Iowa State University

From the Selected Works of Michael C. Dorneich

September 1, 2016

Politeness in Machine-Human and Human-Human Interaction

Joachim Meyer, *Tel Aviv University*

Chris Miller, *Smart Information Flow Technologies*

Peter A. Hancock, *University of Central Florida*

Ewart J. de Visser, *George Mason University*

Michael C. Dorneich, *Iowa State University*

Available at: https://works.bepress.com/michael_dorneich/85/

Politeness in machine-human and human-human interaction

Joachim Meyer (Tel Aviv University), Chris Miller (SIFT), Peter Hancock (University of Central Florida),
Ewart J. de Visser (George Mason University), Michael Dorneich (Iowa State University),

Computers communicate with humans in ways that increasingly resemble interactions between humans. Nuances in expression and responses to human behavior become more sophisticated, and they approach those of human-human interaction. The question arises whether we want systems eventually to behave like humans, or whether systems should, even when much more developed, still adhere to rules that are different from the rules governing interpersonal communication. The panel addresses this issue from various perspectives, eventually aiming to gain some insights into the question of the direction to which the development of machine-human communication and the etiquette implemented in the systems should move.

INTRODUCTION

Since J. C. R. Licklider (1960) wrote about “Human-Computer Symbiosis”, introducing the idea of a partnership between humans and computers, the idea of human-computer teams continues to appear in the literature on automation (e.g., Christofferson & Woods, 2002). In the context of human factors, the notion of trust in automation as resembling interpersonal trust has been widely studied (e. g., Lee & See, 2004; Meyer & Lee, 2013; Muir, 1994).

But not only are computers seen as resembling humans in cognitive functions, researchers also looked at humans’ relationship with devices as something that resembles the relations between people. Clifford Nass with his CASA (computers are social actors) paradigm (Lee & Nass, 2010; Reeves & Nass, 1996) showed that people tend to attribute to computers (and other devices) human properties, integrating into their interactions with devices aspects that are relevant for interpersonal interaction.

The complimentary part in the interaction between humans and machines is the question how machines should act towards humans. Initially, actions were dictated by technological limitations, taking the form of lights, printouts, or indications on the screen. However, attempts to create machines that communicate in ways that resemble interpersonal communication led already in the 1980s to a series of “talking cars” (among others, by the Japanese car manufacturer Datsun). The reception was quite negative, and the idea was rather quickly abandoned.

Developments in computer technologies made it possible to design computer systems that have more and more of the characteristics of the human communication, such as language comprehension and generation. When these computers interact with people, they can adhere to rules of conduct or politeness that also apply to interactions between people (see Hayes & Miller, 2011, for a collection of papers on the topic).

But do the roles of conduct that we expect to apply in interpersonal interactions also apply when we interact with computers or robots? There is evidence for the negative responses that occur at times when people encounter “humanized” technologies. The speaking car and the notorious Microsoft Clippy are examples of human-like communication that were rejected. In robotics, the notion of the “uncanny valley”, a feeling of aversion that arises when an object

becomes too similar to a human, was proposed by Mori in 1970 (Mori, 1970/2012) and has been studied in contexts such as robotics, animation and toys.

So should we strive to make computers increasingly similar to humans, eventually turning them into “team players” that will behave like humans? Should they abide to the same set of rules of conduct and politeness we expect humans to adhere to? Or, perhaps, should the conduct of machines remain distinct from that of humans, perhaps adhering to a different set of rules that are specific to them? The five presentations in this panel address different aspects of this issue, aiming to elicit a discussion on the directions of future research, as well as applications of more or less human-like behavior in technologies.

CONTRIBUTIONS

Etiquette expected from robot and human guards

Joachim Meyer

The interaction between people and robots in unstructured environments can have any level of resemblance to inter-human interactions. The robot can be more or less humanoid, and it can imitate actual human behavior to various degrees. In its most extreme form, as shown in science fiction, robots become indistinguishable from humans, like the devices in the classic “Bladerunner” movie or the more recent “Her” or “Ex Machina”. We are obviously still very far from realizing anything like this, but the issue of robot-human similarity still arises in some contexts even today. It is particularly relevant when robots use speech. Today’s robots can employ modes of speech used in human communication, and they can also use different levels of politeness. But should we try to make robots speak like humans?

