A Computational Approach to Etiquette: Operationalizing Brown and Levinson's Politeness Model

Christopher A. Miller, Peggy Wu, and Harry B. Funk,
Smart Information Flow Technologies

Politeness behaviors are how we manage critical aspects of social interaction. A computational model of culture-specific politeness will support studying, simulating, and aiding such interactions.

Understanding *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—has clear value in many situations. For example, team communications inevitably require communication about power, trust, familiarity, group membership, urgency, importance, and so on—all in addition to explicit task-related content. Cultural differences in how team members convey these social-interaction variables can exacerbate communication difficulties in multicultural teams.

Much basic research has focused on identifying cultural patterns[2] and on how cultural factors affect cognitive processes.[3] However, few of the resulting models directly link cultural factors to human behavior or are readily amenable to computational modeling at a fine 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. To address this need, we're developing a model of observable human behaviors—specifically communication behaviors relating to etiquette and politeness, which are relevant to human performance, attitudes, and broader decision making. We've been leveraging a qualitative model from research in sociology and linguistics[5] for two reasons. First, it accounts for many of the effects of social-interaction behaviors on team interactions. Second, it provides a culturally universal, yet instantiatable, basis on which to develop culture-specific models. Such a universal model will be useful in evaluating, adapting, interpreting, and generating social-interaction behaviors in both human-human and human-machine interactions. We report here on our efforts to develop this quantitative, computational model of politeness interactions and to apply and evaluate it in various different contexts.

Politeness for social interactions

The terms "etiquette" and "politeness" likely 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.[1,6] Politeness is one means by which we convey, interpret, maintain, and alter social relationships.

Etiquette is the code by which we signal politeness.

1541-1672/08/$25.00 © 2008 IEEE
Published by the IEEE Computer Society

It employs verbal, physical, gestural, and even more primitive modes of interaction. For example, a person can express deference by posture, quiet speech, and explicit markers such as titles and honorifics. Specific behaviors vary across cultures, because each culture has its own etiquette code. The cultural etiquette associated with, say, infantry soldiers differs from that of clerical workers, just as the one for marketplace negotiations in the Middle East differs from that for 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

A seminal body of work in the study of politeness is the cross-cultural studies and resulting model that Penelope Brown and Stephen Levinson developed.[5] They 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.

Consider the following simple illustration. As we settle down to a meal, I say to you, "Please pass the salt." The use of "please" is unnecessary for a truthful, relevant, or clear expression of my wish and isn't required to express my overt intent. Over years of cross-linguistic and cross-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 different types and amounts of polite behaviors. We believe their empirically based model can be readily extended to a roughly cognitive model of the perception of threats and threat-mitigating behaviors by any observer, whether or not the observer is involved in the interaction.

In the rest of this article, we will use the terms "speaker" and "hearer" to refer to parties in an interaction. The speaker is the one who makes an utterance, and the hearer is either the speaker's target or a third-party observer. As we'll describe, perceptions of politeness and its constituent elements might differ across these various roles and individuals, but it's a part of the Brown and Levinson model that politeness itself should function similarly for all parties.

Face threats in social interactions

Brown and Levinson assume that two important social wants motivate social actors on the basis of the *face* concept—loosely, as Erving Goffman put it, the "positive social value a person effectively claims for himself."[6] Face can be saved or lost, and it can be threatened or conserved in interactions. Brown and Levinson refine the concept of face into two specific subgoals that all social actors are presumed to have:

- *Positive face*—the desire to be held in high esteem, to have your actions and opinions valued, to be approved of by others, and so on.
- *Negative face*—the desire for autonomy, to have your way, to direct your attention where and when you desire, and so on.

> Politeness and etiquette play a large role in training and predicting social-interaction behaviors and perceptions across cultures.

