A Computational Approach to Etiquette and Politeness: An "Etiquette Engine™" for Cultural Interaction Training

Dr. Christopher Miller
Ms. Peggy Wu
Mr. Harry Funk
Smart Information Flow Technologies
211 First St. N. Ste. 300
Minneapolis, MN 55401
(612) 339-7438, (612) 339-7437
{cmiller, pwu, hfunk}@sift.info

Dr. Lewis Johnson
Dr. Hannes Vilhjalmsson
CARTE Labs USC/ISI
Address: 4676 Admiralty Way
Marina del Rey, CA 90292
(310) 448-8210
{Johnson, Vilhjalmsson}@isi.edu

Keywords:
Social Interaction Behavior, Etiquette, Politeness, Culture,
NPCs, Cross Cultural Training, Training Games

ABSTRACT: *Computational models and simulations of culture-specific social interactions are useful for a variety of applications including training, interaction or perception prediction and interpretation and even the design of machines and systems which will interact verbally with members of different cultures. To date, such models have been achieved via the scripting of specific trajectories of behaviors or, in limited instances, through very complex, "first principles" psychological models of agent motivation and goal-based behaviors. Neither of these is entirely appropriate for the production of a large series of rich and deep social interactions. We have developed a quantitative, computational implementation of a rich, universal theory of human-human "politeness" behaviors and the culture-specific interpretive frameworks for them (labeled "etiquette") from sociology, linguistics and anthropology. This model links observable and inferred aspects of power and familiarity relationships, the degree of imposition of an act (each of which have implications for roles and intents) and the actor's character to produce expectations about politeness behaviors. We have recently demonstrated the ability for this algorithm to produce culture-specific, politeness-appropriate utterances and perceptions of utterances in a game setting. Furthermore, our algorithmic approach to calculating culture-specific attitudes offered distinct advantages over alternate methods of producing such behaviors. Not only was it possible to more rapidly amass a much larger set of alternate trajectories through an interaction setting, but it was also possible to "swap" alternate sets of cultural knowledge to enable giving an agent alternate cultural perceptions with great ease.*

1. Introduction

Social interactions—that is, interactions based on the social characteristics and assumptions of each agent as an intentional entity (Dennett 1989) and drawing from culturally familiar patterns of expectations about appropriate behaviors—between humans and machines is receiving increasing attention (e.g., Preece, 2002; Miller, 2004) as machine and automation capabilities become more complex and more sophisticated. Similarly, awareness of the importance of culture-specific interaction patterns in multi-cultural human-human interaction (e.g., Hofstede, 2001) is driving an increased need for simulation of culture-specific socially interactive agents for training purposes in both military and commercial applications (Chatham & Braddock, 2003).

Yet accurate models and simulations of social interactions are rare. Rarer still are extensive models which offer a wide range of alternate social behaviors, representing a small fraction of the richness that actually exists in human-human social interactions. While there may be literally thousands of different ways that I could greet you, each of them conveying subtle nuances of my attitude toward you, toward our relationship and toward the world in general and my experience of it, most games and simulations will include, at most, one or two such interactions. Even sophisticated language and cultural training systems whose job it is to teach some of this nuance, rarely include more than a handful of such instances, representing a microscopic proportion of the richness which exists in human culture.

There are good reasons for this limitation. Chief among these is the fact that current techniques for producing such interactions are time-consuming and expensive. Most such simulations which do exist involve some form of scripted interactions—which have the drawback that they are costly to produce and encode and are generally ”brittle”—supporting interactions along only a narrow, pre-defined path with minimal variations. It has rarely been cost effective to encode simulations which support anything near the flexibility and breadth of true human-human social interactions. Nevertheless, to achieve the goal of interesting and effective training through social interactions in a game or simulation context will require that the agents used in training be “believable” both in the social interactions they exhibit (which must, in turn, be accurate with regards to the culture the agent is intended to be a member of) and in the breadth of actions which an agent of that sort would exhibit or could recognize and respond to.

Computerized Non-Player Characters (NPCs) don’t currently behave with the richness and fluency of social interaction behavior that we expect of them and are therefore, unbelievable in key ways. For example, it is entirely possible, in most “first-person” games which support any form of face-to-face social interaction besides combat, for me to insult non-player characters in a wide variety of ways with no response on their part except in those rare instances where I trigger a script through the use of a key word.

Accurately simulating cultural differences in social interactions requires “socially-aware” agents. Such agents take offense believably if not addressed in a culturally appropriate fashion, might appear recalcitrant or ignorant when they are merely trying to follow their culturally-derived notions of polite turn-taking in discourse, etc. Relevant social interaction behaviors, even those for different cultures and contexts, can frequently be emulated in hand-written scripts and simple, locally-relevant rules. But such approaches are time- and labor-intensive in their own right and brittle--only limited interaction complexity can be supported if every move has to be hand-scripted in advance.