In last-year’s HFES meeting we presented an experiment on the evaluation of a robotic guard who meets younger and older male or female people (Inbar & Meyer, 2015). The robot asks more or less politely to inspect their bags. We found that the robot was evaluated more positively (as less threatening and functioning better) when it used the politer form of addressing the person. The evaluation was, however, unaffected by the properties of the person the robot met. In a number of subsequent experiments, we explored the issue further, presenting both robot and human guards. We found that characteristics of the person with whom the human

interacted affected responses to human guards, but not the responses to the robot.

These findings may be interpreted as evidence for people expecting robots to be polite in a robotic way. We may not want to have machines that become human-like, but rather machines that become good interactors with people.

The challenge for the human factors profession is to determine what constitutes a good robot. What are human characteristics a user expects to find in a machine, and what are properties that are not really expected in the machine? The elucidation of this issue will not only contribute to the development of more acceptable robotic systems, but it may also help us understand a little better what sets human-machine and human-human interaction apart.

Models of Politeness

Chris Miller

While the effects of etiquette on human perception and reactions to behavior may differ from a robot to a human partner, many differences can be explained by the nature of etiquette itself. Etiquette is the “set of socially understood conventions that facilitate smooth and effective interaction between [actors]” (Hayes and Miller, 2010, p.2), and it is activated in any “social” situation. Social situations (Hayes and Miller, 2010; Miller, 2004) are those in which actors are perceived to have *agency* (that is, intentionality and some autonomy—cf. Dennett, 1987). This perception of agency is far more widespread than we might initially believe—as Nass showed repeatedly in his experiments where computers frequently receive the same forms of reactive social regard as humans (Reeves and Nass, 1996).

But is there a limit to this equation of humans and machines exhibiting identical etiquette behavior? The answer depends on how encompassing our definition of etiquette is.

We have been working with a model of politeness and its role in shaping/interpreting human relations created by Brown and Levinson (1987) and reduced to computational form by us (Miller, Wu & Funk, 2008; Miller, et al., 2007). Here (see Figure 1 below), the perception of politeness/rudeness depends on the power difference and the social distance/familiarity relationships between Speaker and Hearer, the imposition of an act and the “redressive value” of the behaviors or utterances that accompany the act. But all of these values are constantly being negotiated and interpreted. Each Hearer is always discovering their position relative to every Speaker—as well as the accuracy of perceived imposition and redressive value. Any assumptions of power distance, especially in unfamiliar contexts, will be based on perceptual cues and cultural norms stemming from body behaviors, appearance and role. The very values that determine whether I find a behavior “polite” are influenced, at a larger time scale, by my perceptions of politeness. If I assume we are rough power equals and yet you always address me brusquely, I may come to believe that you believe you have power over me. I may even accord it to you.

To the degree that robots do not look or behave identically to humans, they *cannot* obtain exactly the same position in the politeness “space”. But that does not mean

they fail to participate in the same politeness system. Simply because they are metal and plastic, may move or enunciate differently, and because their socio-technical roles may differ—they will afford different perceptions of the pertinent politeness-effecting dimensions. Thus, even if they perform the same function using the words or gestures a human does, their perceived politeness may still vary. This is inevitable precisely because it is not the robot exhibiting politeness (it can only provide redressive behaviors) but rather the humans perceiving it. Until and unless the perceiving human can tell no difference between robot and the human, we should always expect there to be a delta in perceived etiquette for similar actions/behaviors.

But models of human-human politeness and etiquette can still guide the design of robot behaviors to achieve desired perceptions from human users. We simply need to account for the differences in perception of the terms that affect perceived politeness. Indeed, we have been applying these techniques in design and experimental work for nearly a decade now, and are able to provide examples from eldercare, dispatch operations and other fields.

Figure 1: A model of politeness and its role in shaping/interpreting human relations

Just bad manners: Acceptable responses from human and machine

Peter Hancock

Manners are a social bond between living entities, although interestingly insensate objects also possess an inherent etiquette. When we are individually subjected to a dystopic experience, we often trace that negative affect to some nominally causal entity in the environment around us. We might stub our toe on a door stop, and while readily understanding it to be totally inert and without manifest intention, it still raises our anger and ire.

The case of technology is an intermediate one; for it is both object, but one that expresses distal human intentions. The physical structure of the technology might prove equally as cumbersome as the door stop, such as when we accidentally fall against the sharp edge of a display terminal and then strike it with our hand in a meaningless act of revenge. Yet the present concern is not for technology qua physical entity but rather technology as a conduit to desired action. As I write this, the computer may wish to update itself and I find even the least interruption an intolerable intrusion. It is not a physical source of frustration but a cognitive one. What is happening in such a case derives from the bad manners of the designer, especially one who cannot engage in the envisioned world of all subsequent actions of their design. Yet invoked

divorce in space and time cannot be considered a parole for poor etiquette.

