Can Polite Computers Produce Better Human Performance?

Peggy Wu
Smart Information Flow Technologies
211 N 1st St. Suite 300
Minneapolis, MN USA
(612)669-6224
PWu@sift.info

Christopher Miller
Smart Information Flow Technologies
211 N 1st St. Suite 300
Minneapolis, MN USA
(612)716-4015
CMiller@sift.info

ABSTRACT
We claim that the concept from human-human social interactions can be expanded and utilized to facilitate, inform, and predict human-computer interaction and perceptions. By expanding on a qualitative model of politeness proposed by Brown and Levinson we created a quantitative, computational model of etiquette that allows a machine to interpret and display politeness. The results from a human subject study show that the variables included in our model have important effects on subjects' decision making and performance in our experimental tasks. The results also demonstrate that variations in etiquette can result in objective, measurable consequences in human performance.

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human Computer Interaction, Politeness, Task Performance, Socio-linguistic Theory, Computational Models

General Terms
Measurement, Documentation, Performance, Design, Experimentation, Human Factors, Languages, Theory, Verification.

Keywords
Human Computer Interaction, Politeness, Task Performance, Socio-linguistic Theory, Computational Models

1. INTRODUCTION
Etiquette is often defined as a shared code of conduct. Social etiquette, such as how to greet your new boss from Japan, can be seen as a discrete set of rules that define the proper behaviors for specific situational contexts. Those who share the same rules and interpretations of these rules, i.e. those who share the same etiquette model, have similar social expectations and may have similar interpretations of unexpected behaviors. Consequences of a lack of a shared model of etiquette range from interactions that are confusing and unproductive to those that are dangerous (e.g. plane crash [1]).

AFFINE'10, October 29, 2010, Firenze, Italy.

We claim that the concept of etiquette, i.e. social codes of conduct, expectations, and perceptions of appropriateness, can be expanded and utilized to facilitate, inform, and predict human-computer interaction and perceptions. We present a well studied and influential body of work on human-human politeness, and demonstrate that etiquette can be amendable to quantitative modeling and analysis. Further, we claim that variations in etiquette can result in objective, measurable consequences in human+machine performance.

1.1 Brown and Levinson's Theory of Politeness
A seminal body of work in the sociological and linguistic study of politeness is the cross-cultural studies and resulting model developed by Brown and Levinson [2]. Over years of cross-cultural sociological observations, Brown and Levinson found that people across languages and cultures regularly depart from relevant, concise, clear, and even truthful conversation. Consider the example where the word "please" is appended to a request. The use of please is unnecessary for a truthful, relevant or clear message yet it is used regularly. Brown and Levinson speculate that the additional politeness verbiage is necessary to mediate some ambiguities inherent in human-human communication.

The core of Brown and Levinson's model of human-human politeness is based on the social psychology concept of *face*. That is, humans have two important needs - to promote one's own autonomy and to gain social approval and connection with others [3]. All interactions inherently threaten face. In the act of simply speaking to someone, the speaker has requested the hearer's attention, and is therefore threatening the hearer's autonomy. Brown and Levinson theorize that the severity of threat is a function of the power difference between the speaker and hearer, the social distance between the speaker and hearer, and the imposition of the task on the hearer. Brown and Levinson's expression of the degree of face threat of an action is provided by the function:

$$W_x = D(S,H) + P(H,S) + R_x \quad (1)$$

- W_x is the 'weightiness' or severity of the Face Threatening Act (FTA), the degree of threat.
- $D(S,H)$ is the *social distance* between the speaker (S) and the hearer (H). It decreases with contact and interaction, but may also be based on factors such as membership in the same family, clan or organization.
- $P(H,S)$ is the *relative power* that H has over S.