To address this need, we have been pursuing a general theory and computational model of social interactions. The implementation of such a model in an algorithm could greatly enhance the usability and sophistication of NPCs, while improving the speed and/or reducing the cost of their construction.

2. “Politeness” for Social Interactions?

A significant class of social interaction behaviors have been studied in sociology, linguistics and anthropologyl under the heading of “politeness” behaviors. The term “politeness” is likely to evoke notions of formal courtesies and the use of “please” and “thank you”. But in these disciplines, politeness is a technical term having to do with the processes by which we determine and manage the “threat” inherent in communication and interaction between two intentional agents in a social interaction—that is, agents that are presumed to have goals and the potential to take offense at having those goals thwarted in any interaction where those intentional attributes are relevant (cf. Dennet, 1989; Goffman, 1967). As we see below, politeness in this sense is the method by which we signal, interpret, maintain and alter power relationships, familiarity relationships and interpretations of the degree of imposition or urgency of an act.

A seminal body of work in the study of politeness is the cross-cultural studies and resulting model developed by Brown and Levinson (1987). Brown and Levinson noted that people across cultures and languages very regularly depart from strictly efficient conversation by using an array of conversational behaviors designed to mitigate or soften direct expressions of desire, intent or command. A simple example in English will illustrate the point: as we settle down to a meal together and I ask you “Please pass the salt,” the use of “please” in that sentence is unnecessary for a truthful, relevant or clear expression of my wish and is, in fact, an explicit addition of verbiage not required to express my intent (to have the salt passed to me).

Brown and Levinson collected and catalogued a set of politeness behaviors across multiple cultures and language groups and, from this data, developed a general theory and qualitative model of the role of politeness in social interactions. Their model of politeness will be explained next.

2.1 Perceived Politeness as a Function of Threat and Redress

The Brown and Levinson model assumes that social actors are motivated by two important social wants based on the concept of “face” (Goffman, 1967). “Face” is a complex concept in anthropology, but it has been loosely summarized as the "positive social value a person effectively claims for himself" (cf. Cassell and Bickmore, 2002, p. 6). It is the desire to have one's will and interests be seen as important and valuable. Face can be "saved" or lost, and it can be threatened or conserved in interactions. Brown and Levinson further refine the concept of face into two specific subgoals that all social actors can be presumed to have:

1. *Positive face*—an individual's desire to be held in high esteem, to have his/her actions and opinions valued, to be approved of by others, etc.
2. *Negative face*—an individual's desire for autonomy, to have his/her will, to direct his/her attention where and when desired, etc.

Virtually all interactions between social agents involve some degree of threat to face and are, therefore, "Face Threatening Acts" (FTAs). My simple act of speaking to you, regardless of the content of my words, places a demand on your attention that threatens your negative face, for example. This, then, is the reason for the "please" in my request for salt: If I simply state my desire that you give me the salt as bald propositional content (e.g., "Give me the salt") I may efficiently communicate that intent, but I have also been ambiguous about whether or not I have the power or right or can otherwise compel you to give me salt. You might well take offense at the implication that I could demand salt from you.

The "please" in the example above is therefore a politeness strategy used to "redress" or mitigate the threat contained in the request for the salt. Furthermore, the expectation that such a strategy be used in certain contexts is an example of etiquette that enables interpretations. The etiquette which we believe to be in play entitles us to conclude that those who use "please" in an appropriate context are striving to play by the rules—striving to be seen as polite; those who do not are not striving to be polite for various reasons (perhaps they don't believe they need to be, perhaps their notions about politeness are different, perhaps they are just rude).

The core of Brown and Levinson's model is the claim that the degree of face threat posed by an act must be redressed or balanced by the value of the politeness behaviors used if the social status quo is to be maintained. That is:

$$W_x \cong V(\mathbf{A}_x)$$

- Where W_x is the "weightiness" of severity of a face threat x, and
- $V(\mathbf{A}_x)$is the combined redressive value of the set of politness behaviors $(\mathbf{A}_x)$ used in the interaction.

If less redress is used than is perceived as necessary, that is if $W_x >> V(\mathbf{A}_x)$, then the utterance will be perceived as rude and the hearer may seek alternative explanations or interpretations for the behaviors, as will be discussed below. If more politeness behaviors are used than are perceived as necessary, that is if $W_x << V(\mathbf{A}_x)$, then the utterance will be perceived as "over-polite" or obsequious and, again, ulterior motives for the behaviors or ulterior interpretations of the context may be sought.