Let us bring this issue to cases. We are now seeing, and will soon see even more, the spread of automated vehicles on to our roadways. We already witness many incidents of road-rage when individual human drivers find themselves lacking sufficient mutual manners. What happens when the automated vehicle cuts another driver off? More to the point, what of the case of extremely bad manners when that same automated vehicle causes a collision resulting in injury and death?

Here, we will be very much tempted to follow a putative causal chain which will include all involved parties from manufacturer through software sub-contractor, to individual driver. While lawyers will have a field day, the dominance of retrospective attributions of blame over prospective technical improvements enriches the few and disadvantages the many. Thus, how far are technical systems social agents and does their degree of social agency mandate a parallel level of acceptable human social interaction?

I am going to part with tradition and say it should not. Most machines need have no human manners and very little social etiquette. I base this antithesis to the common perception on the rising growth of machine autonomy and inter-machine communication. I do believe that there are necessary manners between the technologies we are creating (such as is evident in packet exchange and inter-connection protocols) but to proceed only down the road of anthropocentric mimesis in technology is a signal mistake and I shall look to defend my position in open discussion.

Mitigating user frustration through adaptive feedback based on human-automation etiquette strategies

Michael Dorneich

Observing communication through human-human interaction can provide inspiration in the design of Human-Computer Interaction (HCI). When humans interact with each other, their social behaviors are governed by expectations based on conventional norms. In this work we study the interplay of emotion, etiquette strategies, and the ability to adapt interaction styles in human-computer interaction.

Human emotion plays a role in many aspects of HCI. Emotion is a key factor in the way humans communicate information (Ferdig & Mishra, 2004). Both positive emotions (e.g. motivation) and negative emotions (e.g. frustration) are key components of learning (Woolf et al., 2009; Fisher & Noble, 2009). Negative emotions, especially frustration, are significant factors, which lead to lower performance in terms of HCI, for example, in task performance (Klein, Moon, & Picard, 2002), decision making (Bechara, 2004), and learning (Graesser, Chipman, Haynes, & Olney, 2005).

Etiquette is defined as a code that indicates conventional requirements for social behavior. Interactions between people without an understanding of etiquette may be confusing, unproductive, or even dangerous since people who share the same model of etiquette expect the same level of social behaviors from each other (Wu, Miller, Funk, & Vikili, 2010). While there are many facets of etiquette, one facet concerns the level of imposition between speaker and hearer. There are

four types of etiquette strategies that vary the level of imposition: bald, negative politeness, positive politeness, and off-record. A bald strategy is a direct way of saying something without any minimization of imposition. Positive politeness minimizes the social distance between speaker and hearer by expressing statements of friendship, solidarity, and compliments. Negative politeness attempts to be respectful however, the speaker also assumes that he or she is in some way imposing on the hearer. Off-record utterances use indirect feedback where the hearer must make some inference to recover what was intended in the feedback (Brown & Levinson, 1978).

One approach to designing responsive interactions between humans and computers is adaptive systems, which can make four types of adaptations (Feigh, Dorneich, & Hayes, 2012): (1) function allocation between the human and the automation, (2) information content; (3) task order, and (4) interaction style between the human and the system. Of these four approaches, changing the interaction style is the least explored, because of the interplay of human factors considerations. For instance, while changing the interaction style is a typical human trait when faced with certain situations, in HCI changing the way information is delivered to users may be a violation of consistency.

Putting it all together, mitigating frustration by adapting the etiquette strategy with which the computer interacts with the human may lead to better outcomes. However, it is not clear under which conditions this would be effective, and it leads to the following research questions:

1. Can the interaction style of feedback, by using etiquette strategies, mitigate user frustration?
 - a. Are there aspects of etiquette strategies that can be varied to mitigate user frustration?
 - b. Do users prefer the same (static) feedback style in all situations, or does the optimal etiquette strategies depend on the situation?
2. Does an adaptive system that employs etiquette strategies to vary the style of feedback effectively mitigate user frustration?

