A Language for Modeling Cultural Norms, Biases and Stereotypes for Human Behavior Models

Steven Solomon
Michael van Lent
Mark Core
Paul Carpenter
Milton Rosenberg

Institute for Creative Technologies
University of Southern California
13274 Fiji Way
Marina Del Rey, CA, USA 90292

Keywords: Human Behavior Model, Cultural Knowledge Representation, Cultural Norms, Biases, Stereotypes, Cognitive Schemas

ABSTRACT: *Increasingly, the military has requirements for teaching cultural awareness, which demands flexible representations of cultural knowledge. The Culturally-Affected Behavior project seeks to define a language for encoding ethnographic data in order to capture cultural knowledge and use that knowledge to affect human behavior models. Having anthropologists encode ethnographic data will validate the language and will result in a library of culture models for immersive training.*

1. Introduction

Recent years have seen much interest in immersive training environments for training users in social and cultural skills. For example, ELECT BiLAT (Hill et al, 2006) trains users how to plan for and conduct bi-lateral meetings or negotiations in a cross-cultural setting. The focus of the present research effort is to define a language for encoding cultural knowledge that can be used by an ethnographer to record a set of ethnographic data such that it can affect the behavior of a human behavior model. Our language focuses on encoding cultural norms and mental models of biases and stereotypes. This language also provides a common representation of cultural knowledge that is separate from a character's domain or task knowledge representation.

The current research effort follows in the mold of cognitive modeling. Just as cognitive modeling is considered a collaboration of artificial intelligence (AI) and cognitive psychology, we seek to form a collaboration between AI and various fields of social science such as cultural and cognitive anthropology.

The effort described here focuses on modeling the socio-cultural values and attitudes of a culture, as well as mental representations of others' beliefs, as described in Section 3. Our approach to building these models draws on social science theories and previous human behavior models, both described in Section 2. We extended one of the previous human behavior models to represent our target aspects of culture. In Section 4, we present this extension and show how our model captures concepts of the social science theories that serve as our foundation. In Section 5, we discuss how our model generates culturally-affected behavior. Finally, we discuss the Culturally-Affected Behavior (CAB) prototype Section 6, and plans for evaluating our approach in Section 7. The result of this work is the language, the extended human behavior model, and two sample socio-cultural models coded using the language.

2. Related Work

2.1 AI and Immersive Training

Virtual Humans

Virtual Humans (Traum et al, 2003; Gratch & Marsella, 2004) is a human behavior model that consists, among other major components, of a task model and a dialogue model. The task model is used to represent domain plans and the underlying world state from the perspective of the human behavior model, while the dialogue model generates speech acts using the task model and the human participant's actions. In Virtual Humans-based applications, users interact through voice, text, or menu input with a virtual character that is implemented using the human behavior model. In the Stability and Support Operations (SASO) application of the Virtual Humans, the user plays the role of a U.S. Army Captain who is tasked with persuading a doctor (the Virtual Human) from the non-governmental organization Doctors Without Borders to move his clinic to a new location. The doctor character has a number of goals including helping his patients and remaining neutral. In the negotiation, the user must build trust with the virtual doctor by building credibility, showing solidarity with the doctor's goals, and establishing familiarity. In addition, the user has a variety of available options including offering to help transport the patients and giving the doctor medial supplies.

The task model formalism, based on a decision-theoretic framework, is the core engine that determines the human behavior model's choices and behaviors. The human behavior model consists of a multiple task models, called plans: a default plan and one or more alternative plans. All plans are authored before runtime, are to be executed in the future, and achieve a goal, e.g. "patients-are-helped". Plans have utility values. The human participant's goal is to get the utility value of an alternative plan ("run-clinic-there") to be greater than the utility value of the default plan ("run-clinic-here") so that the human behavior model switches to the alternative plan the user desires.

Task models consist of tasks, states, and effects. Tasks represent simulation actions, e.g. "donate-vehicle" and "move-clinic", which can be performed by the human participant or the human behavior model. States encode states of the world—e.g. "doctor-is-neutral"—and have intrinsic utility values that represent the relative importance that the human behavior model has for the state weighed against other states. Tasks have effects on states. Effects have a sign (+/-) which represents whether a task reinforces or detracts from a state. For example, "donate-vehicle" positively affects "have-transport", while "cooperate-with-U.S." negatively affects "doctor-is-neutral". States can also be preconditions of tasks. For example, the state "have-transport" is a precondition of "move-clinic".