Thus, perceived politeness is a function of the Imbalance (I) between the degree of face threat in an interaction and the amount or degree of redress used. We can express this as:

$$B_O{:}I_x = B_O{:}V(\mathbf{A}_x) - B_O{:}W_x$$

Where B_O is the Belief of Observer O about the other terms in the equation and I_x is the perceived Imbalance (I) of interaction x. Thus, this equation says that the believed imbalance as perceived by Observer O of interaction x will be the difference between the value of the redressive acts **A** in x (as perceived by O) minus the amount of face threat W (as perceived by O). Imbalance will be positive when more redressive politeness behaviors were used than there was face threat present—corresponding to the overly polite or obsequious condition. I_x will be negative when less redress is used than there was threat—a rude condition.

This model also answers a fundamental question about politeness use—namely, the fact that the same set of politeness behaviors, used in different contexts, may well be perceived as anything from appropriate to rude or over-polite. It is clear that the same degree of redressive value $(V(\mathbf{A}_x))$ may be too much, too little or just right depending on the value of the face threat present. Of course, this leaves open the question of how face threat is determined. This aspect of the Brown and Levinson theory will be discussed next, followed by a discussion of how to assess redressive actions and their values within this framework.

2.2 Computing the severity of a face threat

In the Brown and Levinson model, the degree of face threat of an interaction or utterance is provided by the function.

$$W_x = D(S,H) + P(H,S) + R_x$$

- W_x is the 'weightiness' or severity of the "Face Threatening Action" (FTA) x

- D(S,H) is the social distance between the speaker (S) and the hearer (H). Social distance is roughly the inverse of familiarity—D decreases with contact and interaction, but may also with be based on a priori factors such as membership in the same family, clan or organization and perhaps on being in a "familiar" setting as opposed to a formal one—a sporting event rather than a court.

- P(H,S) is the relative power that H has over S. Power comes from different sources in different cultures and organizations. Clearly, a tutor needs to maintain some power over a student, but NPCs representing commanders, subordinates, or high or low status citizens might all need to act, and to be handled according to different etiquettes if face threats are to be minimized. Power is an asymmetric relationship between S and H.

- R_x is the ranked imposition of the raw act itself. Some degree of imposition is culturally defined—it may be inherently more of an imposition to request food from a host in Western culture than in an Arabic one, for example. But imposition is also dependent upon the roles and duties of the parties involved. One reason a tutor can correct a pupil, even though s/he might have lower power in the society, is that the correction is expected from the tutor and is, therefore, less of an imposition.

2.3 Redressing face threats

Since FTAs are potentially disruptive to human-human relationships, we generally make use of redressive strategies to mitigate the degree of face threat imposed by our actions. The bulk of the Brown and Levinson work is represented by a host of well-researched examples of the use of redressive behaviors from at least three different language/culture groups (English, Tamil and Tzeltal). They have organized these examples into a structure of mutually supporting and incompatible approaches. We do not have space to present their findings in depth, but we note as an example that their categorization of one type of redress strategies ("negative redress"—those strategies focused on minimizing the impact on a hearer's "negative face" by minimizing the impact of the interaction and/or by explicitly recognizing and explicitly offsetting the face threat) contains 10 alternate approaches, some of which are mutually supporting or conflicting, including:

- *Be Pessimistic*—"You're not going to pass me the salt, are you?"
- *Minimize the Imposition*—"Could you just nudge that salt shaker over here?"
- *Give Deference*—"Excuse me, sir, would you pass the salt?"
- *Apologize*—"I'm sorry to interrupt, but would you pass the salt?"

3. Algorithm Implementation and Example Cases

In work funded DARPA under the Force Multipliers for Urban Operations program, we have completed an implementation and demonstration of an algorithm based on Brown and Levinson's work as described above. Our demonstration illustrates the use of this algorithm in the context of a language training game in which our politeness algorithm, which we have labelled the "Etiquette Engine™," enables NPCs to both recognize the degree of politeness directed at them and to reason about the level of politeness to be used in an interaction they themselves issue in keeping with other goals the character may have. Our approach and results will be described in this section.

3.1 The Implementation Environment: TLTS

Our implementation of a computational politeness algorithm has been in the context of the Tactical Language Training System (TLTS) developed by the University of Southern California's CARTE Laboratory (Johnson, Vilhjalmsson and Santani, 2005). TLTS is a first-person game/simulation designed to teach soldiers "tactical" versions of a language (to date, Lebanese Arabic, Iraqi Arabic and Pashto). The trainee navigates scenarios wherein s/he must interact with simulated local inhabitants to accomplish an overall mission: e.g., the trainee might need to ask a group of young men in a café (Figure 1) who the leader of the village is in order to progress to the next scenario, wherein s/he will meet the village leader. By using specific phrases and gestures, the trainee must convince the young men to provide the information. TLTS uses speech recognition techniques to process the trainee's verbal utterances in the selected language, and offers a set of mouse-selectable gestures (such as taking off one's hat or sunglasses, covering one's heart, shakng hand, etc.) to accompany actions. If the trainee's Arabic is not up to the task, s/he will fail, but more than simple language skills are required. TLTS is also concerned with politeness in word and gesture. If the trainee is rude by local cultural standards, the men may well conclude that s/he is a spy and refuse to offer any information.

