# Getting down to business: Teaching cross-cultural social interaction skills in a serious game

H. Chad Lane[1] and Matthew J. Hays[2]

[1]University of Southern California
Institute for Creative Technologies
13274 Fiji Way, Marina del Rey, CA USA
lane@ict.usc.edu

[2]University of California – Los Angeles
Department of Psychology
1285 Franz Hall, Los Angeles, CA USA
matt@hayslab.com

**Abstract.** We consider the use of virtual humans and an intelligent tutoring system (ITS) for the teaching of cultural social conventions. Our learning environment is a serious game that requires the learner to establish trust and reach agreements with virtual characters of a different culture. An intelligent tutoring system provides culturally focused learning support during and after the meetings with these virtual characters. In a study intended to determine the effectiveness of the ITS, we found that guidance provided during meetings led to an improved understanding of culturally-related "phases" in meetings (e.g., when to talk about business) as well as greater success in an unsupported posttest meeting, but with no overall increase in cultural understanding when compared with learning in passive and unguided conditions.

**Keywords:** serious games, intelligent tutoring systems, cultural awareness, intercultural competence, intercultural communication, interpersonal skills

## 1 Introduction

Culture can play a significant role in the success or failure of face-to-face encounters. Many of the expectations we hold going into a conversation or meeting have cultural explanations, or at least are influenced by our cultural values and backgrounds. Similarly, meanings we infer from the communicative acts of others are influenced heavily by our own cultural "lenses." So, when we enter into conversations with people from cultures other than our own, the differences can quickly become a source of confusion, misunderstanding, and at worst, conflict. Awareness of such differences – and a willingness to assume different cultural perspectives – is key for growth towards intercultural competence [2]. In this paper, we consider the use of virtual humans, serious games, and an intelligent tutoring system (ITS) for the learning of cultural social conventions. The aim is to teach trust-building strategies and how to reach agreements with virtual characters from another culture.

## 2 Immersive Environments for Cultural Learning

Immersive learning environments provide new and unique ways in which to learn about a new culture [7]. High-fidelity graphics, sound, and first-person perspectives make it possible to simulate the tangible aspects of a specific culture (e.g., dress, gestures) and provide more authentic practice environments than may otherwise be feasible with more traditional approaches like live role-play or media-based approaches. Further, recent advances in artificial intelligence (AI) and cognitive modeling now permit rich modeling of emotions, language, tasks, and more [13]. When built with cultural accuracy, these models (and the resulting virtual humans) may open new avenues for teaching the cognitive and interpersonal aspects of learning about different cultures.

A number of cultural learning systems exist that take advantage of these capabilities. The Tactical Language and Culture Training System provides a mission practice environment that allows learners to explore a virtual town while speaking to locals in Arabic, make culturally influenced gestures, and accomplish goals such as learning the names of contacts and getting directions [5]. VECTOR, another cultural learning system, also situates learners in a virtual foreign town to explore, but uses English utterances via menu selections for interaction with locals. An example of a goal in VECTOR is to find a bomber and stop him from attacking his next target [9]. Yet another immersive training environment, the Adaptive Thinking and Leadership system, is a team-training system that uses human role players in intercultural scenarios. Learners assume roles as people from different cultures and are given believable back-stories and goals [11].

We have conducted our research in another cultural learning system: ELECT BiLAT (Enhanced Learning Environments with Creative Technologies for Bi-lateral negotiations). BiLAT, a serious-game-based immersive learning environment, teaches the preparation, execution, and understanding of bi-lateral meetings in a cultural context [4]. The primary users of BiLAT have thus far been U.S. Soldiers who are learning about Middle Eastern culture. The full version of BiLAT requires the learner to work in the context of a complex narrative, select virtual characters with whom to meet, conduct background research on those characters and the situation facing the community, choose interpreters, pre-plan meetings, and more. Here, we focus only on the face-to-face aspect of BiLAT because (1) it represents the *practice* phase of learning (when feedback is most essential), and (2) it is most similar to live role-play-based training commonly used in classrooms and intercultural training.

