A Computational Approach to Etiquette:

Operationalizing Brown and Levinson's Politeness Model

Christopher A. Miller, Peggy Wu, Harry B. 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

Abstract

A central source of cultural differences, with powerful impacts on perception and behavior, is communication of "politeness" and its role power and familiarity relationships, urgency, indebtedness, etc. We are operationalizing and making computational a culturally abstract and universal theory of human politeness which combines culture-specific aspects of social context (power and familiarity relationships, imposition, character), to produce expectations about politeness behaviors (also culturally defined). Such a model will enable better training materials, simulations, and even better decision aids. By using observations of politeness behaviors (or their lack), the same model can infer those attributes. We describe our algorithm and results from two validation experiments. We have used this model to guide simulated game agents in interpreting and generating politeness behaviors and have demonstrated promise for reducing software development costs and/or increasing an agent's behavior repertoire through the creation of modular, cross-cultural etiquette libraries.

Keywords: Culture Modeling, Social Interaction, Politeness, Etiquette, Power, Familiarity, Imposition, Urgency

Introduction

We are developing computational models of *social interactions*—that is, interactions based on the social characteristics of each agent as an intentional entity[1] and drawing from culturally familiar patterns of expectations about appropriate behaviors. Understanding social interactions becomes of obvious value when we need to get a job done with team members. Team communications inevitably require communication about power, trust, familiarity, group membership, urgency, importance, etc.—all in addition to explicit task-related content. Communication difficulties in multi-cultural teams can be exacerbated by cultural differences in how these social interaction variables are conveyed.

There has been much theoretical and basic research on identifying cultural patterns[2] and how cultural factors affect cognitive processes[3] but few of the resulting models directly link cultural factors to human behavior, and few are readily amenable to computational modeling at a fine level of granularity. Even in the comparatively richly studied field of culture and negotiation,[4] studies have generally found cultural correlates for negotiation behaviors and attitudes without developing a computational model of cultural effects.

We have instead focused on developing a concrete model of observable human behaviors—specifically communication behaviors having to do with etiquette and politeness—which in turn have relevance for human performance, attitudes and broader decision making. We have been leveraging a qualitative sociolinguistic model both because it accounts for many effects of social interaction behaviors on team interactions and provides a culturally universal, yet instantiable, basis from which to develop culture-specific models. Such a model will be useful in evaluating, adapting, interpreting and generating social interaction behaviors in both human-human and human-machine interactions. We report here our efforts to develop a quantitative, computational model of politeness interactions and to both apply and evaluate it in different contexts.

"Politeness" for Social Interactions?

The terms "etiquette" and "politeness" are likely to evoke notions of dinner forks and curtsies—phenomena that seem to have little impact on practical, work-related activities. But *politeness*[5] is the processes by which we determine and manage the "threat" inherent in communication between intentional actors who are presumed to have goals and the potential to take offense at having those goals thwarted[6] (see also reference 1). Politeness is one means by which we convey, interpret, maintain and alter social relationships. *Etiquette* is the code by which we signal politeness. It makes use of verbal, physical, gestural and even more primitive modes of interaction. For example, deference can be expressed by posture, quiet speech and/or explicit markers such as titles and honorifics—though the specific

behaviors will vary across cultures, since each will have a different etiquette "code." There is a "cultural etiquette" associated with, say, infantry soldiers as opposed to clerical workers, just as there is a one for marketplace negotiations in the Middle East vs. an American shopping center. Politeness and etiquette are thus very much at the forefront of managing social interactions, and they should play a large role in training and predicting social interaction behaviors across cultures.

A Model of Human-Human Etiquette for Politeness

A seminal body of work in the study of politeness is the cross-cultural studies and resulting model developed by Brown and Levinson (see reference 5). Brown and Levinson noted that people across cultures and languages regularly depart from strictly efficient conversation by using behaviors designed to mitigate or soften direct expressions of desire, intent or command. A simple illustration in English would be: as we settle down to a meal, I ask you "Please pass the salt." The use of "please" is unnecessary for a truthful, relevant or clear expression of my wish and is not required to express my overt intent. Over years of cross-linguistic and -cultural studies, Brown and Levinson collected a huge database of such violations, and developed a model to explain their occurrence. Their primary objective was to explain why individuals chose to use different types and "amounts" of polite behaviors. We believe this can be readily extended to a roughly cognitive model of the perception of threats and threat-mitigating behaviors by any observer, whether involved in the interaction or not, as described below.