Virtually all interactions between social agents are potentially face-threatening acts (FTAs). For example, if I simply speak to you, regardless of the content, I place a demand on your attention that threatens your negative face. This, then, is the reason for the "please" in my request: If I simply state my desire as bald propositional content ("Give me the salt"), I would be ambiguous about whether I have the power or right to compel you to comply. You might well take offense. The "please" is thus a redressive strategy that mitigates the threat.

Furthermore, the expectation that I (or any social actor) would use such a strategy is an example of an etiquette code that enables interpretations. The etiquette is the convention entitling us to conclude that those who use "please" are striving to be seen as polite. Those who don't use it aren't striving toward politeness for various reasons (perhaps they don't believe they need to, their notions about politeness are different, or they're just rude).

Computing a face threat's severity

The core of Brown and Levinson's model is the claim that the weightiness or degree of face threat posed by an act is a function of three factors.[5,7] The first is the *relative power* the hearer has over the speaker. Power is an asymmetric relationship. If all other factors are equal, I must use more politeness toward a more powerful hearer to maintain a constant level of threat.

The second factor is the *social distance* between the hearer and speaker. Social distance is a symmetric relationship and roughly the inverse of familiarity. The more familiar I am with my hearer, the less politeness I need to use.

The third factor is the *imposition* of the raw act. Highly imposing acts, requests, or topics demand more redress to mitigate their increased threat level.

In our implementation, we add a term representing the relative weight an individual puts on his or her own goals versus others' face goals. For want of a better term, we call this the speaker's *character*. This refers to the speaker's willingness to value his or her own face more or less than that of others.

Brown and Levinson don't operationalize these parameters (at least in their primary presentation of the model[5]); they offer the parameters as qualitative constructs. Justine Cassell and Timothy Bickmore and Lewis Johnson and Paola Rizzo have created numerical representations of these parameters to guide, respectively, a simulated real estate agent in making small talk[8] and a pedagogical agent in offering advice and criticism.[9] 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

Because FTAs are disruptive, we use politeness strategies to redress or mitigate the degree of face threat our actions impose. Brown and Levinson's model claims that the degree of face threat an act poses must be redressed or balanced by the value of the politeness behaviors the actor uses;

otherwise, the social status quo is disrupted. That is, in our formulation of their model,

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

The face threat's weight (W) or value in an interaction x (recall that W_x itself is a function of power difference, social distance, imposition, and character) is expected to be approximately equal to the combined redressive value (V) of the set of politeness actions ($\boldsymbol{A}$) included in x. If the speaker uses less redress than the hearer perceives as necessary—that is, if W_x is much larger than $V(\boldsymbol{A}_x)$—then the hearer will perceive the utterance as rude and might seek alternative interpretations for the behaviors, as we discuss in a moment. If the value of the politeness behaviors are much more than is perceived as necessary, then the hearer will perceive the utterance as overpolite or obsequious and might again suspect ulterior motives or alternate interpretations.

Interestingly, across the cultures that Brown and Levinson studied, although outward manifestations obviously differ, speakers used identical classes of redressive strategies. Brown and Levinson offer an extensive catalog of universal classes of redressive strategies, organized according to five broad strategies (listed from least to most threatening).

The first type of strategy is to simply not do the FTA, which we might call a *null FTA*. Some FTAs simply can't be performed without insult, if the disparity in power or familiarity relationship or the imposition is large enough.

The second strategy is an *off-record FTA,* which performs the act with "plausible deniability" through innuendo and hints. An off-record way to ask for salt might be, "I find this food a bit bland."

The third strategy is with *negative redress.* A speaker can still mitigate overt FTAs by offering redress aimed at either positive or negative face. Brown and Levinson suggest that negative redress is more effective. Such strategies focus on the target's negative face needs—independence of action and attention. They minimize the impact on the target by being direct and simple and by offering apologies and deference, thus minimizing the imposition's magnitude and explicitly incurring a debt. A way to ask for the salt using several negative redress strategies would be "Excuse me, sir, but could you pass me the salt?"