Humanness as a design variable to calibrate trust

Ewart de Visser

The goal for many designers of automated, intelligent or robotic systems is to emulate humans to the greatest extent possible, causing a tendency to judge designs with anthropomorphic features as inherently good or positive. Others have cautioned against automatically and universally adding humanness in the design of automation (Culley & Madhavan, 2013; Mori, 1970). *Humanness*, the extent to which an agent is designed to act and appear human, has become an important new design variable that designers must deliberately take into consideration much like the material and the algorithms that constitute any automation design. Humanness encompasses the objectively established human *capabilities* (having eyes, a face, or the ability to respond politely) of an agent by design and is distinctively different from anthropomorphism, the psychological *attributions* of human features to non-human agents (Epley, Waytz, &

Cacioppo, 2007; Mori, 1970; Nass & Moon, 2000). Designers of automated agents will have to determine when to increase and decrease humanness strategically based on their design objective (DiSalvo & Gemperle, 2002; Fink, 2012).

When to increase humanness. Increasing trust resilience by increasing humanness will be an important design objective when errors are unique, infrequent, and clumsy. Humanness can function as a buffer against losses in trust resulting from automation failures, both by lowering initial expectations and by encouraging less drastic responses to losses in reliability over time. Because automation bias is particularly damaging with clumsy automation (Wiener, 1988), high levels of automation (Parasuraman, Sheridan, & Wickens, 2000), unfair automation (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003) and high but imperfect levels of reliability (Rovira, McGarry, & Parasuraman, 2007; Sarter & Schroeder, 2001), increasing humanness to an agent could reduce such initial compliance and trust in the system. Lowering expectations of the ability of agents may be a useful design objective if an automated agent is expected to fail frequently, which is often the case when technology is new and untested. If designers expect their robots or aids to be incapable to perform certain tasks, it will help to create a human feature such as an apology or a rationale for why the aid might fail (Dzindolet, Peterson, & Pomranky, 2003) or even deny responsibility if the automation is not at fault (Kim, Dirks, Cooper, & Ferrin, 2006). More extremely, designers can lean even further on people by capitalizing on our inherent altruistic nature to help others. As an example of how this may practically work, an elegant experiment showed that when simple carton robots on wheels, with a destination label and a smiley face, wondered aimlessly in Central Park, people adjusted their direction to help them move towards their goal locations (Melanson, 2009).

When to decrease humanness. There may also be good reason to decrease or exclude humanness in the design of automated agents. The first major reason is that by increasing humanness, such as apologizing, expectations of an expected human response are also raised and if that expectation cannot be met with an actual human behavior, the agent can be seen as deceptive or lacking in integrity, which is extremely damaging to trust in an agent (Kim et al., 2006; Schweitzer, Hershey, & Bradlow, 2006). A notorious example of such an agent is Microsoft's Clippy, universally hated by mankind (Gentilviso, 2010). This research suggests that unless computers reduce errors or somehow change their behavior, trust repair activities have little use. This further suggests that automation that adaptively changes will be most effective in countering unreliable automation, especially in long-term interactions with these types of agents (Kaber, Wright, Prinzel, & Clamann, 2005; Parasuraman, Cosenzo, & de Visser, 2009). Another reason why a designer may choose to reduce or eliminate humanness from their design is if the agent should appear more logical, consistent, fair and without emotion or human judgment in situations where people are particularly vulnerable. For instance, people will disclose more information towards a virtual human when they are told it is controlled by automation compared to a human (Lucas, Gratch, King, & Morency, 2014) or express less guilt towards

machine when presenting it with a disadvantageous decision (de Melo & Gratch, 2015). Finally, humanness may need to be decreased if the addition of humanness makes it more likely that people engage unknowingly in undesirable outcomes such as increased gambling (Riva et al., 2015).

In sum, imbuing automation with human features is not always a good design objective, but should be done in a careful methodical manner (Culley & Madhavan, 2013; DiSalvo & Gemperle, 2002). The contribution to this panel is to discuss an outline for a humanness framework that allows us to study these issues systematically.