Tasks and states have associated probabilities that change during a simulation. The probability of a task is the likelihood that the task will be performed in the future. One factor in the calculation of a task's probability (for tasks to be performed by the human participant) is the confidence level that the human behavior model has in the participant's commitment to perform agreed upon tasks. The probability of a state represents the likelihood that the state is or will be true at a future point. An effect probability is the likelihood that a state will become true given that a task has executed. For example, the "donate-vehicle" task has an effect probability of 0.95 on the state "have-transport". (There is a small chance that the vehicle is not operational.) A negative effect means that executing a task causes a state to become not true. Effect are static, i.e. don't change in the course of the simulation. Figure 1 shows a sample of the tasks and states of the SASO plan "run-clinic-there", effect and precondition relationships, and effect signs and probabilities. The rectangular nodes represent tasks and the rounded nodes represent states.

A state's current utility value represents the utility value to the human behavior model at the current point in the simulation. This value is based on the intrinsic utility value of the state, the tasks that affect the state, and the task and effect probabilities. When a task's probability changes as a result of an action that has occurred, the state probabilities of all affected states are updated. The utility value of a plan as a whole is calculated as the sum of the current utilities of all states belonging to the plan. The human behavior model starts out with an initial, default plan. Through the actions of the human participant (usually speech acts of the human participant), the probabilities of various tasks are affected, which in turn updates the utility value of the states and thus the plans. As stated above, the goal of the human participant is to get the utility value of his desired plan to be higher than the default plan or any other alternative plan.

The Virtual Human system maintains both domain knowledge and cultural knowledge. Domain (task) knowledge represents the knowledge needed for the human behavior model to perform its tasks, e.g. run the clinic, move the clinic, and help the patients. Cultural knowledge represents knowledge that is maintained about socio-cultural states of the world from the human behavior model's perspective, e.g. the agent's view of being familiar with the human participant, the agent's view of being neutral, and agent's view of being insulted. To the degree that the various Virtual Humans-based domains represent cultural knowledge, they do so by embedding the knowledge throughout the dialogue model and the task model, not in a modular fashion. The rules, data, and other software constructs

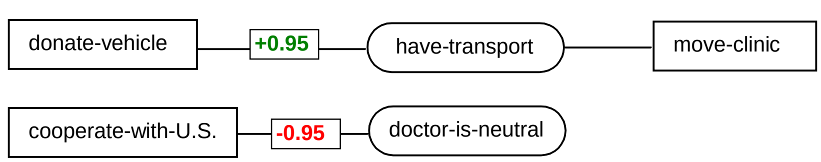

Figure 1 Sample Nodes of the Task Model Network for the SASO Plan "run-clinic-there"

that implement the socio-cultural aspects of the human behavior model's knowledge are not segregated from the constructs that model the domain knowledge.

Cognitive Cultural Architecture

In addition to the work described here, a number of other researchers are investigating how social science theories might be incorporated into human behavior models in order to better simulate social interactions. The Cognitive Cultural Architecture (Taylor et al, 2007) is one such example. Like the work presented here, the Cognitive Cultural Architecture seeks to create a computational model of cultural schemas with an emphasis on event schemas (which are described in more detail below). Taylor et al capitalize on a number of features of the Soar cognitive architecture to represent event schemas. This approach allows multiple schemas to be active at the same time, each with an activation level representing the prominence of the schema in the current situation. The Cognitive Cultural Architecture focuses on event schemas and external behavior while the work presented here focuses more on cognitive schemas and internal reasoning. Thus, these two approaches seem to complement each other and might be combined into a more complete system.

Etiquette Engine™

The "Etiquette Engine™" (Miller et al, 2007) is an implementation of a model of politeness developed by Brown and Levinson (Brown & Levinson, 1987). This model focuses on the perception and communication of relative power and familiarity in social interaction including different types of threat inherent in social interaction and redress strategies. The "Etiquette Engine™" models the degree of threat and degree of redress in an interaction with too little redress being perceived as rudeness and too much redress mapping to over-politeness. Many of the concepts in the "Etiquette Engine™" model, such as familiarity, are also represented in the work described here. However, the "Etiquette Engine™" provides a focused model of politeness which could be used to enhance our approach.

2.2 Social Science

Schema Theory

Schema Theory (D'Andrade, 1992) posits that culture is represented as a collection of schemas, which are a codified version of some abstract behavior or concept. D'Andrade describes at least two important classes of schemas; event schemas and cognitive schemas. Event schemas are behaviors organized in sequential steps for everyday situations. For example, how to go to a restaurant, how to greet someone, and how to purchase an item in a store are all stored as event schemas, to be retrieved, instantiated, and executed at the appropriate time.

