Trust in Automation: Designing for Appropriate Reliance

John D. Lee and **Katrina A. See,** University of Iowa, Iowa City, Iowa

Automation is often problematic because people fail to rely upon it appropriately. Because people respond to technology socially, trust influences reliance on automation. In particular, trust guides reliance when complexity and unanticipated situations make a complete understanding of the automation impractical. This review considers trust from the organizational, sociological, interpersonal, psychological, and neurological perspectives. It considers how the context, automation characteristics, and cognitive processes affect the appropriateness of trust. The context in which the automation is used influences automation performance and provides a goal-oriented perspective to assess automation characteristics along a dimension of attributional abstraction. These characteristics can influence trust through analytic, analogical, and affective processes. The challenges of extrapolating the concept of trust in people to trust in automation are discussed. A conceptual model integrates research regarding trust in automation and describes the dynamics of trust, the role of context, and the influence of display characteristics. Actual or potential applications of this research include improved designs of systems that require people to manage imperfect automation.

INTRODUCTION

Sophisticated automation is becoming ubiquitous, appearing in work environments as diverse as aviation, maritime operations, process control, motor vehicle operation, and information retrieval. *Automation* is technology that actively selects data, transforms information, makes decisions, or controls processes. Such technology exhibits tremendous potential to extend human performance and improve safety; however, recent disasters indicate that it is not uniformly beneficial. On the one hand, people may trust automation even when it is not appropriate. Pilots, trusting the ability of the autopilot, failed to intervene and take manual control even as the autopilot crashed the Airbus A320 they were flying (Sparaco, 1995). In another instance, an automated navigation system malfunctioned and the crew failed to intervene, allowing the *Royal Majesty* cruise ship to drift off course for 24 hours before it ran aground (Lee & Sanquist, 2000; National Transportation Safety Board, 1997). On the other hand, people are not always willing to put sufficient trust in automation. Some operators rejected automated controllers in paper mills, undermining the potential benefits of the automation (Zuboff, 1988). As automation becomes more prevalent, poor partnerships between people and automation will become increasingly costly and catastrophic.

Such flawed partnerships between automation and people can be described in terms of misuse and disuse of automation (Parasuraman & Riley, 1997). *Misuse* refers to the failures that occur when people inadvertently violate critical assumptions and rely on automation inappropriately, whereas *disuse* signifies failures that occur when people reject the capabilities of automation. Misuse and disuse are two examples of inappropriate reliance on automation that can compromise safety and profitability. Although this paper describes reliance on automation as a discrete process of engaging or disengaging, automation can be a very complex combination of many modes, and reliance is often a more graded process. Automation reliance is not a simple binary process, but the simplification makes the discussion of misuse and disuse more tractable. Understanding how to mitigate

Address correspondence to John D. Lee, 2130 Seamans Center, Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, IA 52245; jdlee@engineering.uiowa.edu. **HUMAN FACTORS,** Vol. 46, No. 1, Spring 2004, pp. 50–80.

disuse and misuse of automation is a critically important problem with broad ramifications.

Recent research suggests that misuse and disuse of automation may depend on certain feelings and attitudes of users, such as trust. This is particularly important as automation becomes more complex and goes beyond a simple tool with clearly defined and easily understood behaviors. In particular, many studies show that humans respond socially to technology, and reactions to computers can be similar to reactions to human collaborators (Reeves & Nass, 1996). For example, the similarity-attraction hypothesis in social psychology predicts that people with similar personality characteristics will be attracted to each other (Nass & Lee, 2001).

This finding also predicts user acceptance of software (Nass & Lee, 2001; Nass, Moon, Fogg, Reeves, & Dryer, 1995). Software that displays personality characteristics similar to those of the user tends to be more readily accepted. For example, computers that use phrases such as "You should definitely do this" will tend to appeal to dominant users, whereas computers that use less directive language, such as "Perhaps you should do this," tend to appeal to submissive users (Nass & Lee). Similarly, the concept of affective computing suggests that computers that can sense and respond to users' emotional states may greatly improve human-computer interaction (Picard, 1997). More recently, the concept of computer etiquette suggests that human-computer interactions can be enhanced by recognizing how the social and work contexts interact with the roles of the computer and human to specify acceptable behavior (Miller, 2002). More generally, designs that consider affect are likely to enhance productivity and acceptance (Norman, Ortony, & Russell, 2003). Together, this research suggests that the emotional and attitudinal factors that influence human-human relationships may also contribute to human-automation relationships.