Prior to our involvement, TLTS used ether a traditional scripting approach or a complex, first-principles reasoner emulating human goal-based reasoning and decision making (Prynadith and Marsella's (2005) PsychSim).

Figure 1. An example of interactive NPCs in USC/CARTE Labs' Tactical Language Training System (TLTS—Johnson, et al., 2005).

Neither of these was an entirely satisfactory solution. Traditional scripting appraoches greatly limited the range of etiquette situations which the trainee could experience because they required extensive effort to implement a range of alternatives and, in the absence of such effort, were "brittle"—that is, they provided acceptable behavior only for the narrow path of interactions which had been scripted. When novel, unscripted behaviors were attempted in the context of the game, NPC behaviors would either be inappropriate or absent completely. For example, while a small range of greeting and response options were encoded, it was entirely possible for the trainee to use an insult in the context of the greeting with no ill effects since the game did not recognize insult phrases. Similarly, it had limited sensitivity to the "layering" and combinatorics of politeness behaviors (e.g., using an honorific and an apology and impersonalization vs. using only an honorific).

The use of the detailed goal-based reasoning module, PsychSim, eliminated the problem of brittleness in principle, but not always in practice. PsychSim offers a rich and deep representation of human goal-based reasoning and includes a similarly deep reasoning about the intentions of other agents. This complexity, while far richer and more general than the etiquette and politeness reasoning we proposed and, thus, even less brittle. The problem was that the richness PsychSim offered resulted in a need for substantial development effort and produced comparatively unpredictable results. When development effort was not adquate, incorrect interaction behaviors could result. Even when it was adequate, however, the results were not always as controllable as was desired for a training simulation which, after all, needed to adhere to a reasonable "lesson plan" and progression through the stages of the game.

3.2 Implementation of the Etiquette Engine in TLTS

As an augmentation to these approaches, SIFT created an "etiquette module"—the Etiquette Engine™-- which implemented our computational version of the Brown and Levinson model. Details of our representation and scoring of important politeness dimensions, along with specific redressive behaviors, may be found in Miller, Wu, Funk, Wilson and Johnson, 2006) and are beyond the scope of what can be included in this paper. Instead, we will describe the use of this scoring algorithm in the context of a language training simulation below.

The Etiquette Engine (EE) operated in two alternate modes with TLTS, as illustrated in Figure 2. First, for an NPC to recognize the level of etiquette directed at it, EE received as input from TLTS the recognized utterance (as transcribed text) from TLTS's speech recognition

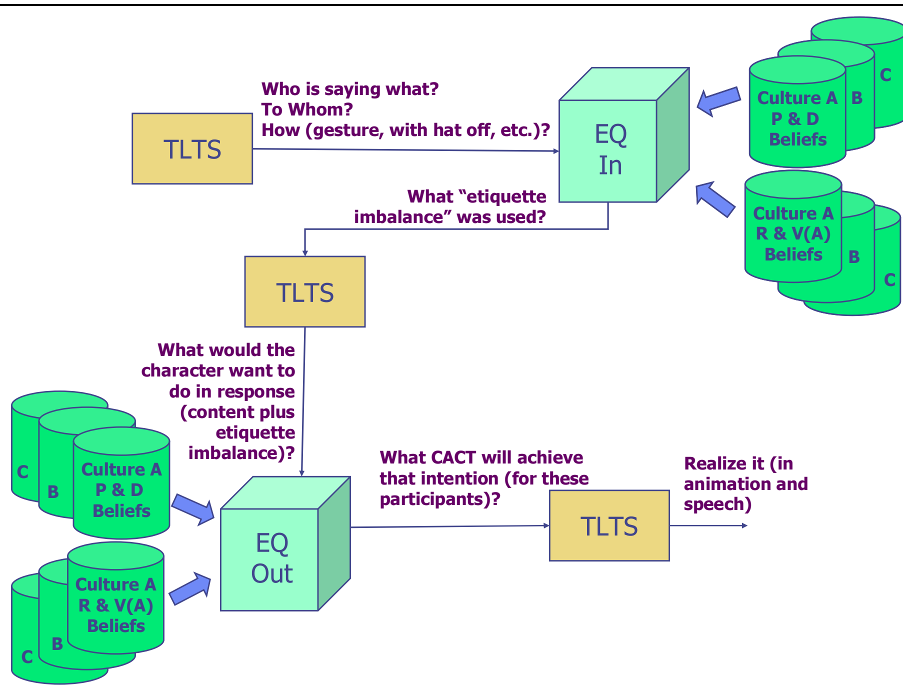

Figure 2. Architecture of SIFT's EE integrated in USC's TLTS.