REFERENCES

- Bechara, A. (2004). The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, 55(1), 30-40.
- Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In *Questions and Politeness: Strategies in Social Interaction* (pp. 56-311). Cambridge University Press.
- Brown, P. and Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge University Press; Cambridge, UK.
- Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.) *Advances in Human Performance and Cognitive Engineering Research*, Volume 2, pp. 1- 12. JAI Press/Elsevier.
- Culley, K., & Madhavan, P. (2013). A note of caution regarding anthropomorphism in HCI agents. *Computers in Human Behavior*, 29(3), 577-579.
- de Melo, C., & Gratch, J. (2015). People Show Envy, Not Guilt, when Making Decisions with Machines. In *Proceedings of the 6th International Conference on Affective Computing and Intelligent Interaction*.
- Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: The MIT Press.
- DiSalvo, C., & Gemperle, F. (2002). All robots are not created equal: the design and perception of humanoid robot heads. In *DIS '02 Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 321-326). New York, NY.
- Dzindolet, M., Peterson, S., & Pomranky, R. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697-718.
- Epley, N., Waytz, A., & Cacioppo, J. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864-886.
- Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems a framework for researchers and system designers. *Human Factors*, 54(6), 1008-1024.
- Ferdig, R. E., & Mishra, P. (2004). Emotional responses to computers: Experiences in unfairness, anger, and spite. *Journal of Educational Multimedia and Hypermedia*, 13(2), 143.
- Fink, J. (2012). Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In *Social Robotics* (pp. 199-208). Springer Berlin Heidelberg.

- Fisher, C. D., & Noble, C. S. (2004). A within-person examination of correlates of performance and emotions while working. *Human Performance*, 17(2), 145-168.
- Gentilviso, C. (2010). The 50 Worst Inventions: Clippy. *Time*.
- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612-618.
- Hayes, C. C., & Miller, C. A. (2011). *Human-Computer Etiquette*. Boca-Raton, FL: CRC Press, Taylor & Francis.
- Inbar, O., & Meyer, J. (2015). Manners matter: Trust in robotic peacekeepers. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting*, September 2015, 59, 186-189.
- Kaber, D., et al. (2005). Adaptive Automation of Human-Machine System Information-Processing Functions. *Human Factors*, 47(4), 730-741.
- Kim, P., et al. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, 99(1), 49-65.
- Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14(2), 119-140.
- Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46 (1), 50-80.
- Lee, J.-E. R., & Nass, C. I. (2010). Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication. In D. Latusek (ed.) *Trust and Technology in a Ubiquitous Modern Environment: Theoretical and Methodological Perspectives*. IGI Global.
- Licklider, J. C. R. (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, 1, 4-11.
- Lucas, G. M., et al. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94-100.
- Melanson, D. (2009). Tweenbots project finds people really do care about robots, for now. Retrieved from <http://www.engadget.com/2009/04/13/tweenbots-project-finds-people-really-do-care-about-robots-for/>
- Meyer, J., & Lee, J. D. (2013). Trust, reliance and compliance. In J. D. Lee and A. Kirlik (eds.) *The Oxford Handbook of Cognitive Engineering*. Oxford University Press.
- Miller, C. (Guest Ed.) (2004). Human-Computer Etiquette: Managing Expectations with Intelligent Agents. *Communications of the ACM*, 47(4), 30-34.
- Miller, C., Wu, P. and Funk, H. (2008). A Computational Approach to Etiquette: Operationalizing Brown and Levinson's Politeness Model". *IEEE Intelligent Systems*, 23(4), 28-35.
- Miller, C., et al. (2007). A Computational Approach to Etiquette and Politeness: An "Etiquette Engine™" for Cultural Interaction Training. *Proceedings of the 2007 Conference on Behavior Representation in Modeling and Simulation*. Norfolk, VA; March 26-29.
- Mori, M. (1970/2012). The uncanny valley (K. F. MacDorman & N. Kageki, Trans.). *IEEE Robotics & Automation Magazine*, 19(2), 98-100.
- Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37 (11), 1905-1922.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81-103.
- Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 30(3), 286-297.
- Parasuraman, R., Cosenzo, K. a., & de Visser, E. J. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*, 21(2), 270-297.
- Reeves, B., & Nash, C. (1996). *The Media Equation: How people treat computers, television and new media like real people and places*. Cambridge, UK: Cambridge University Press.
- Riva, P., Sacchi, S., & Brambilla, M. (2015). Humanizing Machines: Anthropomorphization of slot machines increases gambling. *Journal of Experimental Psychology: Applied*. 21 (4), 313-325.
- Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, 49(1), 76-87.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, 300(5626), 1755-1758.
- Sarter, N., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: the case of in-flight icing. *Human Factors*, 43(4), 573-83.
- Schweitzer, M., Hershey, J., & Bradlow, E. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes*, 101(1), 1-19.
- Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 433-461). San Francisco, CA: Academic Press.
- Woolf, B., et al. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, 4(3-4), 129-164.
- Wu, P., et al. (2010). Computational models of etiquette and culture. *Human-Computer Etiquette: Cultural Expectations and the Design Implications they Place on Computers and Technology*, 63.