Trust, a social psychological concept, seems particularly important for understanding human-automation partnerships. *Trust* can be defined as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability. In this definition, an agent can be automation or another person that actively interacts with the environment on behalf of the person. Considerable research has shown the attitude of trust to be important in mediating how people rely on each other (Deutsch, 1958, 1960; Rempel, Holmes, & Zanna, 1985; Ross & LaCroix, 1996; Rotter, 1967). Sheridan (1975) and Sheridan and Hennessy (1984) argued that just as trust mediates relationships between people, it may also mediate the relationship between people and automation. Many studies have demonstrated that trust is a meaningful concept to describe human-automation interaction in both naturalistic (Zuboff, 1988) and laboratory settings (Halprin, Johnson, & Thornburry, 1973; Lee & Moray, 1992; Lewandowsky, Mundy, & Tan, 2000; Muir, 1989; Muir & Moray, 1996). These observations demonstrate that trust is an attitude toward automation that affects reliance and that it can be measured consistently. People tend to rely on automation they trust and tend to reject automation they do not. By guiding reliance, trust helps to overcome the cognitive complexity people face in managing increasingly sophisticated automation.

Trust guides – but does not completely determine – reliance, and the recent surge in research related to trust and reliance has produced many confusing and seemingly conflicting findings. Although many recent articles have described the role of trust in mediating reliance on automation, there has been no integrative review of these studies. The purpose of this paper is to provide such a review, link trust in automation to the burgeoning research on trust in other domains, and resolve conflicting findings. We begin by developing a conceptual model to link organizational, sociological, interpersonal, psychological, and neurological perspectives on trust between people to human-automation trust. We then use this conceptual model of trust and reliance to integrate research related to human-automation trust. The conceptual model identifies important research issues, and it also identifies design, evaluation, and training approaches to promote appropriate trust and reliance.

TRUST AND AFFECT

Researchers from a broad range of disciplines have examined the role of trust in mediating relationships between individuals, between

individuals and organizations, and even between organizations. Specifically, trust has been investigated as a critical factor in interpersonal relationships, where the focus is often on romantic relationships (Rempel et al., 1985). In exchange relationships, another important research area, the focus is on trust between management and employees or between supervisors and subordinates (Tan & Tan, 2000). Trust has also been identified as a critical factor in increasing organizational productivity and strengthening organizational commitment (Nyhan, 2000). Trust between firms and customers has become an important consideration in the context of relationship management (Morgan & Hunt, 1994) and Internet commerce (Muller, 1996). Researchers have even considered the issue of trust in the context of the relationship between organizations such as those in multinational firms (Ring & Vandeven, 1992), in which cross-disciplinary and cross-cultural collaboration is critical (Doney, Cannon, & Mullen, 1998).

Interest in trust has grown dramatically in the last 5 years, as many have come to recognize its importance in promoting efficient transactions and cooperation. Trust has emerged as a central focus of organizational theory (Kramer, 1999); has been the focus of recent special issues of the *Academy of Management Review* (Jones & George, 1998) and the *International Journal of Human-Computer Studies* (Corritore, Kracher, & Wiedenbeck, 2003b); was the topic of a workshop at the CHI 2001 meeting (Corritore, Kracher, & Wiedenbeck, 2001a); and has been the topic of books such as that by Kramer and Tyler (1996).

The general theme of the increasing cognitive complexity of automation, organizations, and interpersonal interactions explains the recent interest in trust. Trust tends to be less important in well-structured, stable environments, such as procedure-based hierarchical organizations, in which an emphasis on order and stability minimize transactional uncertainty (Moorman, Deshpande, & Zaltman, 1993). Many organizations, however, have recently adopted agile structures, self-directed work groups, matrix structures, and complex automation, all of which make the workplace increasingly complex, unstable, and uncertain. Because these changes enable rapid adaptation to change and accommodate unanticipated variability, there is a trend away from well-structured, procedure-based environments. Although these changes have the potential to make organizations and individuals more productive and able to adapt to the unanticipated (Vicente, 1999), they also increase cognitive complexity and leave more degrees of freedom for the individual to resolve. Trust plays a critical role in people's ability to accommodate the cognitive complexity and uncertainty that accompanies the move away from highly structured organizations and simple technology.