- R_x is the *ranked imposition* of the raw act itself and may be culturally influenced. As an example, the imposition of asking someone for $5 is less than the imposition of asking someone for $500

Based on the severity of face threat, various politeness strategies are selected to mitigate the threat. More precisely, Brown and Levinson claim that the degree of face threat posed by an act must be balanced by the value of the politeness behaviors used if the social status quo is to be maintained. That is:

$$(2) \qquad W_x \approx V(A_x)$$

where $V(A_x)$ is the combined redressive value of the set of politeness behaviors (A_x) used in the interaction. Brown and Levinson collected and catalogued a huge database of mitigation techniques used to redress face threat, i.e. redressive strategies, and created an extensive taxonomy of these politeness behaviors across several languages and cultures. Examples range from adding the word "please" to posing requests as questions. We have used this detailed, empirical but non-quantitative model proposed by Brown and Levinson [2] to create a quantification of politeness use and politeness expectations.

1.2 Quantitative Etiquette Model

Increasingly, anecdotal and empirical evidence support the theory that humans are capable of and naturally interact with machines socially. Reeves and Nass, [4] has conducted a series of experiments demonstrating that humans readily generalize patterns of conduct and expectations from human-human interaction to human-computer interaction—a relationship they call "the media equation". If humans regularly anthropomorphize computers, then violations of etiquette may impact team collaboration and task performance like they would human-human teams. In order for the machine to interpret and display etiquette, a computational model must be in place. Expanding on the Brown and Levinson calculation of face threat, we implemented the use of weights for each component to allow the possibility to value D, P, and R differently, and added another component, character (C), to represent the speaker's general tendencies to be polite.

To translate the qualitative model into a computationally actionable model, we created a coding strategy and manual with which independent coders can evaluate and assign numeric scores to P, D, R, C, as well as politeness strategies. While this mechanism was only tested with three raters, its Robinson's A correlation of .931 was well above traditional thresholds of .7-.8 for multiple-judge rating correlations [5].

2. EXPERIMENTAL DESIGN

2.1 Definition of Variables

We identified five independent variables of interest for the study. They include the following:

- *Fixed Power* – We authored scenarios in which we manipulated power distances between subjects and virtual characters using a back-story and commonly recognized power markers such as job title.
- *Fixed Familiarity* (social distance) – Familiarity between subjects and virtual characters was manipulated in the scenarios using familiarity markers such as group identity.
- *Gender* – This is the gender of the virtual characters defined in our scenarios.
- *Redress* (etiquette) – This is the type of redressive strategy used in virtual character utterances. Each utterance in the scenarios was designed to be perceived as neutral, rude or polite.
- *Subject Type* – Subject type was either novice or professional. Novice subjects were recruited from local universities and the general community, and consisted mostly of students. Professional subjects were professional dispatchers who volunteered from an air control squadron. This variable was included because we wanted to examine the role of etiquette in strict work environments with well defined power hierarchies.

We were interested in measuring both subjective and objective performance metrics. Based on the capabilities of our test environment, we defined the following dependent variables:

- *Compliance*—This variable describes whether or not the subject responded to the requests presented by the virtual characters in the simulation (regardless of accuracy).
- *Reaction Time*—The nature of our testbed enabled us to measure different aspects of reaction time and, therefore, to compute different reaction time statistics. Reaction time measures included:
  - Directive Processing Time: The total amount of time a request was displayed on the screen.
  - Response Determination Time: The time that elapsed between when the subject completed reviewing the directive until just before s/he entered a response.
  - Response Generation Time: The amount of time the subject spent entering a response
  - Total Directive Response Time: The total amount of time the subject spent on reading the request, determining the answer, and typing in a response i.e., the sum of all three times above.
- *Accuracy*—This was calculated as the number of correct responses to a virtual character's directive, divided by the total number of directives given by that virtual character, and expressed as a percentage.
- Subject reported virtual character characteristics – This consists of a set of ratings for various aspects of the subject's perception of the virtual character. They were rated using an 11 point Likert scale and consist of:
  - *Trust in advice* of virtual character
  - *Trust in competence* of virtual character
  - *Likability(affect)* of virtual character
  - *Workload caused* by virtual character