A BiLAT meeting consists of two modes: dialogue and negotiation. A screenshot of the dialogue screen is shown in Figure 1. A menu of conversational actions is used to allow the learner to take communicative actions (e.g., ask questions and make statements – lower left in the screenshot) as well as physical actions (e.g., removing sunglasses, giving a gift). The character responds using a synthesized voice and physical gestures. The spoken utterances also appear as text in the dialogue window in the lower right corner. Although there are dozens of variables governing the actions of the character, the variable of primary importance is trust (it normally appears in upper left corner of the screen, but is blank for reasons of experimental control – this is described later). Although characters may say nice things or display anger in their responses, trust is the persistent record of how well players have used their


**Figure 1.** Screenshot of ELECT BiLAT

interpersonal and intercultural skills. We do not discuss the negotiation aspects here as our focus currently is interpersonal and intercultural skills, except to note that trust is a major factor in whether BiLAT characters will agree to negotiate and what deals they will accept. A mistrusting character will demand more, decline to negotiate, or at worst, kick the player out of a meeting without warning.

Characters each possess a model of how they expect a meetings to progress and the effects of actions are defined in ways that reinforce the cultural expectations. For example, attempting to give alcohol as a gift will have a negative impact on trust for Arabic characters. 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 response is determined. For the Middle Eastern culture model, trust is the primary variable being managed – actions taken in accordance with cultural expectations at appropriate phases tend to increase trust while cultural missteps decrease it.

The dialogue 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), existing trust, meeting phase appropriateness (greeting, social, business), and a small random element.

BiLAT characters respond by selecting from a large collection of hand-authored utterances. These utterances were a result of a long process of analyzing transcripts from Iraqi-American role-playing scenarios involving negotiation and consultation of sources on Middle Eastern culture (e.g., [10]). Twelve possible responses exist for each available user action, along with a set of universal standard responses (e.g., to display confusion). The presence of randomness is intended to simulate the uncertainty of human behavior – more advanced AI, cognitive and emotional modeling techniques are able to simulate this uncertainty in a more principled way [13]. Each action entails a possible change to the trust variable of the character. All content – available actions, utterances, trust changes, etc. – are authored; since the characters behaviors and utterances are data, the BiLAT framework is fundamentally independent of culture. When given new behavior data and graphical assets, other cultures can be represented by the virtual characters (the problem of cross cultural knowledge representation is addressed in [12] where development of a German version of one of the BiLAT characters is discussed).

Because of variability between cultures with respect to how time is treated [10], BiLAT also represents distinct time spans corresponding to a business period and social periods before and after business. Authors are required to indicate the phases when actions are appropriate. If a learner chooses an action that is not appropriate for the current phase of a meeting, the character will respond negatively, which is revealed in the content of the response, through gestures, and with a likely decrement in trust. One of the learning objectives (LOs) underlying BiLAT for Middle Eastern culture is that the learner should follow the lead of his or her host – this is a particular focus of the of the ITS and is discussed in the next section.

## 3 Coaching for culture

In BiLAT, the learner must repeatedly select conversational actions that simultaneously achieve game objectives and respect the cultural norms observed by the virtual counterparts. This can be a significant challenge for a learner who does not understand the new culture or the differences between it and his or her own. Our ITS, an instance of the Intelligent Guided Experiential Learning (IGEL) framework [2], can provide learning support in two ways. *During* meetings, feedback and hints can be delivered by a *coach. After* meetings, an intelligent *reflective tutor* reviews the meeting with the learner, gives more detailed feedback, and asks reflection questions. In the present study, we focus on the details of the IGEL coach that deal with timing aspects of intercultural meetings. This is also the focus of the evaluation described in Section 4.

### 3.1 Feedback and hinting

When the coach decides to provide some form of guidance, a message appears in the dialogue window of BiLAT (lower right hand corner of Figure 1). These messages are intended to promote learning and reduce frustration. For example, if an inappropriate

gift is given to the virtual character, the coach might explicitly state the gift is not an acceptable one for the virtual character's culture (e.g., "alcohol is not generally acceptable in Middle Eastern cultures."; [9]). Coaching messages are generally shorter given the context of a live meeting, while reflective tutoring sessions are used to get into the underlying cultural issues through interactive questioning and explanations.