Trust helps people to accommodate complexity in several ways. It supplants supervision when direct observation becomes impractical, and it facilitates choice under uncertainty by acting as a social decision heuristic (Kramer, 1999). It also reduces uncertainty in gauging the responses of others, thereby guiding appropriate reliance and generating a collaborative advantage (Baba, 1999; Ostrom, 1998). Moreover, trust facilitates decentralization and adaptive behavior by making it possible to replace fixed protocols, reporting structures, and procedures with goal-related expectations regarding the capabilities of others. The increased complexity and uncertainty that has inspired the recent interest in trust in other fields parallels the increased complexity and sophistication of automation. Trust in automation guides reliance when the complexity of the automation makes a complete understanding impractical and when the situation demands adaptive behavior that procedures cannot guide. For this reason, the recent interest in trust in other disciplines provides a rich and appropriate theoretical base for understanding how trust mediates reliance on complex automation and, more generally, how it affects computer-mediated collaboration that involves both human and computer agents.

Definition of Trust: Beliefs, Attitudes, Intentions, and Behavior

Not surprisingly, the diverse interest in trust has generated many definitions. This is particularly true when considering how trust relates to automation (Cohen, Parasuraman, & Freeman, 1999; Muir, 1994). By examining the differences and common themes of these definitions, it is

possible to identify critical considerations for understanding the role of trust in mediating human-automation interaction. Some researchers focus on trust as an attitude or expectation, and they tend to define trust in one of the following ways: "expectancy held by an individual that the word, promise or written communication of another can be relied upon" (Rotter, 1967, p. 651); "expectation related to subjective probability an individual assigns to the occurrence of some set of future events" (Rempel et al., 1985, p. 96); "expectation of technically competent role performance" (Barber, 1983, p. 14); or "expectations of fiduciary obligation and responsibility, that is, the expectation that some others in our social relationships have moral obligations and responsibility to demonstrate a special concern for others' interests above their own" (Barber, p. 14). These definitions all include the element of expectation regarding behaviors or outcomes. Clearly, trust concerns an expectancy or an attitude regarding the likelihood of favorable responses.

Another common approach characterizes trust as an intention or willingness to act. This goes beyond attitude in that trust is characterized as an intention to behave in a certain manner or to enter into a state of vulnerability. For example, trust has been defined as "willingness to place oneself in a relationship that establishes or increases vulnerability with the reliance upon someone or something to perform as expected" (Johns, 1996, p. 81); "willingness to rely on an exchange partner in whom one has confidence" (Moorman et al., 1993, p. 82); and "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that party" (Mayer, Davis, & Schoorman, 1995, p. 712).

The definition by Mayer et al. (1995) is the most widely used and accepted definition of trust (Rousseau, Sitkin, Burt, & Camerer, 1998). As of April 2003, the Institute for Scientific Information citation database showed 203 citations of this article, far more than others on the topic of trust. The definition identifies vulnerability as a critical element of trust. For trust to be an important part of a relationship, individuals must willingly put themselves at risk or in vulnerable positions by delegating responsibility for actions to another party.

Some authors go beyond intention and define trust as a behavioral result or state of vulnerability or risk (Deutsch, 1960; Meyer, 2001). According to these definitions, trust is the outcome of actions that place people into certain states or situations. It can be seen as, for example, "a state of perceived vulnerability or risk that is derived from an individual's uncertainty regarding the motives, intentions, and perspective actions of others on whom they depend" (Kramer, 1999, p. 571).

These definitions highlight some important inconsistencies regarding whether trust is a belief, attitude, intention, or behavior. These distinctions are of great theoretical importance, as multiple factors mediate the process of translating beliefs and attitudes into behaviors.