Face threats in social interactions

Brown and Levinson assume that social actors are motivated by two important social wants based on the concept of face or, loosely, the "positive social value a person effectively claims for himself" (see reference 6, p.5) 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 are presumed to have:

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

Virtually all interactions between social agents are potentially Face Threatening Acts (FTAs). If I simply speak to you, regardless of the content, I place a demand on your attention that threatens your negative face, for example. This, then, is the reason for the "please" in my request: If I simply state my desire as bald propositional content (e.g.,

"Give me the salt") I would be ambiguous about whether I have the power or right to compel you to give it. You might well take offense. The "please" is thus a "redressive" strategy which mitigates the threat. Furthermore, the expectation that such a strategy be used is an example of an etiquette code that enables interpretations. The etiquette is the "rule" that entitles us to conclude that those who use "please" are striving to be seen as polite; those who do not are not striving toward politeness for various reasons (perhaps they don't believe they need to, perhaps their notions about politeness are different, perhaps they are just rude).

Computing the severity of a face threat

The core of Brown and Levinson's model (see reference 5, p. 76, as well as reference 9) is the claim that the "weightiness" or degree of face threat posed by an act is a function of three factors:

- The relative *power* the hearer (H) has over the speaker (S). Power is an asymmetric relationship. If all other factors are equal, I must use more politeness to a more powerful Hearer to maintain a constant level of threat.
- The *social distance* between H and S. Social distance is roughly the inverse of familiarity and is symmetric. The more familiar my H is, the less politeness I need to use.
- The *imposition* of the raw act itself. Highly imposing acts, requests or topics demand more redress if a constant level of threat is to be maintained.

In our implementation, described below, we add a term representing the relative weight an individual puts on his/her own goals vs. the face goals of others. For want of a better term, we have called this:

- The *character* of the speaker; his/her willingness to value own face more or less than that of others,

Brown and Levinson themselves do not operationalize these parameters (at least in their primary presentation of the model in reference 5); instead, they are offered as qualitative constructs. Work by Cassell and Bickmore[7] and by Johnson and Rizzo[8] has created numerical representations for them to guide, respectively, a simulated real estate agent in making small talk and a pedagogical agent in offering advice and criticism. Our goal has been to develop a general computational formulation of the Brown and Levinson algorithm for use in largely free-flowing conversation and social interactions between humans and agents in a simulation or training environment.

Redressing Face Threats

Since FTAs are disruptive, we use politeness strategies to "redress" or mitigate the degree of face threat imposed by our actions. Brown and Levinson's model claims that the degree of face threat posed by an act must be redressed or balanced by the value of the politeness behaviors used – or

else the social status quo will be disrupted. That is (in our formulation of their model):

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

This says that the weight (W) or value of the face threat in an interaction x (recall that W_x itself is a function of power difference, social distance, imposition and character as described above) is expected to be approximately equaled by the combined redressive Value (V) of the politeness actions (**A**) included in x. If less redress is used than is perceived as necessary, that is if W_x is a lot larger than $V(\mathbf{A}_x)$, then the utterance will be perceived as rude and the hearer may seek alternative interpretations for the behaviors, as will be discussed below. If more politeness behaviors are used than are perceived as necessary, then the utterance will be perceived as "over-polite" or obsequious and, again, ulterior motives may be sought.

Interestingly, across the cultures studied by Brown and Levinson, although outward manifestations obviously differ, the classes of redressive strategies used were identical. They offer an extensive catalogue of universal redressive strategies, organized according to 5 broad strategies. From least to most threatening these are:

1. The least threatening approach is simply not to do the FTA. Some FTAs simply can't be performed without insult, if the disparity in power or familiarity relationship, or the magnitude of the imposition is large enough.
2. Off record FTA strategies are means of doing the act with "plausible deniability" by means of innuendo and hints. An "off record" method of asking for salt might be "I find this food a bit bland."
3. Overt FTAs can still be mitigated by offering redress aimed at either positive or negative face. Brown and Levinson suggest that negative redress will be more effective. Negative redressive strategies focus on H's negative face needs—independence of action and attention. They minimize the impact on H by being direct and simple, and by offering apologies and deference, minimizing the magnitude of the imposition and/or explicitly incurring a debt.
4. Positive redressive strategies target the hearer's positive face. These strategies emphasize common ground between S and H by noticing and attending to H, by invoking in-group identity, by joking and assuming agreement and/or by explicitly offering rewards/promises.
5. Finally, the most threatening way to perform an FTA is "baldly, on record," without any redress.