2.2 Hypotheses

To generate the set of hypotheses, we leveraged Hofstede's taxonomy (Power Distance - PDI, Individualism - IDV, and Masculinity - MAS) [6] and paired each with the Brown and Levinson etiquette components of P and D, while holding R constant. We then reasoned about how each cultural dimension might result in variations in the expectations of high, low, and nominal levels of etiquette, and in turn how unexpected levels of etiquette might affect the performance dimensions of compliance, reaction time, accuracy, affect, workload, and trust. For example, a society with a high MAS score is one in which gender roles are clearly distinct: men are supposed to be assertive while women are supposed to be more modest and tender. (Japan has one of the highest MAS scores whereas Sweden has one of the lowest). It follows that if a human observer identifies with a high MAS score, the observer may expect a female speaker to be more polite than a male speaker even if all other variables are constant. The higher expectation then must be met by a higher amount of politeness usage. If the female speaker simply uses the same amount of politeness language as the male speaker, the observer may view the female speaker as less polite due to the gap between expected and exhibited behaviors, even if, in fact, her politeness usage was the same as her male counterpart. A failure to meet the politeness expectation may then lead to measurable consequences such as lower compliance, trust, affect, and reaction time (e.g. if you are rude to me, I will complete the task you asked of me, but I will "drag my feet" doing it). For simplicity, we have summarized our hypotheses in the results section where the findings are listed.

2.3 Methods

We divided the study into five experiments to obtain one control group and four other groups to individually vary and study the effects of cultural dimensions of interest. We varied levels of etiquette (politeness) along with power (Experiment 1), social distance or familiarity (Experiment 2), and the gender of speakers (Experiment 3). We utilized professional subjects and examined social distance in Experiment 4 to compare results with novice subjects from Experiment 2. Experiment 5 served as a control group where politeness served as the only independent variable.

2.4 Testbed

We modified a resource monitoring and management tool (see TTIMR – Tactical Tomahawk Interface for Monitoring and Retargeting [7]) to create the resulting testbed (which we called the Park Asset Management and Monitoring Interface—PAMMI). It enabled us to measure subject compliance, accuracy, and reaction time during experiments while varying independent variables. Memory, trust, affect, and workload were measured in self-report surveys after the subject completed the simulation session in the testbed.

We created a scenario where the subjects played the role of emergency vehicle dispatchers at a national park. PAMMI was their asset tracking interface and conveyed information regarding the location, intended destination, and progress of vehicles. Subjects were told that there was a group of five "field agents", who would periodically request information from them. Subjects were not told whether the field agents, or requestors, were live humans or virtual characters. Information requests arrived in the form of an onscreen dialog showing the requestor's icon and a text message. Icons rather than photos of requestors were used to reduce age, sex, and cultural associations. Messages were presented in text form rather than voice recordings to better isolate the politeness dimension. Occasionally, there would be two simultaneous requestors (speakers) and the subject was instructed to select only one of the requests to fulfill.

2.5 Experimental Stimuli

All subjects were asked to complete a set of online surveys at the beginning of the study. The surveys gathered information regarding the subject's cultural background, tendencies to generate scores (e.g. PDI, IDV, or MAS) pertinent to the experiment in which s/he was randomly assigned, and the perceived politeness of statements made in a given situation along with the subject's generated responses to the same situation. Subjects were then provided with a set of self-paced training materials on how to operate PAMMI and background information about the virtual characters. Subjects were given a 10 minute practice session in the PAMMI environment, and then proceeded to the 45 minute simulation, where one or two simultaneous requests arrived every minute. Subjects then completed a post-test survey which asked them to recall the information requester based on the content of the question (to test for memory; no significant results were found and memory will not be discussed further). The post-test survey also allowed subjects to rank perceived trust, affect, politeness, and the workload caused by each virtual character.

3. RESULTS

Below relevant hypotheses are given followed by confirmatory or contradictory results. Due to the vast amount of analyses run on the data, not all analyses conducted will be discussed.