Ajzen and Fishbein (1980; Fishbein & Ajzen, 1975) developed a framework that can help reconcile these conflicting definitions of trust. Their framework shows that behaviors result from intentions and that intentions are a function of attitudes. Attitudes in turn are based on beliefs. According to this framework, beliefs and perceptions represent the information base that determines attitudes. The availability of information and the person's experiences influence beliefs. An attitude is an affective evaluation of beliefs that guides people to adopt a particular intention. Intentions then translate into behavior, according to the environmental and cognitive constraints a person faces. In the context of trust and reliance, trust is an attitude and reliance is a behavior. This framework keeps beliefs, attitudes, intentions, and behavior conceptually distinct and can help explain the influence of trust on reliance. According to this framework, trust affects reliance as an attitude rather than as a belief, intention, or behavior. Beliefs underlie trust, and various intentions and behaviors may result from different levels of trust.

Considering trust as an intention or behavior has the potential to confuse its effect with the effects of other factors that can influence behavior, such as workload, situation awareness, and self-confidence of the operator (Lee & Moray, 1994; Riley, 1994). Trust is not the only factor mediating the relationship between beliefs and

behavior. Other psychological, system, and environmental constraints intervene, such as when operators do not have enough time to engage the automation even though they trust it and intend to use it, or when the effort to engage the automation outweighs its benefits (Kirlik, 1993). Trust stands between beliefs about the characteristics of the automation and the intention to rely on the automation.

Many definitions of trust also indicate the importance of the goal-oriented nature of trust. Although many definitions do not identify this aspect of trust explicitly, several mention the ability of the trustee to perform an important action and the expectation of the trustor that the trustee will perform as expected or can be relied upon (Gurtman, 1992; Johns, 1996; Mayer et al., 1995). These definitions describe the basis of trust in terms of the performance of an agent, the trustee, who furthers the goals of an individual, the trustor. In this way trust describes a relationship that depends on the characteristics of the trustee, the trustor, and the goal-related context of the interaction. Trust is not a consideration in situations where the trustor does not depend on the trustee to perform some function related to the trustor's goals.

A simple definition of trust consistent with these considerations is *the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability.* This basic definition must be elaborated to consider the appropriateness of trust, the influence of context, the goal-related characteristics of the agent, and the cognitive processes that govern the development and erosion of trust. Figure 1 shows how these factors interact in a dynamic process of reliance, and the following sections elaborate on various components of this conceptual model.

First, we will consider the appropriateness of trust. In Figure 1, appropriateness is shown