A cognitive schema is an organization of knowledge around a particular concept, which can be triggered by symbols or images. For example, members of the U.S. culture associate the image of "golden arches" with "restaurant", and further, they fill in default information such as the fact that the restaurant has a drive-through.

D'Andrade (1984) proposes a type of cognitive schema, the constitutive rules system, to account for shared conceptual meanings. A constitutive rules system is defined as a set of rules that is known, shared, and adhered to by members of a culture and which defines some concept. The execution of certain behaviors, i.e. tasks, infers a cultural state that relates back to the concept. For example, going before a judge and having rites performed reinforces the cultural state "couple is married". What triggers this type of cognitive schema, in addition to the word or a symbol for the concept, is observing or performing any of the actions associated with the schema. When an individual sees a judge perform rites in front of a couple, the individual thinks of the concept of marriage and knows that this observation reinforces the belief that the couple is married.

Theory of Mind

The Theory of Mind, a theory in cognitive psychology (Nichols & Stich, 2003), posits that individuals have an ability to ascribe mental states such as intentions, beliefs, and values, to themselves and to others. In a social interaction individuals generate two mental representations. The first representation consists of the individual's own knowledge, intentions, beliefs, and values. The second representation, which is contained in the first, is a belief model of the other individual's knowledge, intentions, beliefs, and values. Both representations, the individual's own knowledge and his beliefs of the other individual's knowledge, are used in generating behaviors. In the absence of personal information, many of our initial beliefs about another's norms, beliefs and intentions are based on stereotypes about the other person's culture. Thus, modeling Theory of Mind is necessary to model cross-cultural stereotypes and biases.

3. Scoping culture

We define culture as the shared appearance, shared external behavior, and shared internal knowledge and reasoning of members of a group having a common identity. The effort described here seeks to address all three aspects of culture. However, in this paper we focus on the shared internal knowledge and reasoning aspect which we found to be the most challenging and interesting.

The shared internal knowledge and reasoning of a culture refers to socio-cultural values and attitudes ("norms"), as well as the mental representations of other people and cultures (“biases” and “stereotypes”). An example of a socio-cultural norm is the proscription against alcohol in Islamic cultures. Drinking alcohol detracts from one’s view of being observant of Islam. Drinking alcohol also detracts from one’s view of another individual’s view that one is observant of Islam. Suppose person A holds the cultural norm against alcohol in Islam and believes that person B is also aware of the rule. For A to drink alcohol detracts from A’s view of B’s view that A is observant of Islam (given that A thinks that B is aware of A drinking alcohol).

We propose that the social science theories described in Section 2.2 address multiple aspects of culture: external behaviors, cultural norms, and mental representations of others’ beliefs. As demonstrated by the Cognitive Cultural Architecture, event schemas in Schema Theory map directly to the external behaviors of culture. D’Andrade’s (1984) formulation of cognitive schemas addresses the cultural norms which are a key aspect of internal knowledge and reasoning. Finally, the Theory of Mind speaks to mental representations of other’s socio-cultural norms (biases and stereotypes). These mental representations are both a human behavior model’s direct perspective of the world (e.g. an agent’s view that the agent is insulted), and a human behavior model’s perspective of another entity’s views (e.g. an agent’s view of the participant’s view that the agent is insulted.)

D’Andrade’s cognitive schema approach analyzes the cultural norm against alcohol in Islam as follows: Islam as a concept maps to a constitutive system of rules which includes the rule “drinking alcohol violates being observant of Islam”. The constitutive system of rules is known, shared, and adhered to by members of an Islamic culture and, to a lesser degree, by members of other cultures with knowledge of the Islamic culture. For someone from an Islamic culture, drinking alcohol triggers the Islam cognitive schema, and detracts from the cultural state of being observant of Islam.

To continue the analysis farther, being perceived by others as being observant of Islam is also a cultural value or norm in Islamic cultures. In The Theory of Mind, being perceived as X is one’s mental representation of another individual’s mental representation that one is X. Combining a Theory of Mind representation of the cultural norms against alcohol with D’Andrade’s approach suggests this norm be implemented as a second rule: “drinking alcohol detracts from the norm of being perceived as being observant of Islam”.