The coach is capable of two levels of feedback: abstract and more specific. Further, at each of these levels, three categories of coaching actions are possible:

- **hints**: a suggestion pointing the learner to an appropriate next action
- **negative feedback**: statement that an action just taken was problematic and (possibly) a short explanation
- **positive feedback**: praise for a good action taken by the learner and (possibly) a short explanation

Since cultural rules vary across individuals and are ill-defined to begin with, actions often resist clear assessments as "right" or "wrong." Some BiLAT actions can therefore be "mixed," meaning that both negative and positive feedback may be possible (discussed in the next section). Feedback messages in these situations can be concatenated or delivered individually based on other factors, such as if related errors (or related "good" actions) were selected earlier in the meeting. These decisions depend on the configuration of the coaching algorithm. The content of the coaching messages – as well as the settings on how often and at what detail feedback is given – must be determined by a human author ahead of time. For the experiment described in Section 4, we implemented a model/scaffold/fade algorithm to control feedback timing (details are discussed in [6]).

### 3.2 Assessing actions

Of course, it is critical for the coach to assess actions to decide whether to give feedback and if so, what to say. In BiLAT, each time an action is taken in a meeting or the learner responds to a character's question, IGEL's expert model is called to judge the action's quality. Actions are classified as *correct*, *incorrect*, or *mixed* in two stages:

1. The action is checked to see if it is phase appropriate.
2. If no phase mismatch is found, the active LOs are determined along with their positive and negative association with them.

As discussed above, meeting phases are windows of time during a meeting that define when certain categories of actions are appropriate or not. They are culture dependent and the expert module dynamically assigns negative assessments when an inappropriate action is chosen. This is implemented via a link to the appropriate LO.

As discussed in section 2, BiLAT focuses on teaching Middle Eastern business culture. These rules (many are discussed in [10]) involve discussing the right topics at the right time. In addition, they involve respecting the cultural values in ways that indicate interest, understanding, and respect. Timing of actions is an important part of this process – e.g., jumping directly to business before gaining the trust of a negotiation partner can introduce new hurdles, even perhaps insurmountable ones. To

teach these critical skills, we currently use the following set of LOs to classify phase mismatch errors:

1. Don't discuss business during social periods.
2. Don't stray too far from business in a business period.
3. Regarding the opening of meetings:
    a. Opening actions (e.g., greetings) are not appropriate later in the meeting.
    b. The opening is too early for some social actions.
4. Regarding the closing of meetings:
    a. Closing actions (e.g., leaving) are not appropriate earlier in the meeting.
    b. Some social actions are not appropriate in the closing.

The expert model considers in which phases an action is permitted and in which phase it is actually performed. For example, if an action is permitted only in the opening phase (e.g., greeting in Arabic) but is performed in the pre-business phase, the expert model's algorithm will return negative evidence of recognizing LO 3a above. It will also find positive evidence of understanding the LO to greet in the language of your counterpart. The expert model's assessment will be therefore be *mixed*. Another example of a phase-related LO states that it is not ideal to rush into business [9, p.58], which implies conversational actions are needed early in meetings. When a business-phase action is taken in an opening phase, then, the expert model records an *incorrect* for that action (i.e., that the player does not understand that LO).

## 4 Evaluation

The LOs of BiLAT deal with effective negotiation strategy, trust building, and appropriate meeting behavior (i.e., actions relevant to the meeting phase and partner's culture). Together, practice in BiLAT and extrinsic feedback from IGEL (as opposed to intrinsic feedback from the characters' actions and responses) should allow the learner to gather a practical understanding of the LOs. In this section, we report on a study intended to examine the contributions of each component: BiLAT and IGEL.

### 4.1 Research questions: Interactivity and real-time feedback

Our first question was whether actually *playing* BiLAT would be pedagogically beneficial. A common notion is that maximum gains can be produced by *errorless learning*, wherein the conditions of instruction are such that it is impossible for learners to make errors [1]. In the present study, we used a *video-only* condition, in which participants watched videos of perfect gameplay. Because no incorrect choices are made, optimal behavior is modeled for these learners.

Nevertheless, we expected that the video-only condition would suffer because it would lose the benefit of interactivity. Participants who actually play BiLAT would select actions they believed to be correct. Perhaps of even more value: participants would also select *against* actions they believed to be *incorrect*. In this way, they would be evaluating *all* of the available actions – not just those selected by the player in the video of perfect gameplay. We expected that evaluating the actions would

produce general gains when meeting with new characters. We therefore expected interactive conditions to be superior to the video-only condition.