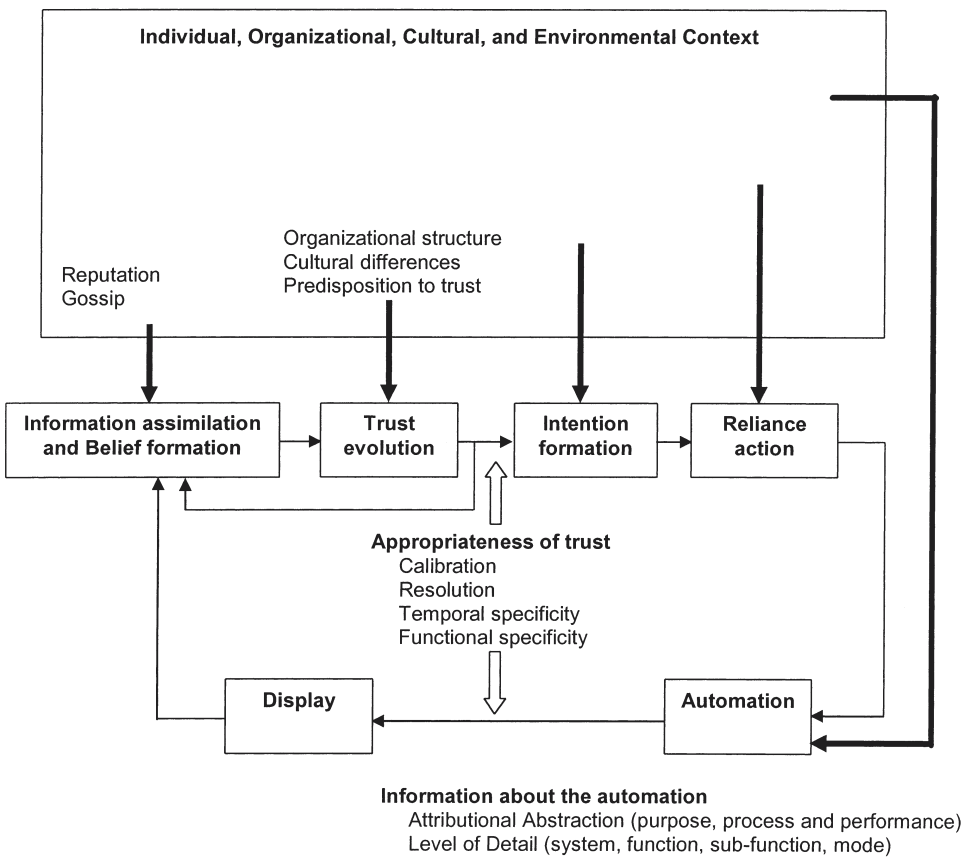

Figure 1. The interaction of context, agent characteristics, and cognitive properties with the appropriateness of trust.

as the relationship between the true capabilities of the agent and the level of trust. Second, we will consider the context defined by the characteristics of the individual, organization, and culture. Figure 1 shows this as a collection of factors at the top of the diagram, each affecting a different element of the belief-attitude-intention-behavior sequence associated with reliance on an agent. Third, we will describe the basis of trust. This defines the different types of information needed to maintain an appropriate level of trust, which the bottom of Figure 1 shows in terms of the purpose, process, and performance dimensions that describe the goal-oriented characteristics of the agent. Finally, regarding the cognitive process governing trust, trust depends on the interplay among the analytic, analogical, and affective processes. In each of these sections we will cite literature regarding the organizational, sociological, interpersonal, psychological, and neurological perspectives of human-human trust and then draw parallels with human-automation trust. The understanding of trust in automation can benefit from a multidisciplinary consideration of how context, agent characteristics, and cognitive processes affect the appropriateness of trust.

Appropriate Trust: Calibration, Resolution, and Specificity

Inappropriate reliance associated with misuse and disuse depends, in part, on how well trust matches the true capabilities of the automation. Supporting appropriate trust is critical in avoiding misuse and disuse of automation, just as it is in facilitating effective interpersonal relationships (Wicks, Berman, & Jones, 1999).

Calibration, resolution, and specificity of trust describe mismatches between trust and the capabilities of automation. *Calibration* refers to the correspondence between a person's trust in the automation and the automation's capabilities (Lee & Moray, 1994; Muir, 1987). The definitions of appropriate calibration of trust parallel those of misuse and disuse in describing appropriate reliance. *Overtrust* is poor calibration in which trust exceeds system capabilities; with *distrust,* trust falls short of the automation's capabilities. In Figure 2, good calibration is represented by the diagonal line, where the level of trust matches automation capabilities. Above this line is overtrust, and below it is distrust.

Resolution refers to how precisely a judgment of trust differentiates levels of automation