4. Modeling Socio-Cultural Norms

We selected ICT’s Virtual Humans architecture as the basis for modeling socio-cultural norms expressed as constitutive systems of rules. We extended the Virtual Human task model formalism to model constitutive systems of rules by creating socio-cultural network that consist of tasks and socio-cultural states. As the task model formalism is already used to model domain states (e.g. “community-is-helped”) we considered the Virtual Humans to be a good fit to model socio-cultural states. Both domain states and socio-cultural states represent a human behavior model’s view of a state of the world. Domain states are more focused on concrete external facts about the situation (e.g. “have-transportation”) while socio-cultural states are more focused on internal facts about the agent’s feelings (e.g. “agent’s-view-that-agent-is-familiar-with-participant”). With our extensions the task model formalism becomes a language for expressing the socio-cultural norms which are a key aspect of cultural knowledge.

We have authored preliminary socio-cultural networks for the Iraqi-Sunni culture and the German culture, and also created a default domain plan and an alternative domain plan for a police captain human behavior model. As in the Virtual Humans, the human behavior model will execute the default plan (“perform-normal-police-duties-plan”) until the utility of the alternative plan (“help-participant-fix-market-problem”) exceeds the utility of the default plan. The socio-cultural network acts as an overlay on top of the default and alternative domain task models: together, both a domain task model and the socio-cultural network form one culturally-specific task model. By swapping the socio-cultural network to a new socio-cultural network, it is possible to achieve cultural knowledge modularity: maintaining the same domain human behavior model but with a different culture model.

The task model formalism is also an effective way to model mental states as posited in the Theory of Mind. By default, the states and utilities of a task model associated with a human behavior model are from the point of view of the human behavior model. Since the state and associated utility belong in the police captain agent, the state "community-is-helped" is more accurately "agent’s-view-that-community-is-helped". To model a mental representation of another entity's knowledge or beliefs only requires authoring parallel states that name the perceiver. For example, "agent’s-view-of-the-participant’s-view-that-community-is-helped" expresses the agent’s mental representation of the (human) participant’s view that the community is helped.

We modeled the socio-cultural norms for Iraqi-Sunni culture by creating a socio-cultural network for the

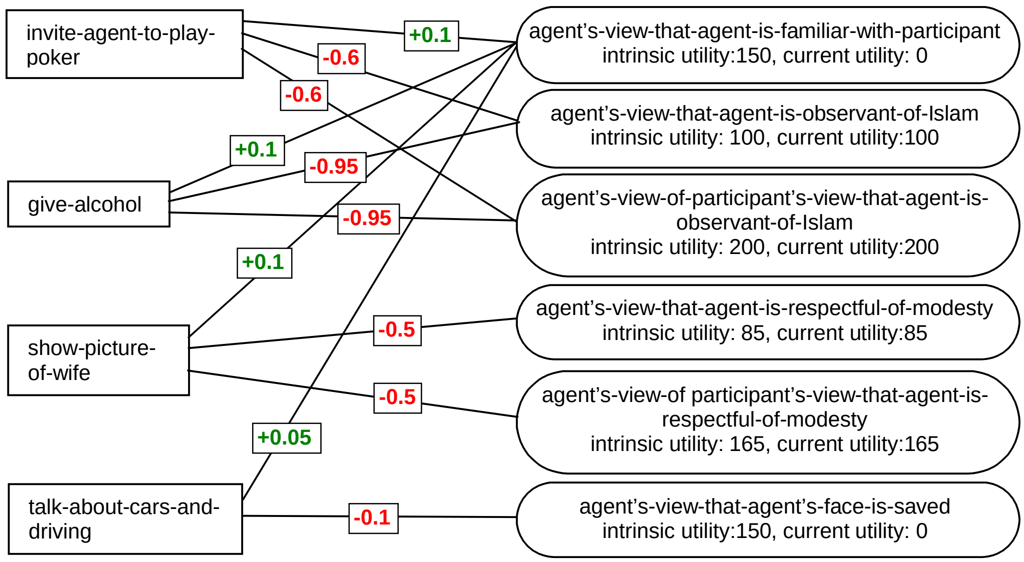

Figure 2 Sample of Socio-Cultural Network for Iraqi Sunni Culture

police captain scenario consisting of the socio-cultural tasks and states required for that scenario and the effects linking the tasks to states. In all, the Iraqi-Sunni socio-cultural network for the police domain consists of 34 tasks, 13 cultural states, and 44 effects between the tasks and states. Figure 2 above shows a representative sample of the Iraqi-Sunni socio-cultural network.