3.1 Pre-test Results

Effects on politeness—the level of politeness should be greater for socially near and male virtual characters. *Results*: Increased familiarity (reduced social distance) was associated with increased perceived politeness in pre-test perceived politeness, $t(74)=6.47$, $p<.001$, and generated politeness questions, $t(71)=6.15$, $p<.001$. In other words, the more familiar a virtual character was, the more polite an utterance was perceived to be. Subjects also tended to judge an utterance as more polite when it came from a male, and less polite when it came from a female, $t(74) =-2.39$, $p<.05$. Similarly, subjects generated more polite utterances when they were spoken by a female asking a male for something compared to when they were from a male to a male, $t(71)=2.150$, $p<.05$.

3.2 Power Distance Index (PDI) in Experiment 1

Effects on compliance—compliance should increase for higher powered virtual characters, and increase with a higher PDI individual. *Results*: Experiment 1 showed a significant main effect of power on compliance rate, $F(1,18)=39.30$, $p<.001$, with high power virtual characters being complied with more than low power virtual characters. An ANOVA also found a significant main effect of PDI, $F(1,17)=7.99$, $p<.05$. Surprisingly, individuals with high PDI tended to comply less overall with non-neutral actors, implying they were less affected by variations in politeness or power than subjects with low PDI scores. This is contradictory to our hypothesis.

Effects on response reaction time—reaction time should decrease (become faster) for a higher powered virtual characters, and increase with a higher PDI individual. *Results*: This hypothesis was supported for response generation time. An ANOVA found a significant interaction between power and PDIVSM, $F(1,17)=6.45$, $p<.05$. High PDI subjects reacted more quickly to high powered actors. The same trend existed, but weaker, for low powered actors. Also, for paired directives, a marginal interaction between power and politeness was found for total directive response time, $F(1,6)=5.74$, $p<.055$. Subjects responded to high power rude virtual characters slower than high power polite virtual characters.

Effects on accuracy—No specific hypotheses relating to accuracy were made. *Results*: For single directives, an ANOVA showed a significant interaction between power and politeness, $F(1,18)=7.74$, $p<.05$. Subjects tended to be more accurate when responding to low power virtual characters that were rude when compared to high power, rude virtual characters.

3.3 Individualism/Collectivism (IDV) in Experiments 2 and 4

Effects on compliance—compliance should increase for a socially close virtual character, and increase for a higher IDV individual. *Results*: In Experiments 2 and 4 a significant main effect of social distance was found, $F(1,19)=15.22$, $p<.001$, $F(1,7)=5.64$, $p<.05$, respectively. For both experiments socially near virtual characters were complied with more than socially distant virtual characters, as expected. For Experiment 2, ANOVA also found a significant main effect of IDV, $F(1,18)=5.19$, $p<.05$. Compliance rates with virtual characters were higher overall for people with higher IDVCDS scores. For Experiment 4, ANOVA also found a significant interaction between IDV and social distance, $F(1,5)=7.16$, $p<.05$. Subjects with high IDV had increased compliance with unfamiliar (socially distant) actors, which is in keeping with our predictions.

Effects on response reaction time—reaction time should increase (take longer) for a socially close virtual character, and increase for a higher IDV individual. *Results*: In Experiment 4, for paired directives, a significant main effect of social distance was found for paired directive response determination time, $F(1,6)=6.92$, $p<.05$. Socially distant virtual characters were responded to faster than socially near virtual characters. For Experiment 2, ANOVA found a significant interaction between IDV and social distance for directive processing, response determination, and total directive response time, $F(1,18)=5.34$, 7.80, 7.80, respectively, $p<.05$. Subjects tended to take longer to respond to socially distant characters, except for those with very high IDV, where they took longer to respond to socially near characters. The interaction between IDV and politeness was also significant for response determination and total directive response time, $F(1,18)=11.11$ and 9.46, respectively, $p<.01$. This effect was marginal for directive processing time, $F(1,18)=4.14$, $p<.058$. In all cases subjects tended to take longer to respond to rude virtual characters, except for those with very high IDVVSM scores, who took longer to respond to polite virtual characters.