The fourth strategy is using *positive redress,* which aims at the target's positive face. These strategies emphasize common ground between the speaker and hearer by noticing and attending to the target, invoking in-group identity, joking and assuming agreement, and explicitly offering rewards or promises. A way to ask for the salt using several positive redress strategies would be "Hey, buddy! You're gonna pass me that salt, aren't ya? I'd appreciate it."

The final type is *no redress.* The most threatening way to perform an FTA is "baldly, on record," without any redress: for example, "Give me the salt."

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

- 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?"

> Perceived politeness is a function of the imbalance between the perceived weight of face threat and the perceived amount of redress in that interaction.

An etiquette metric: Believable levels of politeness

People generally want to accomplish their goals expeditiously—an argument for minimizing redressive strategies. But they also experience a range of social and personal pressures to not threaten the face of those with whom they interact (especially those with greater power or shared familiarity)—an argument for using extensive redress. 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 the speaker's redressive behaviors determines whether the hearer perceives an interaction as nominal, rude, or overly polite. Of course, an individual's perception of an interaction's rudeness will, in turn, depend on that individual's perceptions of the face threat and redress involved—each of which might vary depending on the individual's role in the interaction and a wide range of cultural factors.[4]

We've 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 character of the actors. The imbalance will be positive when the speaker uses more redressive politeness behaviors than the face threat warrants—corresponding to the overly polite or obsequious behavior. To be more precise, a hearer will perceive the imbalance to be positive when the hearer perceives the speaker as using redressive behaviors whose combined value the hearer perceives to be greater than he or she perceives the face threat to warrant. The imbalance will be negative when the hearer perceives the speaker as using less redress than the hearer perceives the face threat to warrant—a rudeness condition.

This model explains a fundamental issue about politeness use: the same set of politeness behaviors (including the specific utterance, its word choice, and verbal and nonverbal behaviors) used in different contexts might well be perceived as anything from appropriate to rude or overly polite. For example, the same set of redressive acts might be too much, too little, or just right, depending on the present face threat's value. Of course, this leaves open the question of how to determine face threat, which is the focus of our computational implementation.

Algorithm implementation

In research funded by a DARPA Small Business Innovation Research grant, we completed initial development of an Etiquette Engine (EE)—an implemented algorithm based on Brown and Levinson's research. We demonstrated the EE's ability to assess expected politeness both in controlled tests involving project team members and in open surveys involving university students unaware of our model.

The EE algorithm

We designed and implemented a computational version of the Brown and Levinson theory to serve as a predictive model of an observer's perceived "degree of imbalance" of an interaction, whether among humans or among humans and nonplayer characters (NPCs). Our algorithm predicts the level of politeness that a hearer will expect. Furthermore, it predicts that actors who don't exhibit the expected degree of polite redress (either by being overpolite or rude) will either appear unbelievable or invite the hearer to rethink what was previously understood about the context so as to rebalance threat and redress levels.

For example, a private bursts in on his captain and issues a bald directive ("Get your coat on") without any redress. An observer might assume a reduced degree of imposition because the private was charged with giving such instructions or because 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 operationalizing the Brown and Levinson model for computing face threat weight (W_x) from knowledge of power difference, social distance, imposition, and character. Our implementation weights each term to allow valuing them differently. Although this weighting extends Brown and Levinson's original model, we suspect it might underlie some systematic cultural differences in politeness usage. Furthermore, we can compute a W_x score on the basis of each observer's perceptions about the context, thus letting us represent differences in viewpoint that might arise because of personal or cultural differences. For example, A might have thought he was more powerful than B, whereas B thought they were equals. So, A might perceive B's behavior as rude (lacking in adequate redress), whereas B intended it as polite.

As we've noted, the model assumes that the perception of an interaction in context will correspond to the balance between face threat weight and redressive actions' value. 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 politeness, we must operationalize the various parameters. We don't have space to present our method in detail, so we'll summarize it.