Brown and Levinson's work doesn't stop at that level, however. For strategies 2-4 they offer a host of well-researched examples from at least three different language/culture groups (English, Tamil and Tzeltal) organized into a structure of mutually supporting and incompatible approaches. We do not have space to present their findings in depth, but we will illustrate a few of the 10 culturally-universal negative redress strategies below:

- 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?"

An "Etiquette Metric"—Believable levels of Politeness

People generally want to accomplish their goals expeditiously-- arguing for minimizing redressive strategies. But they also experience a range of social and personal pressures to not threaten the face of those they interact with (especially those with greater power or shared familiarity)—and this argues for extensive redress use. The balance between these pressures yields the selection of specific strategies in context.

We can therefore assume that the relationship between the amount of face threat and the value of redressive behaviors used determines whether an interaction is perceived as nominal, rude or overly polite. Of course, an individual's perception of the rudeness of an interaction will, in turn, be dependent on that individual's perceptions of the face threat and redress involved—each of which may vary depending on the individual's role in the interaction and on a wide range of cultural factors (as noted in negotiation interations in reference 4). We have expressed this relationship as follows: perceived *politeness* is a function of the perceived *imbalance* (that is, the difference) between the perceived weight of *face threat* in an interaction and the perceived *amount of redress* in that interaction. In turn, perceived face threat is a function of the perceived power difference, social distance, imposition and the character of the interactants. 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. Imbalance will be negative when less redress is used than there was threat—a rude condition.

This model explains a fundamental issue about politeness use—the fact that the same set of politeness behaviors (e.g., the specific utterance, including word choice, verbal and non-verbal behaviors, etc.), used in different contexts, may well be perceived as anything from appropriate to rude or over-polite. It is clear that the same set of redressive acts 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 our implementation will be presented next,

followed by a discussion of how we assess redressive actions and their values.

Algorithm Implementation

In work funded by a DARPA Small Business Innovation Research grant, we have completed initial development of an "Etiquette Engine™" (EE)—an implemented algorithm based on Brown and Levinson's work as described above—and have demonstrated its capability to provide expected politeness assessments both in controlled tests involving project team members and in open surveys involving university students unaware of our model. Our approach and results will be described next.

The EE Algorithm

We have implemented a version of the Brown and Levinson theory to use as a predictive model of an observer's perceived "degree of imbalance" of an interaction, whether among humans or between humans and non-player characters (NPCs)). Actors who do not exhibit the expected degree of polite redress (either by being over- or under-polite) are expected to be seen as either unbelievable or to invite rethinking of what was previously understood about the context. For example, if a private bursts in on his captain and issues a bald directive ("Get your coat on") without any redress, an observer might well assume that the degree of imposition is reduced because the private was charged with giving such instructions, or that the familiarity between them warranted it. Otherwise (and especially in a simulated environment), the observer might simply believe that the private was behaving "unbelievably."

Our algorithm begins by operationalizes the Brown and Levinson model for computing face threat weight (W_x) from knowledge of power difference, social distance, imposition and character. Our implementation uses weights on each term to allow valuing them differently. Although this weighting extends the original model, we suspect it may underlie some systematic cultural differences in politeness usage. Furthermore, a W_x score can be computed based on the perceptions of each separate observer about the context—thus enabling us to represent differences in viewpoint which may arise due to personal or cultural differences. For example, A may have thought he was more powerful than B, while B thought they were equals. Therefore, A may perceive B's behavior as rude (lacking in adequate redress) while B intended it to be polite.

As noted above, the model's assumption is that the perception of an interaction in context will correspond to the balance between face threat weight and the value of redressive actions. We express this as a difference to give us an "incredibility" or "imbalance" metric which also serves as a perceived politeness metric.

To use this metric to evaluate the imbalance between expected and observed levels of politeness, we must operationalize the various parameters. Space does not permit a detailed presentation of our method, but we will summarize the approach below.

Operationalizing EE Terms

To operationalize and quantify the Brown and Levinson model, we first developed scalar values for the *power, social distance* and *imposition* parameters. These scales were initially for basic American culture, but we have since experimented with representing Pashtu culture in a similar approach with reasonable success. We represented the variables (and various parties' perceptions of them) as continuous scalar values ranging from negative to positive 1000. The value of 0 is the "balance point"--a nominal value for each scale, while positive values indicate that the parameter is increased (and contributes to an increasingly "weighty" or potent FTA); negative values indicate that it is decreased (and is building up H rather than threatening him/her). The notion of negative values of power, social distance, etc. having the potential to reduce a "threat" to the point where it becomes of benefit to the Hearer is not explicit in Brown and Levinson's work, but it seems an intuitive addition to us. For Power Difference of H over S, for example, a value of 0 means that the power of H and S are equal, that they are (exact) peers—and that there should be no face threat inherent in their interaction stemming from power differences. Values greater than 0 indicate that H has increasingly greater power relative to S and, therefore, that face threat increases whenever S addresses H. Similar scales were developed for social distance and imposition. The character term was represented as a simple value added or subtracted from the FTA sum.

