

TRUST *and* ETIQUETTE *in* HIGH-CRITICALITY AUTOMATED SYSTEMS

By RAJA PARASURAMAN AND CHRISTOPHER A. MILLER

Whereas the other articles in this section discuss human-computer etiquette in traditional social interactions involving the use of computers that explicitly strive to elicit a perception of “personhood” from the human participant, we focus on computers that occupy more traditional roles as complex and largely unpersonified machines involved in high-criticality working relationships with humans—where the consequences of failure can be catastrophic in terms of lives, money, or both.

Politeness and social niceties are important in many human-human social interactions, but in critical, highly technical work, there is the common misperception that we can “dispense with protocol” and get down to business, even with those who are not particularly courteous. In fact, excessive adherence to polite norms can seem stilted and sometimes frustrating in such settings. Here, we argue the etiquette exhibited¹ by non-personified machines (that is, ones largely without human methods of expressiveness

¹It might, perhaps, be more accurate to say the etiquette is perceived by human users rather than exhibited by the automation itself, but that subtlety is largely irrelevant to the work we review here.

such as facial expressions, speech, voice tones, and gestures) and computer-based automation can profoundly affect users' perception and correct, optimal usage of them in work domains such as aviation, manufacturing, and power plant operation, among others.

A primary candidate for a mechanism through which machine etiquette can influence usage is user trust [2, 6]. Trust influences when and whether users decide to use automation, and can therefore seriously affect the overall human + machine system reliability, efficiency, and safety.

The latter half of the 20th century witnessed explosive growth in computational power, speed, and intelligence. Engineers and designers flocked to the new information technology initially to harness it for efficiency, productivity, and safety in complex and high-criticality domains. Computers were used as faceless (if you discount blinking lights and spinning tape reels), mechanical boxes that calculated ballistic trajectories, controlled an aircraft's flight, or regulated a chemical process, long before they could interact socially in anything like natural language, much less gesture, tone of voice, and facial expression.

There is little doubt that such automation has markedly improved efficiency and productivity in many high-criticality systems. For example, the accident rate in modern automated aircraft is significantly lower than for previous generations. But widespread introduction of automation has not been entirely benign. Indeed, automation has been associated with unforeseen errors and has even been shown to cause catastrophic accidents [8].

Given this record, it is not surprising that automation has been viewed as a double-edged sword [1]. The growth of automation has motivated repeated questions about how human-automation systems should be designed to promote optimal overall system reliability, safety, and efficiency [12] and research has examined a number of factors that influence the effective use of automation. One important, intensively studied factor is trust: that is, users' willingness to believe information from a system or make use of its capabilities.

User Trust in Automation

For at least 30 years, researchers have examined the causes and effects of trust in automation (see [6] and [10] for extensive reviews). Trust is important because operators may not use a well-designed, reliable system if they believe it untrustworthy. Conversely, they may continue to rely on automation even when it malfunctions and may not monitor it effectively if they have unwarranted trust in it. In practice, user trust should be calibrated to the system and context of its use; users should have *accurate* beliefs about the reliability of

automation (and about their own capabilities).

Several factors are known to influence this trust calibration process. The user's experience of the system's reliability and accuracy is clearly important. Intuitively, operator trust should increase as experiences of correct automation behaviors increase. This has been found in studies in which users rated their trust in automation after use [6, 9]. The influence of automation reliability is typically moderated by other factors, however. For example, a user's perception of his or her own ability will be judged against the perceived reliability of the automation, and the user will typically opt for the method believed to be best [5]. Other factors have also proved important, such as the risk associated with the decision and whether a given situation matches previous automation verification experiences, operator workload, and the time criticality of the situation [10].

All of these factors, however, imply an empirical and largely analytical model by which users tune their trust in an automated system. It is also clear that as automation (and the computers that support it) increases in complexity and capabilities, there is a decline in ability and willingness to experience a range of behaviors in context and/or to think deeply about and learn the detailed mechanisms by which automation behaviors are produced. Lee and See [6], in reviewing studies from the sociological literature, argue that trust between humans can also be influenced by shortcuts to the empirical and analytical methods described here. These shortcuts involve analogical or affective responses—that is, at least in human-human relations, we tend to trust those who behave similar to people we find trustworthy and/or those with whom we enjoy interacting. We tend to trust that person more than interacting with someone who produces a negative effect (see [2] for a review of models of the role of social language in producing trust and positive effects).

The Role of Etiquette in Calibrating Trust

We define etiquette as noted in the introduction to this section as “the set of prescribed and proscribed behaviors that permits meaning and intent to be ascribed to actions.” Insofar as etiquette encodes the set of behaviors that mark individuals as members of trustworthy groups or behaviors as pleasing according to cultural norms, adherence to that etiquette can influence trust via the analogic or affective mechanisms that Lee and See [6] described in human-human interactions.