algorithms, along with any user-selected gestures. EE maintained a knowledge base of beliefs that each individual held that enabled it to compute the relative power and social distance between each pair of characters. A second knowledge base contained data on that character's beliefs about the relative imposition of a type of speech act and the redressive value of sets of politeness behaviors. These knowledge bases could be varied from individual to individual with inheritance for cultural groupings. By consulting the knowledge bases appropriate to the character of interest, we could determine that characters' beliefs about each of the values in the equations above and, thereby, compute a perceived imbalance score for that utterance for that character.

The second function of the EE was to recommend "communicative acts" (CACTS—combinations of verbal utterances plus non-verbal content) for NPCs to deliver in accordance with their etiquette goals. The etiquette goal could be expressed as a modulation of the status quo (i.e., the character might want to be more or less polite or neutral for the current context), and therefore, could be expressed as a positive or negative imbalance number ranging from positive to negative 1000. The EE then examined alternate CACTs of the type requested (GREETING_RESPONSE) and selected or composed one whose legal combination of redressive acts was approximately equal to the desired imbalance number.

3.3 EE at Work: A Computed Politeness Example

A brief and simple example (drawn from our work in the Pashto language and culture) may clarify the approach. Let's say that the trainee, playing the role of a sergeant whose mission is to contact the local headman of a Pashto village (the "Malek") and enlist his aid in building a clinic. For this specific example, we will assume that they have just met and exchanged a round of initial greetings ("Salaam aleekum" and "wa aleekum salaam"). After being introduced to the Malek by name, the trainee says "staasee de lidelo tsexa xoshala shwem, saaheb" with a double-handed handshake (DHS) gesture—meaning, "I'm very happy to meet you, sir" with a very warm, friendly gesture. In order to score the Malek's interpretation of this interaction, we need to know what the Malek thinks about the power and familiarity relationship between himself and the trainee, his perception about the imposition of the speech act type the trainee has just used (a RECOGNITION), and finally, his understanding of the various redressive behaviors used and their relative values.

To be brief, P, D, and R are scored on -1000 to +1000 point scales centered around a neutral value of 0. For this example, and derived from the scenario context of the game, we used the following values:

- ❑ The Malek believes he has slightly more power than the trainee does (P = 20)
- ❑ The Malek believes that he and the trainee are not quite as distant as complete strangers would be—they have mutual acquaintances and a mutual objective (D = 15)
- ❑ There is a very slight imposition to the trainee taking the initiative in offering his RECOGNITION speech act before the Malek recognizes him, but this value isn't large in this culture at this stage of the conversation (R=5).

The trainee used several polite redressive strategies here. The values and scores for these are:

- ❑ 15 points of redressive value for the use of an exaggerate approval strategy (a type of positive redress): "I'm very happy to meet you, sir."
- ❑ 30 points of redressive value for the use of formality, a form of giving deference, a negative redress strategy: (this greeting is much more formal and uses formal pronouns than other possible ones)
- ❑ 20 points of redress for an honorific, a negative redress strategy: "I'm very happy to meet you, sir."
- ❑ 35 points for the double hand shake—an enthusiastic, inclusive gesture and therefore, positive redress.

Given these values, we can compute the Malek's perception of this exchange:

$$I_x = B_O{:}V(\mathbf{A}_x) - B_O{:}W_x$$
$$= (15+30+20+25) - (20+15+5) = 60$$

An Imbalance of 60 is moderately over-polite—a bit warmer than the Malek would have expected from the Trainee, but probably not so extreme as to make him question his interpretation of events. In fact, this exchange is likely to put him into a good frame of mind for future negotiations. Nearer term, the Malek will choose his next utterance— a RECOGNITION_ RESPONSE—so as to attempt to match the politeness level the Trainee used.

4. Scale Up with the EE Algorithm

This approach of parsing the perception of politeness into subcomponents (P, D, R, and $V(\mathbf{A}_x)$) opens up the possibility of recombining elements to greatly expand the set of possible utterances captured in a system—just as understanding vocabulary and the rules of syntax enable the construction of all possible sentences in a language. Unlike the linear scalability of traditional scripting approaches, where each subsequent interaction must be developed from scratch with essentially the same cost in