While it should be noted that these scales were developed for American cultural norms, our initial work[9] indicates that Pashtu experts had no difficulties mapping them and their anchor points into Pashtu culture, where they work similarly. The claim is certainly not that, say, parents and children have the same degree of social distance in American culture as in Pashtu culture, but rather than whatever relationship corresponds to the 100 point in both cultures will require the same degree of redress (however that is defined in the culture). While substantial work remains to be done to formalize scales within and across cultures, we believe our approach—and the validation exercises we report below—show promise for cultural universalism, as Brown and Levinson have claimed for their model.

Next, we developed numerical valuations for various redressive behaviors based on the Brown and Levinson guidelines described above. Ranges of values for the broad classes of strategies were defined as follows: the use of an individual *positive redressive strategy* (see reference 5, p. 102, Figure 3, for a list) proview from 1 to 40 "units" of redress, while the use of an individual *negative redressive strategy* (see reference 5, p. 131, Figure 4, for a list) will

provide from 20 to 60 units of redress. We have largely not addressed off record strategies as yet, but Brown and Levinson suggest that they would be generally more potent than either positive or negative strategies. Within the range defined above, a specific score was assigned to individual instances of redress in the category, as illustrated below.

The effects of multiple redressive strategies were scored as simply additive. This is the primary reason that the scales for terms contributing to face threat weight (power, social distance, etc.) and those contributing to the redressive value of an action are so different—because, in practice, any given interaction contains multiple redressive acts and their individual values combine to produce the summed $V(A_x)$ value which is then balanced against Face Threat Weight (W_x). We understand that additive summing is a simplification, and that the efficacy of added redressive behaviors inevitably falls off, eventually becoming simply irritating, but we have not yet attempted to model such subtleties (nor have we needed to). This means that the value of a set of redressive actions contained in interaction is simply the sum of their individual values.

Using and Evaluating the Algorithm

Our approach was tested in a series of sample social interaction vignettes crafted to represent (according to our American cultural intuitions) either normal/balanced politeness, unbelievable over-politeness or unbelievable rudeness. Our goal was to determine if the equation and scoring techniques would track our intuitions. The level of face threat and redress were varied over this set of vignettes so that high face threat situations were paired with high levels of redress (roughly balanced) as well as low levels of redress (highly imbalanced and rude). Similarly, very low levels of face threat were paired with very high levels of redress (over-polite) and with low levels of re-

dress (balanced). Examples of two such vignettes are illustrated in Table 1.

Evaluation 1—Trained Rater Correlations

Each of our eight vignettes was then assessed using the operational scoring tables we had created for situational context parameters (power, social distance, etc.) and for values of the individual redressive actions. For example, for the first vignette the imbalance evaluation proceeded as follows:

- The corporal (as S) has lower power than the mayor; their "power difference" is fairly large—probably larger than an anchor point of 100 we used (the power that a parent has over a 12 to 13-year-old teenager), yet less than the anchor point of 1000 (the power that a parent has over a small child or infant). We scored this as *power difference*= 300.
- There is no particular familiarity between the two individuals, but social distance is not extreme either. They are from slightly different "cultures" (military vs. civilian) and show no evidence of prior relationship, but they are engaged in a common endeavor. The social distance between them is probably only slightly higher than 0. Thus, we decided that *social distance* = 3.
- The imposition of this request could be large. To ask after the location of one's wife so as to pick her up is comparatively threatening, though the fact that this is in the mayor's service should mitigate this (as the corporal reminds him). The "raw" imposition, however, is only a short answer, characteristic of our level 10, so we assigned *imposition* = 10.
- Since we have provided no reason to believe that the character of the corporal is anything other than nominal, we assume that *character* = 0.

This gives us a value of the FTA weight (from the observer's perspective described above) as: $3 + 300 + 10 + 0 = 313$.

For the value of the redress applied we scored the set of redressive actions in Table 2 on the basis of whether they were positive or negative redress, and the relative value we thought they had within the ranges we had specified above.