But does this link hold for human-machine interactions? Reeves and Nass [11] illustrated that people often respond socially to computer technology in ways that are similar to social interaction with humans. As one example, individuals are typically most attracted to others who appear to have personalities similar to them.

This phenomenon, called the *social attraction hypothesis* by psychologists, also predicts user acceptance of computer software [7].

Thus, we might expect aspects of etiquette that moderate the quality of human-human interaction to influence human relations to and use of technology, including automation. This might be especially true in systems where complexity makes it difficult to understand exactly how they will operate in all circumstances and, therefore, makes it attractive to rely on other cues (such as affect, group membership, certification, among others) in determining a level of trust.

Here, we explore evidence that the norms of human-human etiquette can and do affect the calibration of human trust in, and usage of, non-personified automation. First, we examine two instances of miscalibrated trust in high criticality automation and offer an etiquette-based explanation for why that miscalibration occurred. Then we present the results of a preliminary experiment where we explicitly manipulate one dimension of human-machine etiquette in order to examine its effects on user trust and usage decisions.

Etiquette and the Miscalibration of Human-Automation Trust

The consequences of inappropriate calibration of user trust can be catastrophic. An example from the maritime industry illustrates the effects of excessive trust. The cruise ship *Royal Majesty* ran aground off Nantucket after veering several miles off course toward shallow waters. Fortunately there were no injuries or fatalities, but losses totaled \$2 million in structural damage and \$5 million in lost revenue. The automated systems in this ship included an autopilot and an Automatic Radar Plotting Aid (ARPA) tied to signals received by a Global Positioning System (GPS). The autopilot normally used GPS signals to keep the ship on course, but GPS signals were lost when the cable from the antenna frayed. The autopilot then switched to dead-reckoning mode, no longer correcting for winds and tides, which carried the ship toward the shore.

According to the National Transportation Safety Board report on the accident, the probable cause was the crew's overreliance on the automation (ARPA) and management failure to ensure the crew was adequately trained in understanding automation capabilities and limitations. The report stated that "all the watch-standing officers were overly reliant on the automated position display ... and were, for all intents and purposes, sailing the map display instead of using navigation aids or lookout information."

This accident not only represents a classic case of "automation complacency" related to inappropriately high trust in the automation [6], but also suggests the role

of etiquette in the form of expectation violations. If the crew had been interacting with a human responsible for providing course recommendations from a GPS signal, they might reasonably expect to be informed in an appropriate way if the signal was lost. However, the loss of GPS signal was not displayed by the automation in a highly salient manner. As a result, the crew's etiquette-based assumptions about the lack of notification (namely, that all was as expected) proved disastrously wrong.

The opposite of overreliance—disuse of automation due to inappropriate distrust—has also been observed. We often see disuse of automated alerting systems with high false alarm rates [10], even though the systems are designed (and users are trained) to accept and investigate such false alarms. Early versions of the Traffic Collision and Alerting System (TCAS) used in commercial aviation were plagued with disuse following perceived high alarm rates. The added mental workload and perception of false alarms as a nuisance can also be interpreted in etiquette terms. Professional members of a cockpit crew are expected to be highly accurate and not to needlessly sound alarms that demand their superior's attention. A human subordinate who frequently "cries wolf" will be seen not only as unreliable but also as inappropriately demanding a supervisor's time and resources—another etiquette violation.

An Etiquette Experiment

As these case studies illustrate, inappropriate levels of trust in automation can be interpreted as resulting from etiquette violations and can profoundly impact efficiency and safety. But these are merely our interpretations of real-world instances; we wanted results from a more controlled experiment to demonstrate the interaction of etiquette, trust, and non-personified automation in high-criticality environments. As previously discussed, decreased reliability is generally correlated with decreased trust and decreased use of automation, but various factors moderate this general phenomenon. Is etiquette one such factor? Will good etiquette compensate for poor reliability and result in increased usage decisions? Will poor etiquette wipe out the benefits of good reliability? Such phenomena are not uncommon in human-human relationships, but will they extend to human-automation interactions—even in high-criticality domains?

The following experiment illustrates the effects of one dimension of etiquette on human-automation interaction. It also represents an initial attempt to establish a paradigm for investigating etiquette in non-personified human-automation interactions in a simulated high-criticality context. While its results are preliminary and based on a small sample size, they provide initial evidence that machine etiquette can strongly affect

human trust, usage decisions, and overall human + automation system performance.

Experimental design. From the vast range of human etiquette behaviors, we chose to concentrate on a single dimension we call *communication style*, which refers to the “interruptiveness” and “impatience” of delivering relevant text messages. This was chosen as an etiquette dimension available to even fairly primitive and non-personified automation interfaces. The question of when and how a computer should interrupt a user with information is one that Horvitz [3, 4] has considered in his work on “courteous computing;” However, Horvitz’ applications have not been in high-criticality domains, but rather desktop computer applications and email. It is quite possible the definition of “courteous” might either be irrelevant or substantially different on the flight deck than in the office.