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've since experimented with representing Pashtu culture similarly with success. We represented the variables (and various parties' perceptions of them) as continuous scalar values ranging from –1,000 to 1,000. The value of 0 is the balance point—a nominal value for each scale. Positive values indicate that the parameter has increased (and contributes to an increasingly weighty or potent FTA); negative values indicate that it has decreased (and is building up, rather than threatening, the hearer).

We demonstrated the Etiquette Engine's ability to assess expected politeness both in controlled tests involving project team members and in open surveys.

The notion of negative values of power, social distance, and so on having the potential to reduce a threat to the point that it becomes of benefit to the hearer isn't explicit in Brown and Levinson's work, but it seems an intuitive addition to us. For a power difference of the hearer over the speaker, for example, a value of 0 means that the power of the hearer and speaker are equal—that they are (exact) peers—and that no face threat stemming from power differences should be inherent in their interaction. Values greater than 0 indicate that the hearer has increasingly greater power relative to the speaker and, therefore, that the face threat increases whenever the speaker addresses the hearer. We developed similar scales for social distance and imposition and represented the character term as a simple value added or subtracted from the FTA sum.

Although we developed these scales for American cultural norms, our initial work indicates that Pashtu experts had no difficulties mapping them and their anchor points into Pashtu culture, where the scales work similarly.[7] The claim is certainly not that, say, parents and children have the same degree of social distance in American culture as in Pashtu culture. Rather, whatever relationship corresponds to a value of 100 in both cultures will require the same degree of redress (however that is defined in the culture). Although substantial work remains to formalize scales in and across cultures, we believe our approach—and the validation exercises we report on in a moment—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 Brown and Levinson's guidelines. We defined value ranges for the broad classes of strategies as follows: using an individual positive redressive strategy provides 1 to 40 units of redress; using an individual negative redressive strategy provides 20 to 60 units of redress.[5] Within these ranges, we assigned a specific score to individual instances of redress in the category, as we illustrate later. (We largely haven't addressed off-record strategies yet, but Brown and Levinson suggest that they would be generally more potent than either positive or negative strategies.)

We scored the effects of multiple redressive strategies as simply additive. This is the primary reason that the scales for terms contributing to face threat weight (power, social distance, and so on) and those contributing to an action's redressive value differ so greatly. In practice, any given interaction contains multiple redressive acts, and their individual values combine to produce the summed $V(\boldsymbol{A}_x)$ value. We understand that additive summing is a simplification and

Vignette 1: High face threat, high redress, balanced
A low-ranking soldier (a corporal, as his uniform insignia indicates) walks into the mayor's office, and the mayor motions him quickly to a seat. The soldier takes off his hat and sits down, waiting while the mayor continues to write something. The mayor finishes writing, puts down his pen, and looks up at the soldier expectantly. The soldier says, "I'm sorry to interrupt your work, Mayor Fredrickson, but my name is Corporal Jones, and I've been put in charge of your escort to the event tonight. I was wondering if it would be possible for you to let me know where I can meet your wife, so that I can get her there on time?"

Vignette 2: High face threat, low redress, imbalanced
This interaction follows Vignette 1, except the soldier acts and speaks differently. Here, he interrupts the mayor while he is speaking, perhaps by putting a hand on his shoulder, and says loudly, "Tell me where I can meet your wife."

Figure 1. Sample politeness vignettes. To test our approach, we crafted social-interaction vignettes, such as these, designed to represent varying levels of politeness imbalance according to American cultural intuitions.

Table 1. Scoring redressive behaviors used in Vignette 1 in Figure 1.