Thus, the imbalance score for this vignette, as calculated by our equation, would be: $295 – 313 = -18$. Since this vignette was intended to convey both high face threat and high redress and, thus, to be roughly balanced, this score seems about right, falling very near zero. For the second vignette, by contrast, we have a high degree of face threat with virtually no redressive actions. This is unexpected and should be perceived as very rude. This scenario should have a score much less than 0 on our imbalance metric—indicating substantial unredressed threat. This vignette had the same face threat attributes as Vignette 1 but was scored as having only 40 points of redress, thus giving an *imbalance* of $40-313 = -273$. This is strongly

Table 2. Scoring of Redressive Behaviors used in Vignette 1.

Action and Interpretation	Score
1. The soldier waits until the mayor is finished and invites him to speak. This seems to be a very explicit form of negative politeness (putting the other's interests first) and, especially in this instance where the H was not actively engaged in another conversation, seems very potent.	60
2. The soldier takes off his hat. This is a sign of deference, which is in turn a fairly potent negative politeness strategy.	50
3. The soldier apologizes for interrupting. This is also a negative politeness strategy, though arguably a less potent one (though that may be highly mitigated by facial expressions and body language).	30
4. The soldier uses an honorific. Moderately potent negative politeness strategy.	40
5. The soldier poses the FTA as a question. Common negative politeness strategy.	20
6. The soldier offers an explanation/reason for needing the information. Positive politeness strategy.	35
7. The soldier appeals to the Mayor's (H's) interests. Positive politeness strategy Powerful in this context.	30
8. The soldier is hesitant and skeptical about compliance. A common but reasonably potent negative politeness strategy.	30
TOTAL	295

negative—as we expected for an interaction intended to be perceived as rude.

An evaluation similar to that above was carried out for eight vignettes and the quantitative algorithm tracked predictions for rude, polite or nominal perceived etiquette levels very closely. As shown in Figure 1, all vignettes that were intended as "nominal" (that is, using about the amount of redress expected in American culture for the amount of face threat incurred) scored within +/- 100 points of zero. All vignettes that were expected to be over-polite scored well higher than 100 points; while all that were expected to be overly rude scored substantially less than -100 points.

While the above example was based on one individual's assessments (Dr. Miller's), we have since replicated this with two other raters following a brief training session.

Each rater was a member of the project and was generally familiar with the Brown and Levinson model, but not with the specific scores that Dr. Miller had produced. Each rater scored the vignettes and the three scoring sets were then statistically compared. The top-level imbalance metric showed a Robinson's A correlation of .931 among the three raters across the 8 vignettes, and the two major sub-factors (Face Threat Weight and Redress Value) showed correlations of .950 and .863 respectively. These values are all well above traditional thresholds of .7 or .8 for multiple judge rating correlations. Thus, this study supports the claim that we have identified a reliable method of scoring the degree of politeness in social discourse—at least in American cultural settings.

Evaluation 2--Untrained Rater Correlations
While the above evaluation used trained raters, we subse-