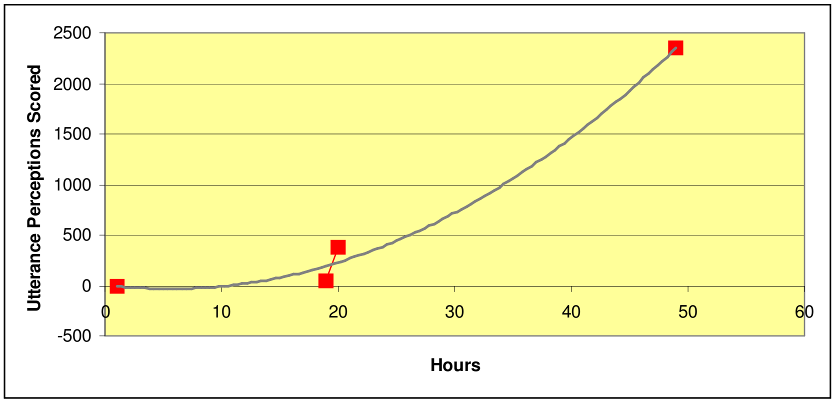

Figure 3. Scale up data for Perception Scores represented during work with TLTS.

labor as the one before it, there is reason to believe that our approach scales geometrically (cf. Figure 3).

By explicitly representing the knowledge to compute how the Malek perceives "staasee de lidelo tsexa xoshala shwem, saaheb" from the Trainee, we have made it very easy to compute how any other character will perceive that speech act from any other speaker—all we need to do is represent the believed P and D values for the new pair. Similarly, knowing the value of adding "saaheb" (an honorific worth 20 points) to the above utterance, means that we know the relative worth of using that positive redress strategy for any utterance in which it makes sense. In fact, as is shown in Figure 3, during the project described above, we demonstrated this scalability by acquiring the knowledge for and encoding our first set of 42 "Perception Scores" (PSs—how one observer perceives the politeness of one CACT uttered by a speaker-hearer pair) at the rate of 2.33 PSs/hour, but the next set were acquired at 19.89 PSs/hour, and the final set of more than 2000 PSs were acquired at the rate of 48.96/hour.

It is important to note that the values used above to score the terms in the Imbalance equation are not arbitrary, nor are they based solely on the programmers' best judgment. Brown and Levinson provide guidance as to the categories and types of redressive behaviors and the dimensions that influence face threat. They even provide guidance about the likely relative value of different classes of strategies. We developed our computational model by making initial claims about the likely value ranges and then adjusting them as novel examples were obtained. We have reduced these techniques to a coding manual and, using it, have obtained interrater reliability scores of .931 using Robinson's A correlation among three raters across 8 vignettes for the top level imbalance score I_x. Similarly, for the two major subfactors (Face Threat Weight—W_x and composite Redress Value—$V(\mathbf{A}_x)$) the Robinson's A correlations were.950 and .863 respectively. These values are all well above traditional thresholds of .7 or .8 for multiple judge rating correlations.

In practice, we obtained the values for Pashto interpretations through a series of interviews and evaluations of the algorithm's outputs conducted with a native Pashto speaker. But acquiring the knowledge to develop these knowledge bases is a process we have not formalized to any significant degree. We believe that the Brown and Levinson model, with our extensions and formalization of it, provides an excellent basis on which to structure efficient and reusable knowledge acquisition, and that role playing is an excellent approach to observing the use of politeness behaviors in a reasonably realistic context.

A final advantage to our structured and computational model of politeness interactions should be noted. Our model-based approach to reasoning about politeness and etiquette separates culture-specific knowledge about what constitutes polite interactions from the core algorithm for computing perceived politeness. Brown and Levinson's claim is that their model of face threat and redress is culturally universal—all cultures use politeness in this way. There are certainly differences in what constitutes redressive behavior from culture to culture—for example, taking off my hat is a polite gesture in America and most western cultures, but it has little meaning in Iraq—where removing one's sunglasses might be a more readily understood polite gesture in most contexts. Furthermore, there are certainly differences from culture to culture in how power, familiarity and imposition are recognized and valued. Nevertheless, the Brown and Levinson claim, based on their years of data acquisition and analysis, is that these factors will work in similar ways across all cultures.

In our implementation, this means that the core algorithm for computing perceived politeness can be used to emulate the politeness behaviors and perceptions of any culture with no modifications. Culture-specific knowledge about redressive behaviors and their values, as well as about the factors that constitute power, social distance and imposition, can be stored in separate knowledge bases which we refer to as "cultural modules". This architecture is depicted in Figure 4. By swapping cultural modules we can, almost literally, provide an NPC with a "brain transplant". The NPC portraying the Malek