Action and interpretation	Score
1. The soldier waits until the mayor is finished and invites him to speak. This seems a very explicit form of negative politeness (putting the other's interests first) and, especially in this instance where the hearer wasn't 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, although arguably less potent. However, it can be highly mitigated by facial expressions and body language.	30
4. The soldier uses an honorific, which is a moderately potent negative politeness strategy.	40
5. The soldier poses the face threat action as a question, which is a common (and somewhat weak) negative politeness strategy.	20
6. The soldier offers an explanation or reason for needing the information. This is a reasonably potent positive politeness strategy.	35
7. The soldier appeals to the mayor's (the hearer's) interests. This is a positive politeness strategy that's powerful in this context.	30
8. The soldier is hesitant and skeptical about compliance. This is a common but reasonably potent negative politeness strategy.	30
Total	**295**

that the efficacy of added redressive behaviors inevitably falls off, eventually just becoming irritating. However, we haven't 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. We then subtract the computed W_x from the summed $V(\boldsymbol{A}_x)$ value to provide an imbalance metric, which conveys varying degrees of rudeness or overpoliteness by its magnitude and size, as we described earlier.

Using and evaluating the algorithm

We tested our approach in a series of sample social-interaction vignettes crafted to represent (according to our American cultural intuitions) either normal, balanced politeness; unbelievable overpoliteness; or unbelievable rudeness. Our goal was to determine whether the equation and scoring techniques would track our intuitions. The level of face threat and redress varied over this set of vignettes so that high face threat situations paired with high redress levels (roughly balanced) as well as low redress levels (highly imbalanced and rude). Similarly, very low face threat levels paired with very high redress levels (overpolite) and with low redress levels (balanced). Figure 1 illustrates two such vignettes.

Evaluation 1: Trained rater correlations

Next, we assessed each of our eight vignettes using the operational scoring tables we created for situational context parameters (power, social distance, and so on) and for the individual redressive actions' values. For example, for the first vignette, the imbalance evaluation proceeded as follows.

The corporal (as the speaker) has lower power than the mayor. Their power difference is fairly large—probably larger than the anchor point of 100 we used (a parent's power over a 12- to 13-year-old teenager) but less than the anchor point of 1,000 (a parent's power over a small child or infant). We scored this as power difference = 300.

The two individuals have no particular familiarity, but social distance isn't extreme either. They are from slightly different cultures (military and civilian) and show no evidence of a prior relationship, but they're engaged in a common endeavor. The social distance between them is probably only slightly higher than 0. So, we decided that social distance = 3.

The request's imposition could be large. To ask after the location of one's wife so as to pick her up is comparatively threatening, although the fact that this task 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 determined that imposition = 10.

Because we've provided no reason to believe that the corporal's character is anything other than nominal, we assume that character = 0.

These scores result in an FTA weight (from the observer's perspective, as we've described) of 3 + 300 + 10 + 0 = 313.

For the redress value of the corporal's utterance and behavior, we scored the set of redressive actions in Table 1 on the basis of whether they were positive or negative strategies and the relative value we thought they had within the ranges we specified in the

"Operationalizing EE terms" section.

So, this vignette's imbalance score, calculated through our equation, would be 295 – 313 = –18. Because we designed this vignette 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, in 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 lower than 0 on our imbalance metric, indicating substantial unredressed threat. Vignette 2 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 imbalance is strongly negative, meeting our expectations for an interaction intended to be perceived as rude.

We carried out a similar evaluation for eight vignettes, and the quantitative algorithm closely tracked predictions for rude, polite, or nominal perceived etiquette levels. As Figure 2 shows, 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 we expected to be overpolite scored well higher than 100 points, whereas all those we expected to be overly rude scored substantially fewer than –100 points.

Whereas this example was based on one individual's assessments (Miller's), we have since replicated it with two other raters following a brief training session. Each rater was a member of the project and generally familiar with the Brown and Levinson model but unfamiliar with the specific scores that Miller produced. Each rater scored the vignettes, then we statistically compared the three scoring sets. The top-level imbalance metric showed a Robinson's A correlation of .931 among the three raters across the eight vignettes, and the two major subfactors (face threat weight and redress value) showed correlations of .950 and .863, respectively. These values are all well above the traditional thresholds of .7 or .8 for multiple-judge rating correlations. So, this study supports the claim that we've identified a reliable method of scoring the degree of politeness in social discourse—at least in American cultural settings.