The rectangular nodes in the network represent tasks that may be performed in the simulation and the rounded nodes represent states. States have an intrinsic utility value and a current utility value. Starting current utility values are shown. Intrinsic utility values are similar to intrinsic utility values in the Virtual Humans. While an intrinsic utility in the Virtual Humans represents the importance that the human behavior model places on a domain state weighed against other states, in our extension to the Virtual Humans it represents the shared importance that members of a culture place on a socio-cultural norm weighed against other norms. An intrinsic utility value can be negative, in which case the state represents a socio-cultural norm with a negative connotation, i.e. the state is undesirable to members of a culture. Examples are "agent's-view-that-agent-is-insulted" and "agent's-view-that-agent-is-threatened".

A state's current utility value represents the utility value at any given time during the simulation. Current utility values fall within the range ($-|u_{int}|$, $|u_{int}|$), where u_{int} is the intrinsic utility. A negative current utility value means that the value of the state is opposite the meaning implied by the name of the state. For example, if the current utility value of "agent's-view-that-agent-is-respectful-of-modesty" is -85, then the agent believes that he is not respectful of modesty to the maximal degree.

The association lines are effects from a task to a state. Effects have a sign (+/-), which determines whether the effect reinforces or detracts from the state, and a degree. We use the term "degree" in a socio-cultural network where the Virtual Humans uses "probability", since in our cultural models a state degree is the degree to which a state is true, and an effect degree is the degree to which a task affects a state, not the likelihood that a state will be true at a future point or the likelihood a future task will affect a state.

Multiple effects on a single state are combined by summing the current utility value with the change in utility value caused by each effect, bounded within the range ($-|u_{int}|$, $|u_{int}|$). The change in the utility value caused by an effect, Δu, given that a task has executed, is calculated as:

$$\Delta u = s_{eff} \cdot d_{eff} \cdot u_{int}$$

where s_{eff} is sign of the effect, d_{eff} is the effect degree, and u_{int} the intrinsic utility of the state.

Each tuple consisting of (task, effect, state) models a rule of D'Andrade's constitutive rules approach to representing cultural concepts. For example, the tuple ("give-alcohol", "-0.95", "agent's-view-that-agent-is-observant-of-Islam") models the rule "drinking alcohol detracts from being observant of Islam" [1] and the tuple ("show-picture-of-wife", -0.5, "agent's-view-that-

[1] It should be noted that we have taken some shortcuts in these models. For the alcohol rule in Islam, we assume that giving someone alcohol implies a belief that the person regularly drinks or will drink alcohol at a future point.

agent-is-respectful-of-modesty") models a rule "looking at pictures of women in Western clothing detracts from being respectful of modesty" for the Iraqi-Sunni concept of modesty.

The task model quantifies the constitutive rules approach by assigning degrees to the effects and utility values to the states, which allows an author of a cultural model to reflect the fact that, within a culture, some effects and norms are more important that others. The values currently assigned to the effects and states are based on our non-expert knowledge of Iraqi-Sunni culture, and below, of German culture. To be considered valid, socio-cultural networks would have to be based on actual ethnographic data. Our plans for evaluating the current approach (see Section 7) includes having ethnographers create culture models using actual ethnographic data on cultural norms.

The tuple expression (task-name, (+|-) effect-degree, state-name) constitutes the grammar of a language for representing cultural norms and mental representations for a given culture. To test the flexibility of our language for modeling cultural knowledge, we authored a second socio-cultural network for German culture, also designed to overlay the police captain domain. To author a new socio-culture network, one has to identify:

- the socio-cultural states
- the intrinsic utility values of the states
- the starting current utility values for the states
- the effects from simulation actions to the states
- the effect degrees.

There is a modeling distinction to be drawn between the absence of a state and the presence of a state with zero intrinsic utility. Not having a cultural state implies that the state is not shared in the culture; having a cultural state where the intrinsic utility is 0.0 implies that the state is shared in the culture but no importance is assigned relative to other states. An example of the latter would be used to model religious norms held by secular members of a culture.

Figure 3 below shows a representative sample of the socio-cultural network that we authored for German culture.

The main differences between the Iraqi-Sunni socio-cultural network and the German network are the absence of the cultural states for Islam and the lack of associations between any actions and the cultural states for modesty. While the German character would have knowledge of Islamic cultural norms, the socio-cultural states for those norms are not activated in our scenario, since the German character is himself not a Muslim and is meeting with a non-Muslim American participant. Also, the German culture has the value of modesty, but none of the current set of simulation actions affects modesty as they do in the Iraqi-Sunni culture.