Our second question was whether the IGEL-driven coach adds any pedagogical value. A concern that accompanies any form of extrinsic support is dependence on the help. Such reliance might lead the player to neglect feedback from the meeting partner; if that support was removed, the player might flounder. Instead, we believed that the coach would produce significant learning gains because of its ability to identify incorrect actions and the reasons those actions are incorrect. Plainly, this type of feedback could generalize to other actions – and the learner's ability to evaluate whether those actions would be advisable. To get at this question, we included two interactive conditions: one in which the coach was active (*yes-coach*) and one in which it was not (*no-coach*). In all three conditions, the reflective tutor was active to ensure that participants in all three conditions received extrinsic pedagogical support.

### 4.2 Method

**Participants.** Participants were thirty U.S. Citizens recruited by flyer from the campus of the University of Southern California. As compensation for their three hours of participation, they were paid $60.

**Procedure.** All participants received an instructions packet, watched a video on how BiLAT worked, took the pretest, conducted meetings with three different BiLAT characters (including reading background information about each and reflective tutoring sessions), conducted a fourth meeting with no coaching or reflective tutor, and finally took the posttest.

**Design.** The manipulation between the three conditions – *video-only*, *no-coach*, and -*yes-coach* – occurred in the three meetings. For video-only, participants observed expert play with coaching and reflective tutoring active. For no-coach, they held the meetings and received reflective tutoring. For yes-coach, they played with coaching and reflective tutoring. Otherwise, participants' experiences were identical.

**Measures.** We made two comparisons of the three between-subjects conditions. The first was *success*, which was measured on the uncoached fourth meeting. This measurement was binary: whether the participant was able to achieve the mission objective in 25 minutes or less (as with all of their meetings, participants could repeatedly meet if prior meetings were unsuccessful – up until time expired). A successful meeting required achieving given objectives identified in the reading materials and by choosing actions that increased the meeting partner's trust to the point where agreements were possible.

The second measurement was the pretest-posttest improvement. A *situational judgment test* (SJT) was used for the pre- and posttests, which present a series of small scenarios. For each scenario, the participant is asked to rate the quality of a small set of responses on a Likert scale [8]. Each action is rated between 0 (never take this action) up to 10 (definitely a good action). The test was given to three subject-

matter experts (SMEs) and the means of their answers were used as the gold standard. Two measurements were used to evaluate participants' mastery of the LOs. One is the *correlation* between participants' answers and SMEs'. The correlation represents the degree to which the two groups correspond in their overall assessment of the responses, but does not capture valence data. To illustrate: let us say that a SME rates an answer as 10 and that is the highest rating the SME gives. If a participant rates that answer as 4, but 4 is the highest rating that participant gives, the two may be perfectly correlated. To capture information about the valence of answers, we calculated what we have called a *ballpark* score. To illustrate: if an SME rates a response as 8, the participant would be said to be "in the ballpark" if the participant provides a rating of 7, 8, or 9. The change in each score was measured from pretest to posttest.

### 4.3 Results and Discussion

**Fourth meeting.** Three participants – one from each condition – were omitted from this analysis because of experimenter error. As a result, this analysis included a total of 27 participants (nine per condition). Figure 2 presents the success data from our experiment. As can be seen, the combination of interactive gameplay and feedback from the coach helped players to be successful in the fourth meeting (in which both the coach and reflective tutor were deactivated). 56% of the participants in the video-only condition were successful. With the addition of interactivity – and the ability to make errors – 67% of the participants in the no-coach condition were successful. Finally, those who had coaching during practice had the highest rate of success in the fourth meeting at 89%.The number of participants per cell (nine) in this analysis was too small for a reliable statistical analysis to be run. Nevertheless, we view these results as encouraging.