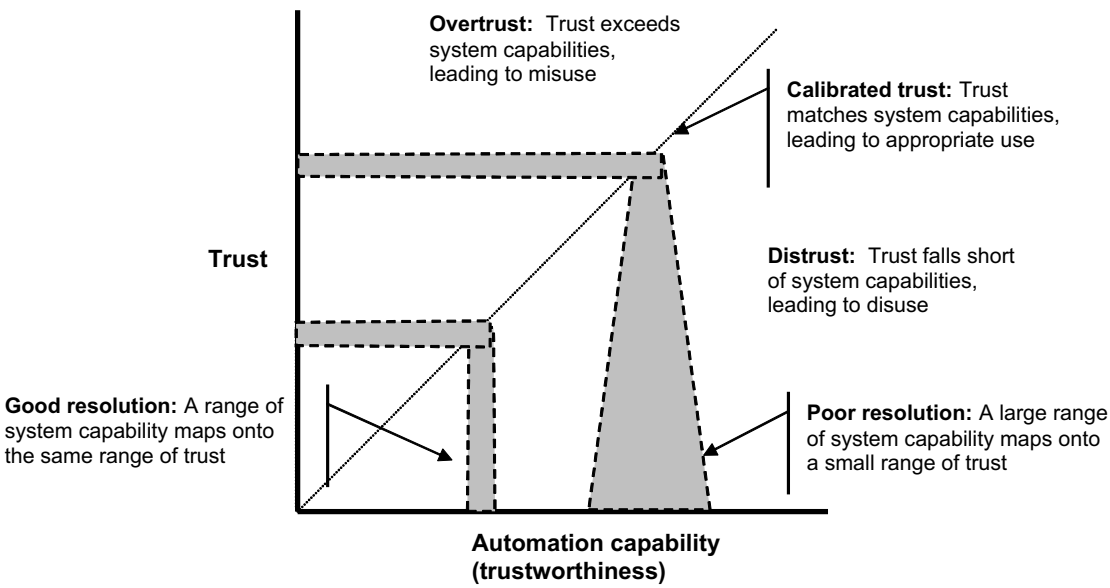

Figure 2. The relationship among calibration, resolution, and automation capability in defining appropriate trust in automation. Overtrust may lead to misuse and distrust may lead to disuse.

capability (Cohen et al., 1999). Figure 2 shows that poor resolution occurs when a large range of automation capability maps onto a small range of trust. With low resolution, large changes in automation capability are reflected in small changes in trust. *Specificity* refers to the degree to which trust is associated with a particular component or aspect of the trustee. Functional specificity describes the differentiation of functions, subfunctions, and modes of automation. With high functional specificity, a person's trust reflects capabilities of specific subfunctions and modes. Low functional specificity means the person's trust reflects the capabilities of the entire system.

Specificity can also describe changes in trust as a function of the situation or over time. High temporal specificity means that a person's trust reflects moment-to-moment fluctuations in automation capability, whereas low temporal specificity means that the trust reflects only long-term changes in automation capability. Although temporal specificity implies a generic change over time as the person's trust adjusts to failures with the automation, temporal specificity also addresses adjustments that should occur when the situation or context changes and affects the capability of the automation. Temporal specificity reflects the sensitivity of trust to changes in context that affect automation capability. High functional and temporal specificity increase the likelihood that the level of trust will match the capabilities of a particular element of the automation at a particular time. Good calibration, high resolution, and high specificity of trust can mitigate misuse and disuse of automation, and so they can guide design, evaluation, and training to enhance human-automation partnerships.

Individual, Organizational, and Cultural Context

Trust does not develop in a vacuum but, instead, evolves in a complex individual, cultural, and organizational context. The individual context includes individual differences such as the propensity to trust. These differences influence the initial level of trust and influence how new information is interpreted. The individual context also includes a person's specific history of interactions that have led to a particular level of trust. The organizational context can also have a strong influence on trust. The organizational context reflects the interactions between people that inform them about the trustworthiness of others, which can include reputation and gossip. The cultural context also influences trust through social norms and expectations. Understanding trust requires a careful consideration of the individual, organizational, and cultural context.

Systematic individual differences influence trust between people. Some people are more inclined to trust than are others (Gaines et al., 1997; Stack, 1978). Rotter (1967) defined trust as an enduring personality trait. This conception of trust follows a social learning theory approach, in which expectations for a particular situation are determined by specific previous experiences with situations that are perceived to be similar (Rotter, 1971). People develop beliefs about others that are generalized and extrapolated from one interaction to another. In this context, trust is a generalized expectancy that is independent of specific experiences and based on the generalization of a large number of diverse experiences. Individual differences regarding trust have important implications for the study of human-automation trust because they may influence reliance in ways that are not directly related to the characteristics of the automation.

Substantial evidence demonstrates that the tendency to trust, considered as a personality trait, can be reliably measured and can influence behavior in a systematic manner. For example, Rotter's (1980) Interpersonal Trust Scale reliably differentiates people on their propensity to trust others, with an internal consistency of .76 and a test-retest reliability of .56, with 7 months between tests. Interestingly, high-trust individuals are not more gullible than low-trust individuals (Gurtman, 1992), and the propensity to trust is not correlated with measures of intellect (Rotter, 1980); however, high-trust individuals are viewed as more trustworthy by others and display more truthful behavior (Rotter, 1971). In fact, people with a high propensity to trust predicted others' trustworthiness better than those with a low propensity to trust (Kikuchi, Wantanabe, & Yamasishi, 1996). Likewise, low- and high-trust individuals respond differently to

feedback regarding collaborators' intentions and to situational risk (Kramer, 1999).