Effects on accuracy— No specific hypotheses relating to accuracy were made. *Results*: In Experiment 2, for paired directives, an ANOVA showed a significant interaction between social distance and politeness, $F(1,11)=6.81$, $p<.05$. Subjects tended to be more accurate when responding to socially distant virtual characters that were rude compared to rude, socially near virtual characters. A three-way interaction was also found in Experiment 2 between social distance, politeness and IDV, $F(1,18)=5.57$, $p<.05$. When the virtual character was polite and socially distant, accuracy rates were higher for people with higher IDV scores. When the virtual character was polite and socially near, accuracy rates were higher for people with lower IDV scores.

3.4 Masculinity/Femininity (MAS) in Experiment 3

Effects on response reaction time—reaction time should decrease (get shorter) for a higher MAS individual for male actors. *Results*: In Experiment 3, for single directives, a significant main effect of virtual character gender was observed in total directive response time, $F(1,12)=9.09$, $p<.05$, directive processing time, $F(1,12)=6.81$, $p<.05$, and response determination time, $F(1,12)=5.61$, $p<.05$. In all cases subjects took longer to respond to directives from female virtual characters than directives from male virtual characters. An ANOVA also found a three way interaction between gender, politeness, and MAS for response determination time, $F(1,11)=5.11$, $p<.05$. Breakdown of the interaction indicated that Subjects tended to take longer to respond to female rude virtual characters when they had a low MAS score, but when the MAS score was high subjects took less time to respond to female rude virtual characters. For response generation time a significant interaction between politeness and MAS was found $F(1,11)=4.90$, $p<.05$. Subjects with low MAS scores tended to take longer to respond to polite virtual characters, but those whose MAS score was high took longer to respond to rude virtual characters.

Effects on accuracy— No specific hypotheses relating to accuracy were made. *Results*: An ANOVA found a significant interaction between masculinity and MAS, $F(1,10)=6.9$, $p<.05$. When virtual characters were masculine, MAS had a negative effect on accuracy. However, when virtual characters were feminine a positive relationship between MAS and Accuracy was found.

3.5 Novice vs. Expert in Experiment 4

We predict that while professional subjects may have better overall performance, there will be no politeness effect differences between novice and professional subjects. *Results*: As predicted, professional subjects were significantly more accurate than novice subjects, $F(1,26)=21.79$, $p<.001$. However, the same ANOVA also found a significant interaction between politeness and professionalism, $F(1,26)=5.43$, $p<.05$. When virtual characters were rude there was no significant difference between experience levels, however when virtual characters were polite the accuracy of experts was greater than that of novices. Professional subjects also reacted to politeness differently than novice subjects when looking at compliance rates, as explained below.

Politeness consistently improved compliance, at least for non-professionals. Specifically, in our paired-comparisons where subjects had to choose either a polite vs. a nominal virtual character or a rude vs. a nominal virtual character, the subjects in Experiments 1, 2, 3 and 5 all chose to comply with polite virtual characters more frequently than with the rude virtual characters on average. This effect reached significance only for Experiments 2 ($F(1,19)=23.267$, $p<.001$) and 3 ($F(1,12)= 7.467$, $p<.05$). Effect sizes ranged from about 5% in Experiment 1 and 5, to nearly 40% in Experiments 2 and 3. Interestingly, though, for professional subjects in Experiment 4, politeness actually marginally decreased compliance (~10%, $p=.06$). This difference between novices' and professionals' response to politeness in directives proved significant ($F(1,26)= 12.747$, $p=.001$) in an ANOVA for compliance with paired directives.

3.6 Post-test Results

Perception of virtual character Politeness—Our manipulations of politeness in the testbed stimulus sets were effective. *Results*:

Directives that were designed to be polite were, in fact, rated as significantly more polite by subjects in Experiments 1, 2, 3 and 4, as expected; $F(1,21)=22.65$, $p<.001$; $F(1,19)=34.91$, $p<.001$; $F(1,12)=75.34$, $p<.001$; $F(1,7)=6.23$, $p<.05$; respectively. In Experiment 5, there was also a significant main effect of politeness, $F(2,36)=17.58$, $p<.001$. Subjects perceived polite virtual characters as significantly more polite than nominal and rude virtual characters, as expected, $p<.03$ and $p<.001$, respectively. Subjects also perceived nominal virtual characters as significantly more polite than rude virtual characters, as expected, $p<.01$.