We tested 16 participants (both general aviation pilots and non-pilots) on the Multi-Attribute Task (MAT) Battery, a flight simulator extensively used in previous high-criticality automation research [10]. The MAT incorporates primary flight (that is, joystick) maneuvering, fuel management, and an engine monitoring/diagnosis task (see Figure 1).

Participants always performed the first two tasks manually to simulate the busy operator of a high-criticality system. Intelligent automation support, modeled after the Engine Indicator and Crew Alerting System (EICAS) common in modern automated aircraft, was provided for the engine monitoring/diagnosis task. Automation monitored engine parameters, detecting potential engine faults, and advised participants when and what to examine to diagnose the faults. An example advisory message, presented as text, might be, “The Engine Pressure Ratio (EPR) is approaching Yellow Zone. Please check. Also, cross-check Exhaust Gas Temperature (EGT). There is a possible flame out of Engine 1.”

We operationally defined good automation etiquette in this experiment as a communication style that was “non-interruptive” and “patient” while poor automation etiquette was “interruptive” and “impatient.” In the non-interruptive case, the automation support was provided after a five-second warning and not at all when the operator was already doing the requested action (querying the EICAS system for engine param-

eters). Also, in this condition, automation was “patient” in that it would not issue a new query until the user had finished the current one. In the interruptive/impatient case, automation provided advice without warning and came on when the user was already querying EICAS. Automation also urged the next query before the user was finished with the current one (impatience).

These good and poor etiquette (in terms of communication style) conditions were crossed with the effects of automation reliability. Based on previous research [9], we chose two levels of reliability: low, in which automation provided correct advice on 6 out of 10 malfunctions, and high, when it was correct on 8 out of 10 malfunctions.

Results of experiment. Open-ended post-session interviews largely confirmed that participants saw our automation etiquette manipulations as intended: as either good or poor etiquette. For example, comments in the poor etiquette condition included “I hated it when it kept interrupting me when I was in the middle of diagnosis.” And “I

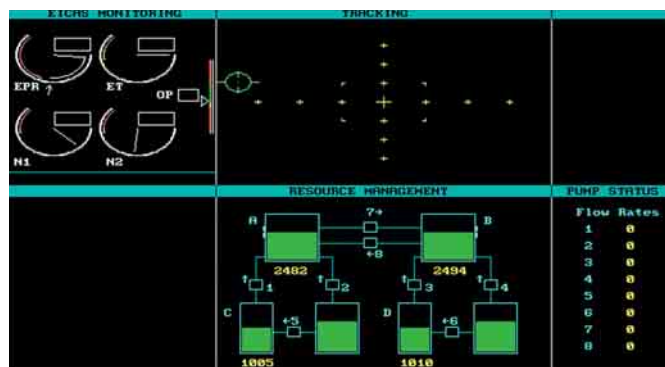


Figure 1. Sample interfaces from the MAT Battery and the EICAS display used in this experiment.

wished the computer would wait for me before asking the next question; that really bugged me.” Few such comments were made in the good etiquette condition, which elicited statements such as “The computer was nice to direct me to the next question—that was helpful,” or “One of the features I liked about the expert aid was that it was not intrusive—it came on when I needed it but not at other times.”

We were primarily interested in the effects of etiquette and reliability on users’ performance and on their rated trust in the automation. The percentage of correct diagnoses of engine malfunctions in all four conditions is shown in Figure 2. As expected, user diagnostic performance was significantly ($p < 0.01$) better when automation reliability was high (80%) than low (60%). Less obviously, good automation etiquette significantly ($p < 0.05$) enhanced diagnostic performance, regardless of automation reliability. Perhaps most interestingly, the effects of automation etiquette were powerful enough to overcome low reliability ($p < 0.05$). As the dotted line in Figure 2 indicates, performance in the low reliability/good etiquette condition was almost as good as (and not significantly different from) that in the high reliability/poor etiquette condition. These findings on diagnostic performance were mirrored in the results for

user trust, shown in Figure 3. High reliability increased trust ratings significantly, but so did good automation etiquette.

A possible objection to these findings is that any interruption might be expected to degrade user performance. While this might itself be seen as an aspect of human-machine etiquette, we wondered whether our findings were due to the “rudeness” of this automation (which interrupted to “nag” users with instructions they were already performing) or due to the simple interruption itself. It seemed to us that “rudeness” came primarily from offering redundant, task-specific information that lent itself to a perception of being told to “hurry up.” Accordingly, we ran a control group of four participants using non-specific interruptions, for example, “Maintaining primary flight performance is important, but do not forget to check engine parameters for possible malfunction.” These interruptions were varied in their intrusiveness—they were either preceded by a warning or were not given at all if the user was engaged in diagnosis (non-intrusive) or were given with no warning regardless of user activity (intrusive). Under these conditions, as expected, diagnosis of engine malfunctions and user trust were both significantly higher in the high-reliability than in the low-reliability conditions. However, neither of these measures was significantly affected by the intrusiveness factor, in contrast to the main experiment. Apparently, less rude, non-specific interruptions were more easily ignored and did not adversely affect user-system performance or trust.