These findings may explain why individual differences in the general tendency to trust automation, as measured by a complacency scale (Parasuraman, Singh, Molloy, & Parasuraman, 1992), are not clearly related to misuse of automation. For example, Singh, Molloy, and Parasuraman (1993) found that high-complacency individuals actually detected more automation failures in a constant reliability condition (53.4% compared with 18.7% for low-complacency individuals). This unexpected result is similar to findings in studies of human trust, in which highly trusting individuals were found to trust more appropriately. Several studies of trust in automation show that for some people trust changes substantially as the capability of the automation changes and for other people trust changes relatively little (Lee & Moray, 1994; Masalonis, 2000). One possible explanation that merits further investigation is that high-trust individuals may be better able to adjust their trust to situations in which the automation is highly capable as well as to situations in which it is not.

In contrast to the description of trust as a stable personality trait, most researchers have focused on trust as an attitude (Jones & George, 1998). In describing interpersonal relationships, trust has been considered as a dynamic attitude that evolves along with the developing relationship (Rempel et al., 1985). The influence of individual differences regarding the predisposition to trust is most important when a situation is ambiguous and generalized expectancies dominate, and it becomes less important as the relationship progresses (McKnight, Cummings, & Chervany, 1998). Trust as an attitude is a history-dependent variable that depends on the prior behavior of the trusted person and the information that is shared (Deutsch, 1958). The initial level of trust is determined by past experiences in similar situations; some of these experiences may even be indirect, as in the case of gossip.

The organizational context influences how reputation, gossip, and formal and informal roles affect the trust of people who have never had any direct contact with the trustee (Burt & Knez, 1996; Kramer, 1999). For example, engineers are trusted not because of the ability of any specific person but because of the underlying education and regulatory structure that governs people in the role of an engineer (Kramer, 1999). The degree to which the organizational context affects the indirect development of trust depends on the strength of links in the social network and on the trustor's ability to establish links between the situations experienced by others and the situations he or she is confronting (Doney et al., 1998). These findings suggest that the organizational context and the indirect exposure that it facilitates can have an important influence on trust and reliance.

Beyond the individual and organizational context, culture is another element of context that can influence trust and its development (Baba, Falkenburg, & Hill, 1996). *Culture* can be defined as a set of social norms and expectations that reflect shared educational and life experiences associated with national differences or distinct cohorts of workers. In particular, Japanese citizens have been found to have a generally low level of trust (Yamagishi & Yamagishi, 1994). In Japan, networks of mutually committed relations play a more important role in governing exchange relationships than they do in the United States. A cross-national questionnaire that included 1136 Japanese and 501 American respondents showed that the American respondents were more trusting of other people in general and considered reputation more important. In contrast, the Japanese respondents placed a higher value on exchange relationships that are based on personal relationships. One explanation for this is that the extreme social stability of mutually committed relationships in Japan reduces uncertainty about transactions and diminishes the role of trust (Doney et al., 1998).

More generally, cultural differences associated with power distance (e.g., dependence on authority and respect for authoritarian norms), uncertainty avoidance, and individualist and collectivist attitudes can influence the development and role of trust (Doney et al., 1998). Cultural variation can influence trust in an on-line environment. Karvonen (2001) compared trust in E-commerce service among consumers from Finland, Sweden, and Iceland. Although all trusted a simple design, there were substantial

differences in the initial trust in a complex design, with Finnish customers being the most wary and Icelandic customers the most trusting. More generally, the level of social trust varies substantially among countries, and this variation accounts for over 64% of the variance in the level of Internet adoption (Huang, Keser, Leland, & Shachat, 2002). The culturally dependent nature of trust suggests that findings regarding trust in automation may need to be verified when they are extrapolated from one cultural setting to another.

Cultural context can also describe systematic differences in groups of workers. For example, in the context of trust in automation, Riley (1996) found that pilots, who are accustomed to using automation, trusted and relied on automation more than did students who were not as familiar with automation. Likewise, in field studies of operators adapting to computerized manufacturing systems, Zuboff (1988) found that the culture associated with those operators who had been exposed to computer systems led to greater trust and acceptance of automation. These results show that trust in automation, like trust in people, is culturally influenced in that it depends on the long-term experiences of a group of people.