Perception of virtual character likeability (affect)—affect should increase for socially close virtual characters. *Results*: In Experiment 2, socially near virtual characters were perceived as being significantly more likeable than socially distant virtual characters, $F(1,19)=7.11$, $p<0.05$. Additionally, more polite virtual characters were generally perceived as more likeable. In Experiments 1, 2, 3, and 5 results also showed that polite virtual characters were perceived as significantly more likeable than rude virtual characters, $F(1,21)=29.79$, $p<0.001$, $F(1,19)=26.08$, $p<0.001$, $F(1,12)=25.59$, $p<0.001$, $F(2,36)=13.61$, $p<0.001$, respectively. In Experiment 5, Subjects also perceived nominal virtual characters as significantly more likeable than rude virtual characters, $p<.001$.

Perception of trust—trust should increase for a socially close virtual characters. *Results*: In Experiment 2, subjects said they trusted the advice and competence of socially near virtual characters more than socially distant virtual characters, $F(1,19)=20.40$ and 9.81, respectively, $p<.01$. Additionally, politeness generally increased subjects' rating of trust in advice and competence. In Experiments 1, 2, 3 and 5 politeness significantly increased the trust subjects said they would have in advice given by virtual characters, $F(1,21)=16.04$, $p<.001$, $F(1,19)=26.75$, $p<.001$, $F(1,12)=58.58$, $p<.001$, $F(2,36)=5.56$, $p<0.01$, respectively. In Experiment 5, subjects also trusted the advice of nominal virtual characters more than the advice of rude virtual characters, $p<.05$. In Experiments 1, 2, 3, and 5 politeness significantly increased the trust subjects would have the competence of the virtual characters, $F(1,21)=4.51$, $p<.05$, $F(1,19)=9.81$, $p<.01$, $F(1,12)=17.62$, $p< .01$, $F(2,36)=5.71$, $p<.01$, respectively. In Experiment 5, subjects also trusted the competence of nominal virtual characters more than the competence of rude virtual characters, $p<.01$.

Perceived workload—Workload should increase for either polite or rude virtual characters. *Results:* Subjects generally reported greater perceived workload with rude virtual characters than with polite ones, though this trend was frequently not significant. In Experiment 1, a significant effect of politeness was found, $F(1,21)=4.54$, $p<.05$, with polite virtual characters resulting in less perceived workload than rude virtual characters. In Experiments 2, 3 and 4 no significant effects on perceived workload were found, $p>.07$, although the trend for reporting greater workload with rude virtual characters was observed in all cases. While the main effect of politeness was not significant in Experiment 5, $F(2,36)=2.48$, $p<.098$, the difference between polite and rude virtual characters was marginally significant, $p<.052$, with rude virtual characters resulting in a higher perceived workload.

3.7 Summary and Discussion

Our results indicate that the variables included in our model have important effects on subjects' decision making and performance in our experimental tasks. The more familiar a virtual character is perceived to be, the more polite an otherwise identical utterance delivered by that virtual character is perceived to be and the less polite one needs to be in providing an utterance to that virtual character. This is exactly as predicted by Brown and Levinson. Also as predicted, power, politeness and familiarity were associated with increased compliance rates. Unexpectedly, rude virtual characters that were powerful and familiar sometimes produce much lower accuracy than any other type of virtual character, while it makes little difference if a polite virtual character is familiar/unfamiliar or powerful/not powerful. Reaction times did not vary for polite and rude virtual characters; however familiar virtual characters yield longer reaction times on some components, but primarily only for professional subjects. Furthermore, subjects' ratings indicate that they found polite and familiar virtual characters more likeable and more trustworthy. Subjects also felt they experienced less workload when interacting with polite virtual characters.