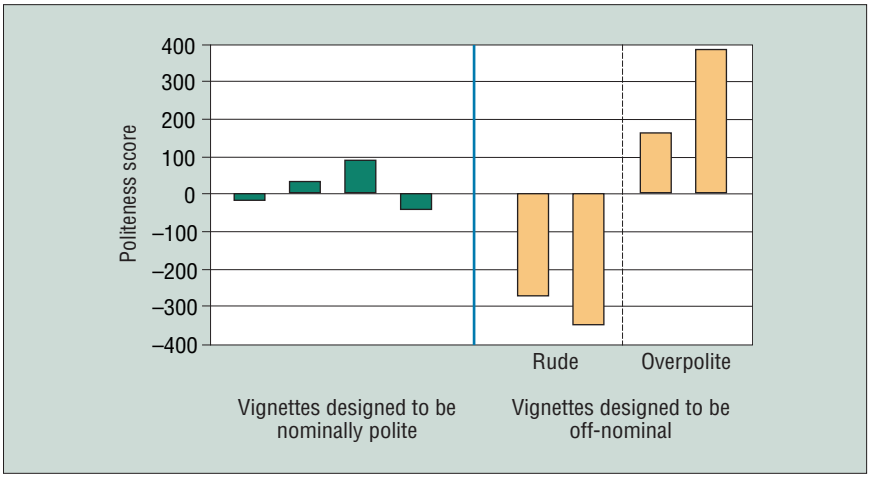

Figure 2. A priori expectations based on vignette designs along with politeness scores. We calculated imbalance scores for eight vignettes and found that they closely tracked our predictions.

Evaluation 2: Untrained rater correlations

We subsequently conducted an experiment wherein 22 American college students (respondents to flyers and unaware of our theory or model) rated the same eight vignettes. Participants reviewed a backstory describing the vignette participants and answered a series of questions about their relationships. They then read the specific, verbal interaction (utterance) and answered (using Likert scales) questions about their perceptions of the actors, their relationships, the degree of politeness they used, and whether the participants 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). Although the number of participants involved was limited, that merely reduced the probability of a significant finding, making this a conservative test.

We also asked participants whether they changed beliefs about the power and social-distance values after they had seen and heard the actors' utterances. Some vignettes used utterances expected to be nominal, whereas others used either unexpectedly high or low politeness levels. So, we hypothesized that if our model was 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.

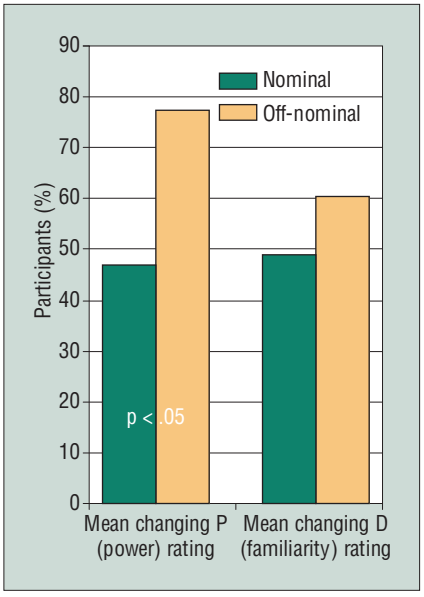

Figure 3. The effect of nominal versus off-nominal vignettes. The politeness scores for all eight vignettes fell in their expected ranges. As our model predicted, more participants wanted to change their estimates of power difference and social distance after hearing an unexpected (off-nominal) level of politeness.