The three "morality" actions in the Iraqi-Sunni socio-cultural network in Figure 2 (i.e. "invite-agent-to-play-poker", "give-alcohol", and "show-picture-of-wife") only result in a positive effect on familiarity in the German socio-cultural network. Another difference in the samples shown concerns the action "talk-about-cars-and-driving". This action refers to small talk about one's own car, the car of the character, and driving in general. The action has a larger positive effect on familiarity in the German socio-cultural network based on the intuition that German culture is car-oriented, and that car and driving small talk builds greater familiarity than other topics. In Iraqi-Sunni culture we authored the same action as causing the police captain to lose face based on the idea that reminding the character that cars in present-day Iraq are a luxury lowers his self esteem and hence causes him to lose face.

When overlaying a socio-cultural network on top of a task model network, it is necessary to identify effects from domain tasks to socio-cultural states. This occurs when a simulation action created as part of the domain task model has socio-cultural side effects. For example,

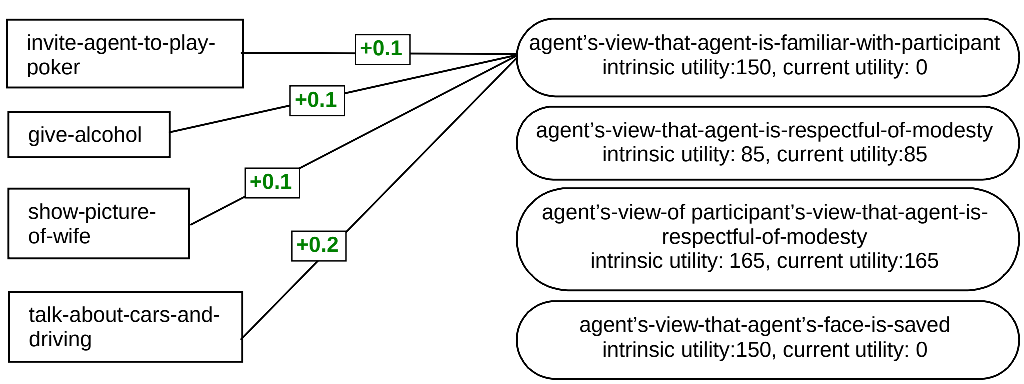

Figure 3 Sample of Socio-Cultural Network for German Culture

in our domain, there is an action “offer-police-uniforms-consisting-of-khaki-shorts” which has a positive effect on a domain state “agent’s-view-that-precinct-is-helped” but a negative effect on the two modesty states “agent’s-view-that-agent-is-respectful-of-modesty” and “agent’s-view-of-participant’s-view-that-agent-is-respectful-of-modesty” in the Iraqi-Sunni socio-cultural network. These links between the domain and socio-cultural network currently have to be hand authored. An automated process for finding the links, based on culturally-shared symbols, is an interesting topic for future research.

Finally, as we author additional culture models we will be compiling a library of socio-cultural states which an author could re-use when authoring new socio-cultural networks. We assume that related cultures share many cultural norms and hence socio-cultural states. We expect that this library will grow quickly as new socio-cultural states are added to address norms of new cultures; over time the number of new socio-cultural states that need to be added to the library for each new culture should drop off as states are increasingly reused. An interesting avenue of research would be to test this hypothesis as experts create culture models of multiple, related cultures using actual ethnographic data.

5. Generating Culturally-Affected Behavior

One purpose of the socio-cultural network is to calculate the utility of the network, which we call the Socio-Cultural Satisfaction (SCS) of the human behavior model. SCS represents the human behavior model's appraisal of the current situation against the set of his socio-cultural norms. Currently, we use SCS to affect the utility of domain plans, and as an input to the dialogue model. SCS affects domain plan utilities in that it is used to calculate intention probability, one of the factors in the calculation of task probability in the Virtual Humans. Intention probability represents the confidence that the human behavior model has in the human participant carrying out tasks to which he has committed. The lower the value of intention probability, the less likely these tasks are to happen, which in turn lowers the utility of states affected by the tasks and the overall utility of a domain plan. Thus, the overall effect of low SCS is to make the human behavior model less likely to agree to courses of action that the participant wants. In the future, different subsets of state utilities may be used for different purposes. For example, the calculation of intention probability may be restricted using the utility values of a subset of states, such as those related to familiarity, credibility and solidarity. Another use could be basing the acceptable phase change of a meeting from small talk to business on the utility of the familiarity set of states.