The gender of the virtual character can also impact perceived politeness and compliance. Subjects tended to judge an utterance as more polite when it came from a male, and less polite when it came from a female. Similarly, subjects tended to generate more polite utterances when they were spoken by a female asking a male for something, than when they were from a male to a male. Male virtual characters also tended to be complied with more quickly than female ones.

Subjects' scores on Hofstede's cultural dimensions were found to impact performance. Keeping with predictions, high PDI individuals were more prompt in responding to high power virtual characters. Contrary to predictions, the higher a subject's PDI score, the less willing s/he was to comply with off-nominal directives. We predicted that high PDI should be associated with more discriminating selections in favor of high power individuals.

IDV scores were associated with higher overall compliance with non-neutral virtual characters and interaction effects that generally confirmed our predictions. These interactions indicate that those with high IDV scores are more likely to comply with, respond more accurately (when the virtual character is polite) and quicker to unfamiliar virtual characters. Those with high IDV are also more likely to respond quickly to rude virtual characters.

Finally, females seemed more threatening to those with higher MAS scores. This was correlated with at least some of the expected effects of higher face threat: increased accuracy in response and increased reaction time up to a point where extreme threat provokes decreased RT.

Professionals in our experiments frequently behaved similarly to non-professionals. However, professionals tended to be more accurate than non-professionals, particularly when the virtual character was polite. Additionally, while politeness tended to improve compliance rates for non-professionals, it tended to decrease compliance for professionals. Anecdotally, some professional subjects told us that they used rudeness as a cue in their interactions with pilots that the pilot was stressed and his or her need was urgent, leading us to believe that politeness nonetheless played a role in compliance for this population.

These findings are not novel to cross-cultural studies or general sociology, but instead, demonstrate the feasibility of collecting objective metrics in disciplines that are highly dependent on subjective data and self reports. We have shown that humans respond to etiquette language even in low fidelity simulations such as text based chat. Further, we have provided evidence that such responses can be measured in quantifiable ways in terms of task performance.

These findings can be used to help guide interactions. For example, if compliance with a virtual character is desired, the probability of compliance can be enhanced by using a polite, male virtual character that is both familiar and more powerful than the person receiving the directive. However, if a professional is receiving the directive the odds of compliance will increase with a rude virtual character. Additional exceptions include using neutral virtual characters for high PDI individuals and unfamiliar virtual characters for high IDV individuals. Unfortunately some characteristics that enhance compliance can result in decreased accuracy (familiarity and power) and increased reaction time (familiarity), so the desired outcome needs to be considered when choosing the best directive approach.

4. ACKNOWLEDGMENTS
We would like to thank the Air Force Research Laboratories, especially Rik Warren and Kellie Turner at AFRL Wright Patterson AFB for their support.

This work was funded in whole or in part by Department of the Air Force contract number FA8650-06-C-6635. The U.S. Government has for itself and others actin on its behalf an unlimited, paid-up, nonexclusive, irrevocable worldwide license to use, modify, reproduce, release, perform, display, or disclose the work by or on behalf of the Government.

5. REFERENCES
[1] Baron, Robert. Barriers to Effective Communication: Implications for the Cockpit. http://www.airlinesafety.com/editorials/BarriersToCommunication.htm Accessed July 2010.

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

[3] Goffman, E. (1955). On face-work: An analysis of ritual elements of social interaction. Psychiatry: Journal for the Study of Interpersonal Processes, 18, 213-231.

[4] Reeves, B., & Nass, C. (1996). *The Media Equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press.

[5] Miller, C., Wu, P., Funk, H., Wilson, P., Johnson, W. L. (2006). A Computational Approach to Etiquette and Politeness: Initial Test Cases. In *Proceedings of 2006 BRIMS Conference*, May 15-18, Baltimore, MD.

[6] Hofstede, G. (1980). *Culture's consequences*. London: Sage Publications.

[7] Cummings, M. L. (2004). The need for command and control instant message adaptive interfaces: Lessons learned from tactical tomahawk human-in-the-loop simulations. *CyberPsychology & Behavior*, 7, 653-661.