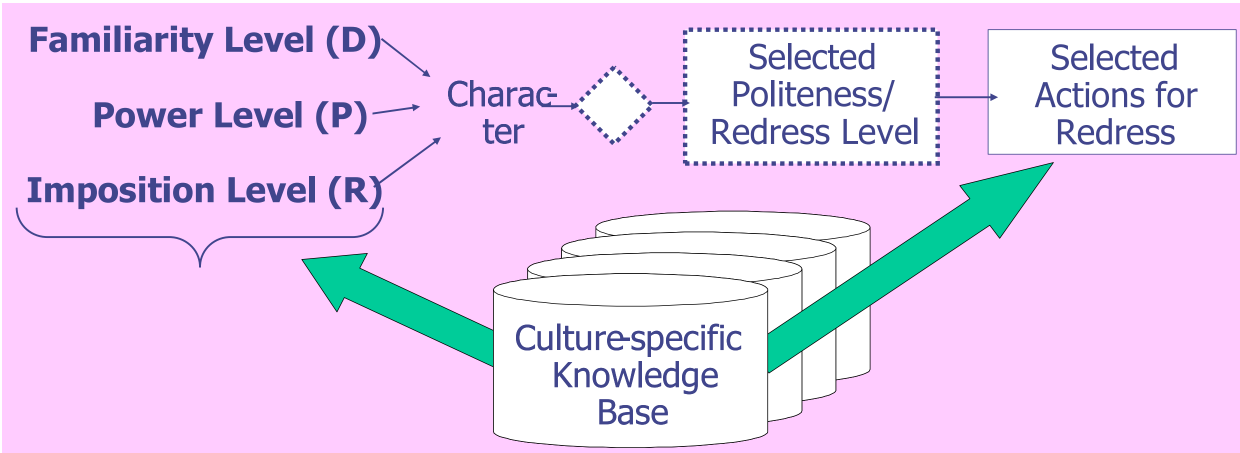

Figure 4. The role of "culture modules" of culture-specific politeness knowledge in our implementation of the Brown and Levinson model.

in Pakistan can, with the flick of a software switch, instead think with the perceptions and biases of, say, a Kosovar village priest. While the construction of culture-specific modules of relevant etiquette knowledge is a non-trivial task in its own right, the framework of our model can serve to guide and limit such knowledge acquisition and representation and, once implemented, both the scaling arguments made above and the reusability of the core reasoning algorithm provide further advantages in developing believable characters representing alternate cultural backgrounds and perspectives. Granted, this does nothing to alter the graphical appearance of the character or its surroundings (arguably more resource intensive software development considerations), it nevertheless represents an important innovation which enables more cost-effective development of culture-specific emulations in gaming and training software.

In our work developing and demonstrating the EE approach, we have illustrated this capability to swap cultural modules in the context of the scenario concerning introductions with the Malek. Via a simple user interface to the EE which we have developed (see Figure 5), a user (e.g., a software developer) may assert that various alternate cultural modules be used in computing individual's perceptions. Then, in one example, we have captured U.S. perceptions and redressive behaviors to be used for the U.S. characters in the scenario, while capturing Pashto behaviors and perceptions for the Pashto characters. This then allows us to, for example, inspect the perception of different characters of an observed behavior or, alternatively, to generate alternate behaviors which different characters might choose to use to achieve the same intended level of politeness. For example, as mentioned above, a naïve American might be represented as assuming that taking off his hat will be a gesture with a redressive value of 30 points. A naïve Pashto character will not have this gesture among his repretoire (as represented in the cultural module used by that individual), thus it will have no value in his perceptions. Instead, he might have the knowledge that keeping one's sunglasses on in personal interaction is an offensive gesture (worth -30 points). Since the naïve American does not have this alternate piece of knoweldge about redressive behaviors, he might well greet the Malek by taking his hat off while leaving his sunglasses on—expecting and intending this to be perceived as a moderately polite interaction. Instead, the Malek (using the naïve Pashto module) would fail to perceive any politeness from the hat removal, but would see rudeness in failing to take off the sunglasses.

5. Conclusions

This work represents the development of a simple, yet powerful, algorithm for reasoning about important aspects of culture-specific social interactions. NPCs that exhibit realistic social interaction behaviors—and a realisic range of variation in such behaviors—are important for the next level of believability and engagement in a wide variety of commercial and entertainment applications. Beyond that, the ability to exhibit culturally realistic social interaction behaviors are absolutely critical for the development of language and culture training applications.