The Socio-Cultural Satisfaction of a socio-cultural network changes over time as the current utilities of the network states change in response to the execution of tasks. Recall that the change in a state’s current utility value caused by an effect, Δu, given that a task has executed, is calculated as:

$$\Delta u = s_{eff} \cdot d_{eff} \cdot u_{int}$$

where s_{eff} is sign of the effect, d_{eff} is the effect degree, and u_{int} the intrinsic utility of the state.

Consider the participant (human user) action “give-alcohol”. In the Iraqi-Sunni socio-cultural network (Figure 2), “give-alcohol” has a positive effect on “agent’s-view-that-agent-is-familiar-with-participant” and a negative effect on “agent’s-view-that-agent-is-observant-of-Islam” and “agent’s-view-of-participant’s-view-that-agent-is-observant-of-Islam”. When this action is executed, the value of Δu for “agent-is-familiar-with-participant” is calculated as $(+1) \cdot 0.1 \cdot 150.0 = 15.0$. Similarly, the value of Δu for “agent’s-view-that-agent-is-observant-of-Islam” is calculated as $(-1) \cdot 0.95 \cdot 100.0 = -95.0$ and for “agent’s-view-of-participant’s-view-that-agent-is-observant-of-Islam” as $(-1) \cdot 0.95 \cdot 200.0 = -190.0$ respectively. The total effect of “give-alcohol” is to decrease the SCS of the Iraqi-Sunni socio-cultural network by 270.0 (i.e. $\Delta u = -270.0$). This represents a drop of 34%, assuming that the current utility values of all states in the socio-cultural network are the starting values. Thus, giving alcohol to someone from an Islamic culture is a very negative social action.

In the German socio-cultural network (Figure 3), “give-alcohol” only has the positive effect on “agent’s-view-that-agent-is-familiar-with-participant”. The effect degree and state intrinsic utility are the same as in the Iraqi-Sunni model; the SCS for the German socio-cultural network increases by 15.0. This represents an increase of 4%, again assuming that the current utility values of states are the starting values. Thus, giving alcohol to someone from a German culture is a positive social action.

The action “talk-about-cars-and-driving” has different effect degrees in the two socio-cultural networks modeled. In the Iraqi-Sunni socio-cultural network the effect on the state “agent’s-view-that-agent-is-familiar-with-participant” is 0.05, while in the German network it is 0.20. Also, in the Iraqi-Sunni network the action has a negative effect of 0.10 on the state “agent’s-view-that-agent’s-face-is-saved”. Comparing the action in the two socio-cultural networks, we see a decrease in SCS of 7.5 (1%) in the Iraqi-Sunni network and an increase in SCS of 30.0 (8%) in the German network.

Figure 4 Farid and Fritz

As previously described, one way that SCS is used is in the calculation of intention probability. Another way that SCS is used in the current system is to determine the response of the human behavior model to user simulation inputs. The set of responses of the human behavior model to user actions consists of speech acts and associated animations; each speech act behavior is determined from a look-up table indexed by user action and the value of SCS. The less that the human behavior model is satisfied with socio-cultural state of the meeting, the more unfavorable the response, in tone and content. A third way that SCS can be used is as a trigger for the human behavior model to end the meeting, should the value of SCS fall beneath a set minimum floor.

6. Prototype

The CAB prototype consists of the CAB language, the extension to the Virtual Humans behavior model, and culture models integrated with the graphics front-end and animation and voice capabilities of ELECT BiLAT. We implemented our extended Virtual Humans architecture, task models, and culture models in Java and Jess, a rule engine.

The human behavior model implements a version of a character of ELECT BiLAT, a police captain, overlaid with either the Iraq Sunni or German socio-cultural network. In the prototype, the user can choose to meet with either the Iraqi police captain, named Farid, or the German police captain, named Fritz. When the human behavior model is loaded, first a file containing the police captain task model is loaded, followed by a file containing the socio-cultural network. Lastly, a third file containing any effect links between the domain and socio-cultural network is loaded. By loading a different socio-cultural file and links file, we are able to swap the culture of the human behavior model.

As in ELECT BiLAT, the user selects predetermined actions from a set of menus. The police captain starts out executing a default plan, “perform-normal-police-duties-plan”. The user’s goal is to persuade the police captain to provide police cooperation in solving a problem with a local market. By performing a mix of domain and socio-cultural actions (tasks), the utility (as influenced by the SCS of the socio-culture network) of the police captain’s alternative plan, “help-participant-fix-market-problem” is updated. If the derived utility of this plan exceeds the utility of the default plan, the police captain agrees to cooperate with the user. The simulation continues until the user exits the meeting (either having won the agent’s cooperation or giving up).