A priori expectation based on vignette design
Bal. = Balanced, Imbalance scores near 0
Neg. = Negative, Imbalance scores << 0, Rude
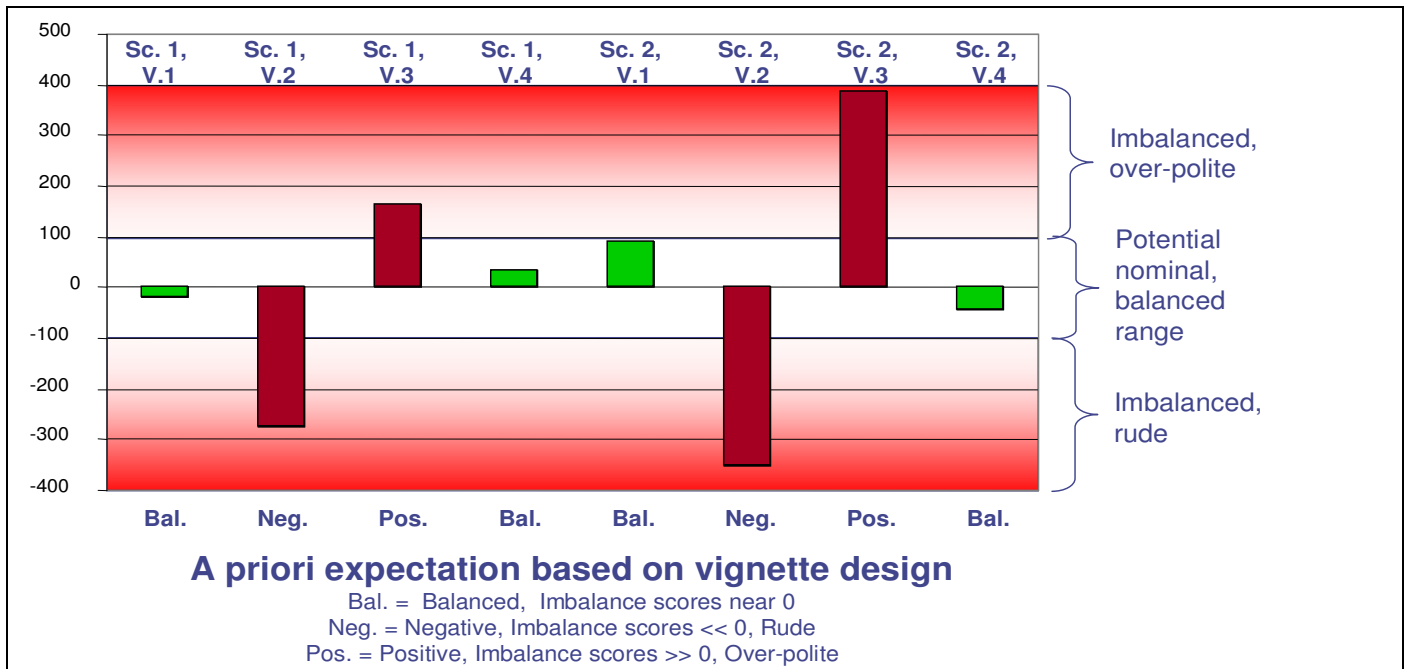Pos. = Positive, Imbalance scores >> 0, Over-polite

Figure 1. Figure 1. Imbalance scores for all eight vignettes.

quently conducted an experiment wherein 22 American college students (respondents to flyers and unaware of our theory or model), also rated the same 8 vignettes. Participants reviewed a "backstory" describing the participants in the vignette and then answered a series of questions about their relationships. They then read the specific, verbal interaction (i.e., utterance), and answered (using Likert scales) questions about their perceptions of the actors, their relationships, the degree of politeness used and whether or not they regarded the interaction as normal, rude or overly polite.

Correlations between participants' ratings of the model parameters and our own remained very high. Comparing the project team's mean rating with the untrained participants' mean rating using Pearson's coefficient showed correlations of .867 for power ratings, .881 for social distance, .766 for imposition and .892 for overall imbalance/politeness ratings. The overall politeness correlation was significant at the $p<.01$ level (two-tailed). Note that while the number of participants involved may have been limited, that merely reduces the probability of finding a significant finding, making this a conservative test.

Participants were also asked whether they changed beliefs about the power and social distance values after they had seen the utterances used. Since some vignettes used utterances expected to be nominal while others used either unexpectedly high or low politeness levels, we hypothesized that if our model were correct, then more participants would be willing to change their ratings after seeing the vignettes with "off-nominal" politeness rather than those with nominal politeness.

As shown in Figure 2, a higher percentage of participants reported willingness to change their ratings in response to the off-nominal vignettes, though this effect was more pronounced for Power Difference (P) than for Social Distance (D) ratings. A paired-samples t-test on the mean values for the four nominal vs. four off-nominal vignettes showed that significantly more participants wanted to change their estimate of Power Difference after reading/viewing the Off-nominal behaviors than the Nominal ones (t=-4.85, df=3, p<.05). A similar test for the Social Distance parameter was not significant (t=-1.186, df=3, p>.2) but trended in the same direction.

In general, these data support our interpretation and implementation of Brown and Levinson's model and the claims that unexpected (i.e., off-nominal) amounts of redress prompt people to reinterpret their beliefs about context—specifically, their beliefs about the power and social distance parameters. In this study, participants proved more willing to review their perceptions of power relationships than social distance. This may be a function of the marked power relationships in the vignettes (involving, as they did, soldiers and civilians) or it may reflect a more general tendency among Americans to seek explanations for politeness variations in power dimensions first. Resolving this question must await future research.

