hays, M., Core, M., Gomboc, D., Forbell, E., Auerbach, D., & Rosenberg, M. (2008). Coaching intercultural communication in a serious game. In T.W. Chan (Ed.) *Proceedings of the 16th International Conference on Computers in Education* (pp. 35-42). Jhongli City, Taiwan: APSCE.

Lynch, C.F., Ashley, K., Aleven, V., & Pinkwart, N. (2006). Defining "ill-defined" domains: A literature survey. In V. Aleven, K. Ashley, C. Lynch, & N. Pinkwart (Eds.) *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the 8th International Conference on Intelligent Tutoring Systems* (pp. 1-10). Jhongli, Taiwan: National Central University.

Marsella, S.C., Pynadath, D.V., & Read, S.J. (2004). PsychSim: Agent-based modeling of social interactions and influence. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.) *Proceedings of the Sixth International Conference on Cognitive Modeling (*pp. 243-248). Mahwah, NJ: Lawrence Erlbaum Associates.

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59, 14–19.

Morrison, J.E., & Meliza, L.L. (1999). *Foundations of the After Action Review Process*. (ARI Special Report 42*)*. Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Nobel, O., Wortinger, B., & Hannah, S. (2007). *Winning the War and the Relationships: Preparing Military Officers for Negotiations with Non-Combatants* (ARI Research Report 1877), Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L., III, & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment*, 15, 19 – 29.

Nydall, M.K. (2006). *Understanding Arabs: A guide for modern times* (4th ed.). Boston: Intercultural Press.

Peters, S., Bratt, E.O., Clark, B., Bon-Parry, H., & Schultz, K. (2004). Intelligent systems for training damage control assistants. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*.

Polanyi, M. (1983). *The Tacit Dimension*. Gloucester, MA: Peter Smith. First published by Doubleday & Company, Inc., 1966.

Pynadath, D.V. and Marsella, S.C. (2005). PsychSim: Modeling theory of mind with decision-theoretic agents. In L. Kaelbling & A. Saffiotti (Eds.) *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence* (pp. 1181-1186). Edinburgh: Professional Book Center.

Raiffa, H. (1982). *The Art and Science of Negotiation*. Cambridge, MA: Harvard University Press.

Schön, D. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
Sternberg, R., Forsythe, G., Hedlund, J., Horvath, J., Wagner, R., Williams, W., Snook, S., & Grigorenko, E. (2000). *Practical Intelligence in Everyday Life*. New York: Cambridge University Press.
Tressler, D. (2007). *Negotiation in the New Strategic Environment: Lessons from Iraq*. Carlisle, PA: Strategic Studies Institute, U.S. Army War College.
Wunderle, William D. (2007). *Through the Lens of Cultural Awareness: A Primer for US Armed Forces Deploying to Arab and Middle Eastern Countries*. Fort Leavenworth, KS: Combat Studies Institute Press.