Etiquette Matters, Even for High-Criticality Automation

The results of this experiment provide what we believe is the first empirical evidence for the effects of automation etiquette in a simulated high-criticality system. Strong, if preliminary, evidence was obtained for the influence of automation etiquette on both user performance and trust in using an intelligent fault management system to diagnose engine malfunctions. The

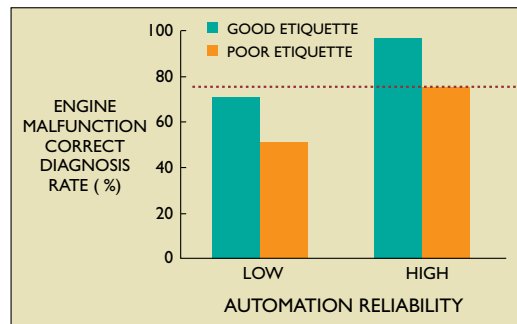


Figure 2. Effects of automation etiquette and automation reliability on the rate of correct diagnosis of engine malfunctions.

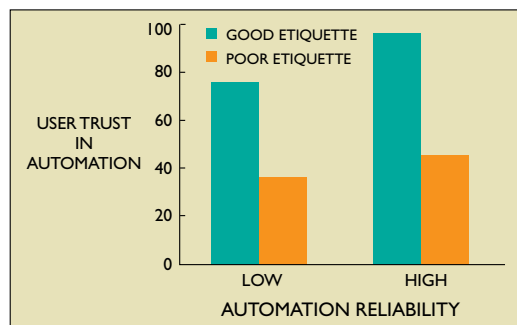


Figure 3. Effects of automation etiquette and automation reliability on subjective reports of trust in automation.

results clearly show that building reliable automation may not be enough for overall human + machine system efficiency: both user diagnostic performance and trust were lowered by poor automation etiquette even when the reliability of automation advice was high.

The results also provide support for the intriguing notion that good

automation etiquette can compensate for low automation reliability. Some may find this result disturbing, since it suggests that developing robust, sensitive, and accurate algorithms for automation—a challenging task under the best of conditions—may not be necessary as long as the automation “puts on a nice face” for the user.

We think not, for it was clear the best user performance (and the highest trust) was obtained in the high-reliability condition in which the automation also communicated its advice to the user in a polite and friendly manner. **■**

REFERENCES

- Bainbridge, L. Ironies of automation. *Automatica* (1983), 775–779.
- Cassell, J. and Bickmore, T. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and Adaptive Interfaces 12* (2002), 1–44.
- Horvitz, E. Principles of mixed-initiative user interfaces. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, PA, May 1999).
- Horvitz, E., Kadie, C., Paek, T. and Hovel, D. Models of attention in computing and communication: From principles to applications. *Commun. ACM* 46, 3 (Mar 2003), 52–59.
- Lee, J.D., and Moray, N. Trust, control strategies, and allocation of function in human-machine systems. *Ergonomics* 35 (1992), 1243–1270.
- Lee, J.D., and See, K.A. Trust in computer technology: Designing for appropriate reliance. *Human Factors* (2003).
- Nass, C., Moon, Y., Fogg, B.J., Reeves, B., and Dryer, D.C. Can computer personalities be human personalities? *International J. Human-Computer Studies* 43 (1995), 223–239.
- Parasuraman, R., and Byrne, E.A. Automation and human performance in aviation. *Principles of Aviation Psychology*. P. Tsang and M. Vidulich, Eds. Erlbaum, Mahwah, NJ, 2003, 311–356.
- Parasuraman, R., Molloy, R., and Singh, I. L. Performance consequences of automation-induced “complacency.” *International J. Aviation Psychology* 3 (1993), 1–23.
- Parasuraman, R., and Riley, V.A. Humans and automation: Use, misuse, disuse, abuse. *Human Factors* 39 (1997), 230–253.
- Reeves, B., and Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press/CSLI, NY, 1996.
- Sheridan, T.B. *Humans and Automation*. Wiley, NY, 2002.

RAJA PARASURAMAN (parasuraman@cua.edu) is a professor of psychology and director of the Cognitive Science Laboratory at The Catholic University of America, Washington, D.C.

CHRISTOPHER A. MILLER (cmiller@sift.info) is chief scientist at Smart Information Flow Technologies (SIFT) Inc., Minneapolis, MN.