Figure 4 shows side-by-side screenshots of the Farid and Fritz characters.

7. Conclusion and Evaluation

We have presented an extension of the human behavior model of the Virtual Humans, providing a language for encoding cultural knowledge in the areas of norms, biases and stereotypes. Our approach also supports modular models of cultural knowledge, enabling the separation of domain knowledge from cultural knowledge and the ability to overlay characters with different culture models. We based our approach on cognitive schemas, in particular on D’Andrade’s (1984) constitutive system of rules approach, and on the Theory of Mind, and have identified event schemas as a foundation for modeling the external behaviors aspect of culture.

In the next phase of research we plan to evaluate our approach to modeling culture. We seek to determine whether the language for modeling cultural norms and mental representations is expressive and intuitive enough for an ethnographer to encode cultural knowledge. In a second experiment, we will determine whether a user can learn detailed knowledge about a fictional culture by using our prototype. To avoid contamination from previous cultural knowledge we will construct a completely artificial culture (e.g. aliens from Mars) for this experiment. Lastly, we will evaluate our characters for believability. If test subjects, who are native to the modeled culture, find our characters believable, then it would suggest that our approach to constructing human behavior models as a base domain behavior model overlaid with a culture model is valid.

8. Acknowledgements

We would like to thank Bill Swartout for conducting many lively discussions of research in the related realms of psychology, anthropology, and linguistics. Thanks also to Jon Gratch and Arno Hartholt for help in understanding the Virtual Humans architecture, and to Eric Forbell and Kelly Christoffersen for help with ELECT BiLAT. We would also like to thank Kim LeMasters for his support and Vadim Bulitko of the University of Alberta, Edmonton for his collaboration.

The project or effort 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.

9. References

Brown, P. & Levinson, S. (1987). Politeness: Some Universals in Language Usage. Cambridge, U.K.: Cambridge Univ. Press.

D’Andrade, R. (1984). Cultural Meaning Systems. In Schweder, R. and Levine, R. (Eds.), Culture Theory. Essays on Mind, Self, and Emotion. Cambridge, U.K.: Cambridge Univ. Press.

D’Andrade, R. (1992). Schemas and motivation. In D’Andrade, R. and C. Strauss (Eds.), Human motives and cultural models, pp. 23-44. Cambridge, UK: Cambridge Univ. Press.

Press.Gratch, J., & Marsella, S. (2004). A domain independent framework for modeling emotion. Journal of Cognitive Systems Research, *5*(4), 269-306.

Hill, R.W., Belanich, J., Lane, H.C., Core, M.G., Dixon, M., Forbell, E., Kim, J., & Hart, J. (2006). Pedagogically Structured Game-based Training: Development of the ELECT BiLAT Simulation. In Proceedings of the 25th Army Science Conference (ASC 2006), Orlando, FL.

Miller, C., Wu, P., Funk, H., Johnson, L. and Vilhajalmsson, H. (2007). A Computational Approach to Etiquette and Politeness: An "Etiquette Engine™" for Cultural Interaction Training. In Proceedings of the 2007 Conference on Behavior Representation In Modeling and Simulation. Norfolk, VA; March 26-29.

Nichols, S. & Stich, S. (2003). Mindreading. An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds. Oxford, U.K.: Oxford Univ. Press.

Taylor, G., Quist, M., Furtwangler, S. and Knudsen, K., (2007). Toward a Hybrid Cultural Cognitive Architecture. CogSci Workshop on Culture and Cognition, Nashville, TN, Cognitive Science Society.

Traum, D., Rickel, J., Gratch, J., & Marsella, S. (2003). Negotiation over tasks in hybrid human-agent teams for simulation-based training. Paper presented at the International Conference on Autonomous Agents and Multiagent Systems, Melbourne, Australia.

Author Biographies

STEVEN SOLOMON is a Research Programmer at the ICT.

MICHAEL VAN LENT has left the ICT and is now the Chief Scientist of Soar Technology, Inc. in Ann Arbor, MI.

MARK CORE is a Research Scientist at the ICT.

PAUL CARPENTER is a Research Programmer at the ICT.

MILTON ROSENBERG is a Special Project Manager at the ICT.