Implementation Results and Payoffs

The use of Brown and Levinson's model in a module for reasoning about social interaction behaviors ensures universal reasoning about and scoring of abstract politeness "moves". Any such module will need to be equipped with culture-specific knowledge bases, however, to enable reasoning from the observable behaviors in a culture (e.g., pursed lips or a rigid hand-to-eyebrow salute) to the abstract etiquette "moves" (and therefore, politeness implications) over which the model's parameters are scored. This has the practical implication that the general social interaction reasoning of an automated system can be effectively modularized, and, thus, large savings in simulation code development can be realized. Furthermore, basic game storylines or training modules and even specific characters can be easily transposed from one cultural milieu to another—enabling the village priest who the player had to interact with to get intelligence information in a Kosovo training game to take on the culture-specific behaviors and reactions (though not, without further work not reflected here, the appearance) of an imam in an Iraqi training game—simply by loading a new module of cultural knowledge. In each case, new knowledge bases of culture-specific politeness behaviors would need to be developed (and, of course, checked for accuracy) for each new simulation, but the core game story-
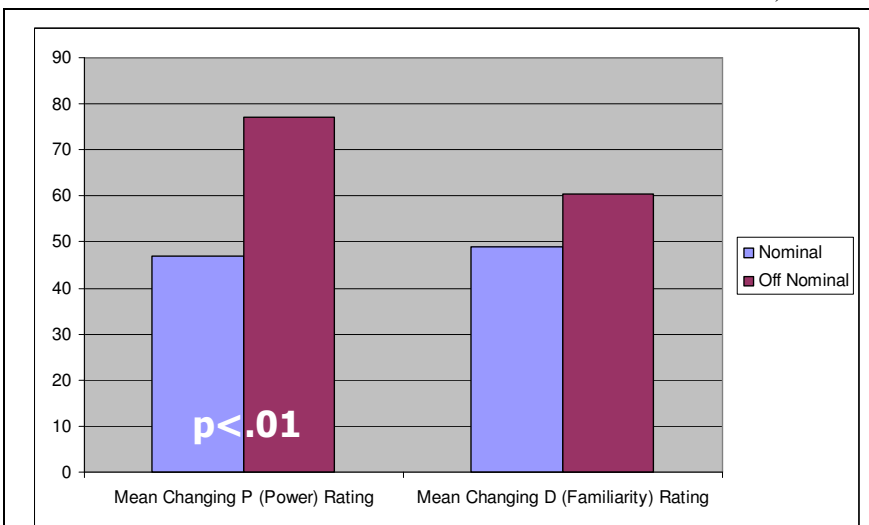

Figure 2. Percentage of participants wanting to change their estimates of Power Difference (P) and Social Distance (D) after hearing the actors' utterances.

line(s) and character roles, general actions, motivations, capabilities, etc., could remain unchanged.

In work reported elsewhere (see reference 9) we have detailed our success in integrating this implemented algorithm into a language training game (the Tactical Language Training System developed by USC's CARTE Labs[10]). This work has demonstrated our algorithm's ability to inform both the perceptions and the reactions of simulated characters—and to do so with less software development time than traditional scripting approaches. Moreover, it has also demonstrated the ability of our approach to provide at least reasonable knowledge and use of politeness levels in a culture different from American English, (specifically, the Pashtu language spoken along the Afghanistan/Pakistan border).

Perhaps most interestingly, this approach of parsing the perception of politeness into subcomponents (power, social distance, imposition, character and redressive values) opens 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 to representing social interaction behaviors in games or simulations, where each subsequent interaction must be developed from scratch with essentially the same cost in labor as the one before it, there is reason to believe that our approach scales geometrically (cf. reference 9).

By explicitly representing the knowledge to compute how one observer perceives an utterance from one speaker, 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 power and social distance values for the new pair. Similarly, knowing the value of adding a specific honorific to an utterance, means that we now know the relative worth of using that positive redress strategy for any utterance in which it makes sense.

In fact, during our work with the TLTS system described in reference 10, 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 specific communication uttered by a specific 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. In short, it becomes easy to recombine previously scored elements to generate the product set of possible communicative acts each of which will then be automatically scored on the basis of those previously scored elements.

A final advantage of our model of social interaction politeness and etiquette is that it allows explicit and quantifiable predictions about people's perceptions of utterances in context. We are currently using this aspect of our model to inform a series of human interaction experiments exploring human reactions to variations in the politeness with which a directive is delivered. "Directives" here are meant in the speech act sense of any utterance whose intent is to direct the hearer to perform some act. Thus, "directives" covers everything from beseeching requests to direct and forceful commands and advice or warnings from a decision aid or human trainer. In short, our experiments are exploring whether, for example, saying "please" has an effect on whether and how a human hearer chooses to comply with the directive. We suspect that using directives that are perceived as more polite will increase the overall likelihood that the hearer will comply with the directive as well as result in increased trust for the directive giver and in the perception of reduced workload. On the other hand, reaction time may be slower since increased politeness is associated with reduced urgency according to our model. We will also be looking at whether these trends differ cross culturally and at whether the directive giver's status as human or machine makes any predictable difference in a human hearer's directive compliance behavior.