While our approach cannot be said to embody all aspects of human etiquette and politeness interactions, much less provide a detailed model of human goal-based reasoning as it applies to human-human social interactions, it nevertheless holds promise for dramatically improving the ability to rapidly create computer training simulations or games with realistic, culture-specific social interaction models for their NPCs. We have sought, and the data and examples presented above provide initial arguments that we have found, a "sweet spot" between robust but complex and difficult to use first-principles models of

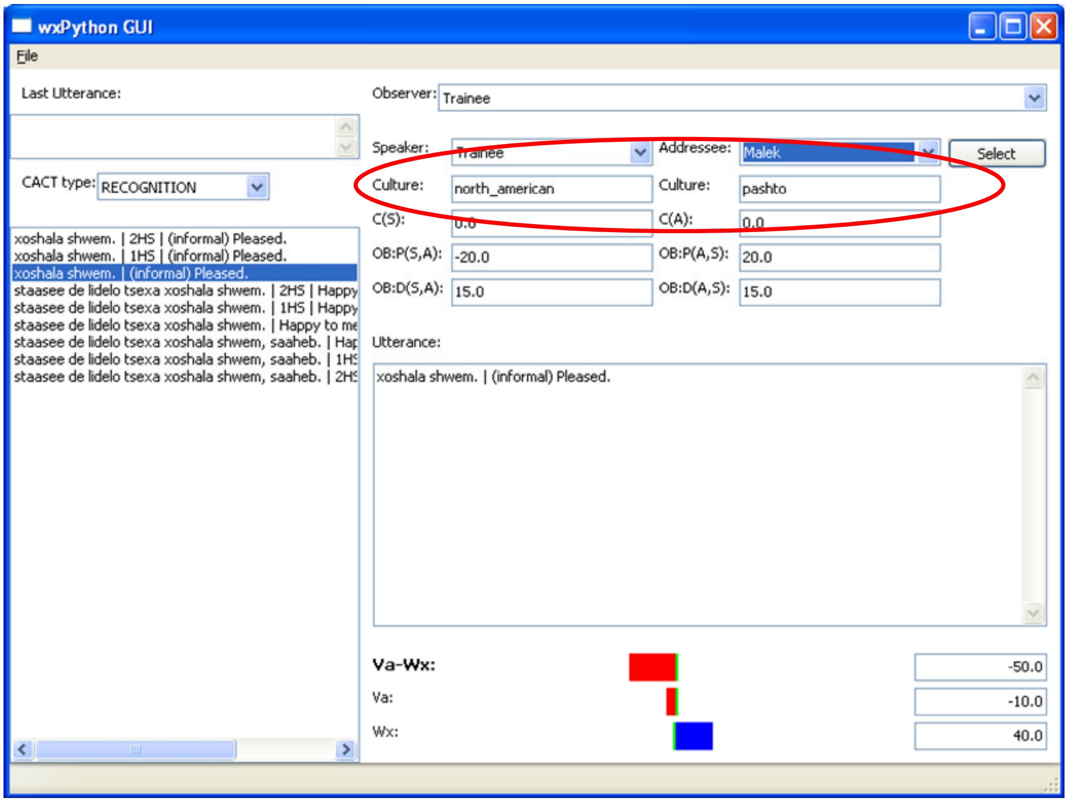

Figure 5. Interface for selecting the cultural module to be used with each agent and for visualizing differences in etiquette perception.

human interaction, and the easy to use but brittle and costly aspects of traditional scripting approaches.

6. Acknowledgements

This material is based upon work supported by the Defense Advanced Research Projects Agency and U.S. Army Aviation and Missile Command under contract number W31P4Q-04-C-R221. We would like to thank Dr. Ralph Chatham, our Program Monitor.

This document was approved for public release, distribution unlimited by the DARPA Technical Information Office.

7. REFERENCES

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

Cassell, J. and T. Bickmore. (2003). Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. *User Modeling and User-Adapted Interaction. 13(1):* 89-132

Chatham, R., and J. Braddock, (2003). *Training for Future Conflicts*, Report of the Defense Science Board.

Dennet, D., (1989). *The Intentional Stance*, Cambridge, MA; MIT Press.

Goffman, E. (1967). *Interaction Ritual: Essays on Face to Face Behavior.* Garden City; New York.

Hofstede, G. (2001). *Cultures and Consequences, 2nd Ed.* Thousand Oaks, CA; Sage Publications.

Johnson, L., Vilhjalmsson, H. and Samtani, P. (2005). The Tactical Language Training System (Technical Demonstration), *The First Conference on Artificial Intelligence and Interactive Digital Entertainment*, June 1-3, Marina del Rey, CA

Miller, C. A. (Ed.), (2004). Human-Computer Etiquette. *Communications of the ACM, 47(4*). 30-61.

Miller, C., Wu, P., Funk, H., Wilson, P, and Johnson, L. (2006). A Computational Approach to Etiquette and Politeness: Initial Test Cases. In *Proceedings of the 15th Conference on Behavior Representation in Model-ing and Simulation (BRIMS*). 15-18 May 2006; Baltimore, MD.

Preece, J. (Ed.) (2002). Supporting community and building social capital. *Communications of the ACM 45(4)*, 36-73.

Pynadath, D. and Marsella, S. (2005). PsychSim: Modeling Theory of Mind with Decision-Theoretic Agents. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1181-1186.