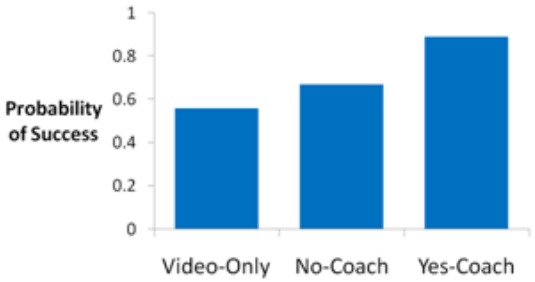


**Figure 2. Probability of success (by condition)**

**Situational Judgment Test.** The uncoached meeting suggested learners who interacted and were coached improved their ability to play BiLAT. To check for learning beyond the game, we used the SJT to determine whether participants were mastering the LOs. The SJT therefore served as a non-ELECT-BiLAT measure of learning from BiLAT (and IGEL). These data initially painted a different picture. We did not find that the conditions produced a differential increase in correlation with SMEs: $F(2, 27) = .805$, $p = .457$. Neither did we find a significant difference between the conditions on the ballpark measure: $F(2, 27) = .111$, $p = .896$.

Immediately, we wondered why there would be such a disparity between performance in the game and improvement on the SJT. One possibility is that the pretest SJT ratings affected posttest ratings (i.e., that learning occurred from the test itself). It is unlikely that participants were able to remember the exact numbers that they had provided hours before. Nevertheless, pretest ratings may have served as anchors for posttest responses. To the extent that this interference reduced variance in posttest responses, it reduced the ability for between-group differences to emerge.

Another, more likely possibility is that this particular experiment and the SJT were not perfectly aligned. The "correct" answers on the SJT were provided by subject matter experts, whose knowledge spans all of the elements of BiLAT, all of the LOs in IGEL, and much, much more. The SJT itself was not originally designed as a measure for this experiment. As a result, the SJT taps knowledge about many more issues of negotiation and culture than we could have expected to address in three hours with our participants. The SJT may be too broad in scope to examine gains in knowledge *in our participants* as a result of our experimental procedure.

We therefore ran a second analysis of learning gains in the SJT. In this secondary analysis, we culled data from questions that did not address LOs addressed by IGEL. For example, we omitted responses to prompts about the reliability of information during the meeting preparation phase (which was skipped in our experiment; participants were instead provided with a dossier). By removing these responses from our analysis, we increased our power to detect learning in two ways. First, to the extent that participants' responses to culled answers remained fixed from pre-test to post-test, estimates of change would be deflated and would therefore underestimate the amount of learning in the subset of SME knowledge tapped by BiLAT and IGEL. Second, to the extent that participants' responses to culled answers wandered randomly, pre- and posttest correlations would be attenuated and within-group variance in both of our measures would increase. This secondary analysis therefore reflected only changes in knowledge in domains addressed by BiLAT gameplay and the IGEL coach. The domain of meeting-phase-specific behavior was highlighted for this analysis. It draws on cultural understanding and appropriate use of negotiation strategies.

**Correlation data for phase-specific questions.** Figure 3 presents the correlation data for the three conditions in our experiment. As can be seen, interactive gameplay and the coach together produced significant gains in understanding of meeting-phase-specific information: $F(2, 27) = 2.062$, $p = .147$. Post-hoc tests revealed this result to be driven by the difference between the superior yes-coach and the inferior no-coach conditions: $p = .054$.

This finding is not surprising; an action's phase-appropriateness is the branching point from which the coach begins to decide to provide feedback. If the coach were to produce benefits in only one way, this would be it. Regardless of any other elements of the action chosen by the player, the phase-appropriateness is a primary determinant of coach intervention.

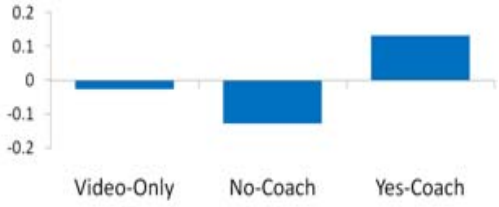


**Figure 3. Changes in correlation with experts on meeting-phase-specific prompts (by condition)**


**Figure 4. Changes in "ballpark" score on meeting-phase-specific prompts (by condition)**

It should be noted that the no-coach condition is actually *worse* than the video-only condition (as shown in Figure 3). This pattern of results may reflect the inadequacy of discovery (i.e., trial-and-error) learning. Without guidance from the coach or a model of ideal gameplay (e.g., the video in the video-only condition), players in the no-coach condition are left to wander through actions. It appears that this unguided wandering is unproductive, perhaps even harmful.