Conclusions

While etiquette and politeness are far from the only aspects of culture that should be modeled in computational tools and approaches, they are a pervasive aspect of virtually all interactions that matter in cross-cultural collaboration and interaction. In this work, thanks to the basic model developed by Brown and Levinson, etiquette and politeness are proving amenable to quantitative, computational modeling. Furthermore, the resulting models are providing predictions which correlate well with both trained and naïve users of our modeling framework, at least for American cultural sensibilities.

An ability to score the politeness of the interaction behaviors of actors in a cultural and social context is important because it allows for quantitative reasoning (and, ultimately, for machine aids and predictions) about how social interaction "moves" will be perceived. Thus, this model holds the potential to equip a wide range of simulated characters representing different cultures with the culture-specific ability to perceive and evaluate human behaviors and then to react accordingly. Similarly, it can equip machine systems with the ability to determine appropriate behaviors in order to further ends such as trust, usability and user acceptance. Finally, it may also assist in managing or even filtering and translating the "politeness" in conversations between members of different cultures.

Acknowledgements

Biographies

Dr. Christopher A. Miller is Chief Scientist at Smart Information Flow Technologies. His interests include human automation integration, human performance modeling, and politeness and etiquette across cultures and in both human-human and human-machine interaction. Dr. Miller's Ph.D. was received from the Committee on Cognition and Communication in the Psychology Department at the University of Chicago. He is a member of the Human Factors and Ergonomics Society and of the Association for Computing Machinery. He can be reached at cmiller@sift.info or 211 First St. N., Suite 300, Minneapolis, MN 55108, USA.

Ms. Peggy Wu is a Research Scientist at Smart Information Flow Technologies. Her research interests include how culture influences and constrains HCI designs, and understanding how and when modeling the user can improve interfaces. Her roles include acting Principal Investigator, technical lead, designer, and programmer. Prior to joining SIFT, Ms. Wu was a Research Associate at Honeywell's Research Laboratories, and was a human factors and user interface designer for Roxio International. She received her B.Sc. in Computer Engineering from the University of Toronto. She can be reached at 211 N 1st St. #300, Minneapolis, MN 55401 PWu@SIFT.info.

Mr. Harry B. Funk is a Senior Research Scientist and Vice President of R&D at Smart Information Flow Technologies. His research interests include human-automation integration, human-computer interfaces, human performance modeling, and neurophysiological approaches to human-machine interaction. Mr. Funk has a BA in Psychology from the University of California at Berkeley, and a BS in computer science from the University of Minnesota. He can be reached at hfunk@sift.info or 211 First St. N., Suite 300, Minneapolis, MN 55108, USA.

References

1. D. Dennet, *The Intentional Stance*, MIT Press, 1989.
2. G. Hofstede, *Culture's Consequences*, 2nd Ed. Sage Publications, 2001.
3. R. E. Nisbett, *The geography of thought: How Asians and Westerners think differently…and why.* Free Press, 2003.
4. Gelfand, M. J. & Brett, J. M. (Eds.) *Handbook of negotiation: Theoretical advances and cultural perspectives.* Stanford University Press, 2004.
5. P. Brown and S. Levinson, *Politeness: Some Universals in Language Usage,* Cambridge University Press, 1987.
6. E. Goffman, *Interaction Ritual: Essays on Face to Face Behavior.* Garden City; 1967, pp. 5-9.
7. J. Cassell and T. Bickmore. "Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents," *User Modeling and User-Adapted Interaction*, vol. 13, no. 1, 2003, pp. 89-132.
8. L. Johnson, and P. Rizzo, P., "Politeness in Tutoring Dialogs: 'Run the Factory, That's What I'd Do',". *Intelligent Tutoring Systems*, 2004, pp. 67-76.
9. C. Miller, P. Wu, H. Funk, P. Wilson, and L. Johnson, "A Computational Approach to Etiquette and Politeness: An 'Etiquette Engine™' for Cultural Interaction Training," In *Proceedings of BRIMS 2007.* SISO, paper #045.
10. Johnson, H. Vilhjalmsson, and M. Marsella, "Serious Games for Language Learning: How Much Game, How Much AI?," In *Proceedings of the 12th International Conference on Artificial Intelligence in Education* . IOS Press, pp. 306-313.