As Figure 3 shows, a higher percentage of participants reported a willingness to change their ratings in response to the off-nominal vignettes, although this effect

was more pronounced for power difference (P) than for familiarity (which Brown and Levinson have called "social distance" or D) ratings. A paired-samples t-test on the mean values for the four nominal versus four off-nominal vignettes showed that significantly more participants wanted to change their estimate of power difference after reading or viewing the off-nominal behaviors than after 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. Moreover, they support the claims that unexpected (that is, off-nominal) amounts of redress prompt people to reinterpret their beliefs about context—specifically, their beliefs about the power and familiarity parameters. That participants proved more willing to review their perceptions of power relationships than social distance might be a function of the marked power relationships in the vignettes (involving, as they did, soldiers and civilians). Or, it might 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

Using Brown and Levinson's model in a module for reasoning about social-interaction behaviors ensures universal reasoning about and scoring of abstract politeness actions. However, any such module must be equipped with culture-specific knowledge bases to enable reasoning from a culture's observable behaviors (such as pursed lips or a rigid hand-to-eyebrow salute) to the abstract etiquette actions (and therefore, politeness implications) over which raters score the model's parameters. This separation of reasoning framework from culture-specific scoring knowledge has the practical implication that developers can effectively modularize an automated system's general social-interaction reasoning and thus realize large savings in simulation code development.

Using this approach, developers can easily transpose basic game storylines, training modules, and even specific characters from one cultural milieu to another. For example, the village priest the player had to interact with to get intelligence information in a Kosovo training game could take on the culture-specific behaviors and reactions of an imam (the appearance would require further work not reflected here) in an Iraqi training game simply by loading a new module of cultural knowledge. New knowledge bases of culture-specific politeness behaviors would need to be developed (and, of course, checked for accuracy) for each new simulation. However, the core game storylines and character roles, general actions, motivations, capabilities, and so on, could remain unchanged.

In previous research,[7] we've detailed our success in integrating this implemented algorithm into the Tactical Language Training System (TLTS), a language training game developed by the University of Southern California's Center for Research in Technology for Education.[10] In this research, we used our algorithm as part of the reasoning of simulated characters and demonstrated the algorithm's ability to provide both perceptions of and reactions to interactions from other simulated characters and from human game players. Furthermore, developmental efforts using our modular algorithm took less software development time than traditional scripting approaches. This work also demonstrated our approach's ability to provide at least reasonable knowledge and use of politeness levels in a culture that's not American—specifically, that of the Pashtu-speaking people along the Afghanistan-Pakistan border.

> Owing to the recombinable nature of the knowledge we capture for our etiquette algorithm, we have reason to believe that our approach scales geometrically.

Perhaps most interesting is that parsing the perception of politeness into subcomponents (power, familiarity, imposition, character, and redressive values) makes it possible to recombine these subcomponents and greatly expand the set of possible utterances a system captures—just as understanding vocabulary and syntax rules enables construction of all possible sentences in a language. By explicitly representing the knowledge to compute how one observer perceives a speaker's utterance, we simplify computing how any other characters will perceive that speech act from any other speaker. To accomplish this, we simply represent the believed power and familiarity values for the new pair. Similarly, because we know the value of adding a specific honorific to an utterance, we also know the relative worth of using that positive redress strategy for any utterance in which it makes sense.

Traditional scripting approaches to representing social-interaction behaviors in games or simulations have linear scalability; developers must create each subsequent interaction from scratch with essentially the same cost in labor as the one before it. Owing to the recombinable nature of the knowledge we capture for our etiquette algorithm, we have reason to believe that our approach scales geometrically.[7]

In fact, during our work with TLTS, we demonstrated this scalability by acquiring the knowledge for and encoding *perception scores* (PSs) that could be generated by our Etiquette Engine. A PS indicates how one observer perceives the politeness of one specific communication that a specific speaker-hearer pair uttered. We encoded our first set of 42 PSs on the TLTS project at a rate of 2.33 PSs per hour, but we acquired the next set of 378 PSs at 19.89 PSs per hour and the final set of more than 2,000 PSs at a rate of 48.96 per hour. In short, it becomes easy to recombine previously scored elements to generate the product set of possible communicative acts. The system will then automatically score each act on the basis of those previously scored elements.