**Ballpark scores on phase-specific questions.** Figure 4 presents the ballpark data for the three conditions in our experiment. Again, playing BiLAT and receiving real-time feedback from IGEL resulted in marginally superior comprehension and retention of the LOs related to meeting phase: $F(2, 27) = 1.681$, $p = .205$. Post-hoc tests revealed a marginally significant difference between the superior yes-coach and the inferior video-only conditions: $p = .086$. This finding provides additional (albeit preliminary) support for our conclusions from the correlation data.

On the other hand, the reversal of the video-only and no-coach relationship in the correlation and ballpark measures suggests some instability in our measures. This is likely due to the small number of participants and to the relatively small number of data points per participant; there were only five phase-specific prompts on the SJT. Nevertheless, in combination with the success data, we find these results encouraging, and will conduct a more comprehensive investigation of benefits from IGEL in BiLAT in the coming months.

## 5 Conclusion

This paper described a serious-game-based approach to teaching intercultural communication skills with virtual human characters and with the support of an intelligent tutoring system. Our results suggest that learning occurs and is likely due to the interactivity of the game and from the intelligent tutoring support during face-to-face time with the virtual characters. We intend to continue experimentation to further explore these preliminary findings by considering different patterns of coach feedback, more advanced reflective tutoring tactics, and through adjusting character behaviors (this idea is developed further in [6]). We also intend to revise our pre- and posttests to better match the intended aims of the IGEL tutoring framework and its instantiation in BiLAT.

**Acknowledgements.** We thank Dave Gomboc, Daniel Auerbach, Ashish Karnavat, Milton Rosenberg, and Eric Forbell for their help in making the modifications necessary for us to complete our study. We also thank Jim Belanich and Paula Durlach from the U.S. Army Research Institute for their thoughtful feedback and suggestions on the study. 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.

## References

1. Anderson, N.D., Craik, F.I.M.: The mnemonic mechanisms of errorless learning. Neuropsychologia, 44, pp. 2806-2813 (2006)
2. Bennett, M.J.: Towards ethnorelativism: A developmental model of intercultural sensitivity. In: R.M. Paige (ed.) Education for the Intercultural Experience, pp. 27-71. Intercultural Press, Yarmouth (1993)
3. Gomboc, D., Lane, H.C., Core, M., Karnavat, A., Rosenberg, M.: Towards an ITS architecture for simulations. In: Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference (FLAIRS-21). AAAI Press (2008).
4. Hill, R.W., Belanich, J., Lane, H.C., Core, M.G., Dixon, M., Forbell, E., Kim, J., Hart, J.: Pedagogically structured game-based training: Development of the ELECT BiLat simulation. In: Proceedings of the 25th Army Science Conference. Orlando, FL (2006)
5. Johnson, W.L.: Serious use of a serious game for language learning. In: The 13th International Conference on Artificial Intelligence in Education, pp. 67-74. IOS Press (2007)
6. Lane, H.C., Core, M.G., Gomboc, D., Karnavat, A., Rosenberg, M.: Intelligent tutoring for interpersonal and intercultural skills. In: Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2007). Orlando, FL (2007)
7. Lane, H.C.: Metacognition and the development of intercultural competence. In: Proceedings of the Workshop on Metacognition and Self-regulated Learning in Intelligent Tutoring Systems at the 13th International Conference on Artificial Intelligence and Education, pp. 23-32. Marina del Rey, CA (2007)

8. Legree, P., Psotka, J. Refining situational judgment test methods. In: Proceedings of the 25th Army Science Conference (2006)
9. McCollum, C., Deaton, J., Barba, C., Santerelli, T., Singer, M.J., & Kerr, B.W.: Developing an immersive, cultural training system. In: Proceedings of I/ITSEC: Interservice/Industry Training, Simulation, and Education Conference (2004)
10. Nydall, M.K.: Understanding Arabs: A guide for modern times (4th ed.). Boston, Intercultural Press (2006)
11. Raybourn, E.M.: Applying simulation experience design methods to creating serious game-based adaptive training systems. Interacting with Computers, 19, pp. 206–214 (2007)
12. Solomon, S., van Lent, M., Core, M., Carpenter, P., & Rosenberg, M.: A language for modeling cultural norms, biases, and stereotypes. In: Proceedings of the 17th Conference on Behavior Representation in Modeling and Simulation (2008)
13. Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., Rickel, J.: Toward virtual humans. AI Magazine, 27, 2, 96–108 (2006)