Our model of politeness and etiquette allows explicit, quantifiable predictions about people's perceptions of utterances in context. Such work has multiple payoffs. In addition to the reductions in software engineering time and effort required for games and simulations described previously, we're using our model to inform a series of human-interaction experiments, exploring human reactions to variations

in the politeness with which a directive is delivered. By "directives," we mean the speech act sense of any utterance intended to direct the hearer to perform some act. So, this term covers everything from beseeching requests to direct, forceful commands and advice or warnings from a decision aid or human trainer. In short, our experiments are exploring if, 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 targeted hearers perceive as more polite will increase the overall likelihood that the hearers will comply with the directive. This should also result in increased trust for the directive giver and the perception of reduced workload. On the other hand, reaction time might be slower, because increased politeness is associated with reduced urgency, according to our model. We'll also look at whether these trends differ cross-culturally and whether the directive giver's status as human or machine makes any predictable difference in a human hearer's compliance.

Etiquette and politeness are far from the only aspects of culture that researchers should model in computational tools and approaches. However, they're a pervasive aspect of virtually all interactions that matter in cross-cultural collaboration and interaction.

An ability to score the politeness of actors' interaction behaviors in a cultural and social context allows for quantitative reasoning (and, ultimately, for machine aids and predictions) about how social-interaction moves will be perceived. So, our interpretation of Brown and Levinson's model and implementation of it in a computational form 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 react accordingly. Similarly, it can enable machine systems to determine appropriate behaviors to further ends such as trust, usability, and user acceptance. Finally, the model might also assist in managing or even filtering and translating politeness in conversations between members of different cultures. ■

Acknowledgments

This material is based on research supported by DARPA and the US Army Aviation and Missile Command under contract W31P4Q-04-C-R221. We thank Ralph Chatham, who was our program monitor, Lewis Johnson, who was our partner in much of this research, and Ning Wang, who ran the experiment we describe in the "Evaluation 2" section and performed initial data analysis. This material has been "Approved for Public Release, Distribution Unlimited."

The Authors

Christopher A. Miller is chief scientist at Smart Information Flow Technologies. His research interests include human automation integration, human performance modeling, and politeness and etiquette across cultures and in human–human and human–machine interaction. Miller received his PhD in cognition and communication from the University of Chicago. He's a member of the Human Factors and Ergonomics Society and the ACM. Contact him at cmiller@sift.info.

Peggy Wu is a research scientist at Smart Information Flow Technologies. Her research interests include how culture influences and constrains human–computer interaction designs and how and when modeling the user can improve interfaces. Wu received her BSc in computer engineering from the University of Toronto. Contact her at pwu@sift.info.

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. Funk received his BA in psychology from the University of California at Berkeley and his BS in computer science from the University of Minnesota. Contact him at hfunk@sift.info.

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. M.J. Gelfand and J.M. Brett, eds., *Handbook of Negotiation: Theoretical Advances and Cultural Perspectives*, Stanford Univ. Press, 2004.
5. P. Brown and S. Levinson, *Politeness: Some Universals in Language Usage*, Cambridge Univ. Press, 1987.
6. E. Goffman, *Interaction Ritual: Essays on Face-to-Face Behavior*, Doubleday, 1967, pp. 5–9.
7. C. Miller et al., "A Computational Approach to Etiquette and Politeness: An 'Etiquette Engine' for Cultural Interaction Training," *Proc. Behavior Representation in Modeling and Simulation Conf.* (BRIMS), Simulation Interoperability Standards Org., 2007, paper 45.
8. 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.
9. L. Johnson and P. Rizzo, "Politeness in Tutoring Dialogs: 'Run the Factory, That's What I'd Do,'" *Intelligent Tutoring Systems*, LNCS 3220, Springer, 2004, pp. 67–76.
10. W.L. Johnson, H. Vilhjalmsson, and M. Marsella, "Serious Games for Language Learning: How Much Game, How Much AI?" *Proc. 12th Int'l Conf. Artificial Intelligence in Education*, IOS Press, 2005, pp. 306–313.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.