Trust between people depends on the individual, organizational, and cultural context. This context influences trust because it affects initial levels of trust and how people interpret information regarding the agent. Some evidence shows that these relationships may also hold for trust in automation; however, very little research has investigated any of these factors. We address the types of information that form the basis of trust in the next section.

Basis of Trust: Levels of Detail and of Attributional Abstraction

Trust is a multidimensional construct that is based on trustee characteristics such as motives, intentions, and actions (Bhattacharya, Devinney, & Pillutla, 1998; Mayer et al., 1995). The basis for trust is the information that informs the person about the ability of the trustee to achieve the goals of the trustor. Two critical elements define the basis of trust. The first is the focus of trust: What is to be trusted? The second is the type of information that describes the entity to be trusted: What is the information supporting trust? This information guides expectations regarding how well the trustee can achieve the trustor's goals. The focus of trust can be described according to levels of detail, and the information supporting trust can be described according to three levels of attributional abstraction.

The focus of trust can be considered along a dimension defined by the level of detail, which varies from general trust of institutions and organizations to trust in a specific agent. With automation, this might correspond to trust in an overall system of automation, as compared with trust in a particular mode of an automatic controller. Couch and Jones (1997) considered three levels of trust – generalized, network, and partner trust – which are similar to the three levels identified by McKnight et al. (1998). *General trust* corresponds to trust in human nature and institutions (Barber, 1983). *Network trust* refers to the cluster of acquaintances that constitutes a person's social network, whereas *partner trust* corresponds to trust in specific persons (Rempel et al., 1985).

There does not seem to be a reliable relationship between global and specific trust, suggesting that this dimension makes important distinctions regarding the evolution of trust (Couch & Jones, 1997). For example, trust in a supervisor and trust in an organization are distinct and depend on qualitatively different factors (Tan & Tan, 2000). Trust in a supervisor depends on perceived ability, integrity, and benevolence, whereas trust in the organization depends on more distal variables of perceived organizational support and justice. General trust is frequently considered to be a personality trait, whereas specific trust is frequently viewed as an outcome of a specific interaction (Couch & Jones, 1997). Several studies show, however, that general and specific trust are not necessarily tied to the "trait and state" distinctions and may instead simply refer to different levels of detail (Sitkin & Roth, 1993). Developing a high degree of functional specificity in the trust of automation may require specific information for each level of detail of the automation.

The basis of trust can be considered along a dimension of attributional abstraction, which varies from demonstrations of competence to

the intentions of the agent. A recent review of trust literature concluded that three general levels summarize the bases of trust: ability, integrity, and benevolence (Mayer et al., 1995). *Ability* is the group of skills, competencies, and characteristics that enable the trustee to influence the domain. *Integrity* is the degree to which the trustee adheres to a set of principles the trustor finds acceptable. *Benevolence* is the extent to which the intents and motivations of the trustee are aligned with those of the trustor.

In the context of interpersonal relationships, Rempel et al. (1985) viewed trust as an evolving phenomenon, with the basis of trust changing as the relationship progressed. They argued that *predictability,* the degree to which future behavior can be anticipated (and which is similar to ability), forms the basis of trust early in a relationship. This is followed by *dependability,* which is the degree to which behavior is consistent and is similar to integrity. As the relationship matures, the basis of trust ultimately shifts to *faith,* which is a more general judgment that a person can be relied upon and is similar to benevolence. A similar progression emerged in a study of operators' adaptation to new technology (Zuboff, 1988). Trust in that context depended on trial-and-error experience, followed by understanding of the technology's operation, and finally, faith. Lee and Moray (1992) made similar distinctions in defining the factors that influence trust in automation. They identified performance, process, and purpose as the general bases of trust.

Performance refers to the current and historical operation of the automation and includes characteristics such as reliability, predictability, and ability. Performance information describes *what* the automation does. More specifically, performance refers to the competency or expertise as demonstrated by its ability to achieve the operator's goals. Because performance is linked to the ability to achieve specific goals, it demonstrates the task- and situation-dependent nature of trust. This is similar to Sheridan's (1992) concept of robustness as a basis for trust in human-automation relationships. The operator will tend to trust automation that performs in a manner that reliably achieves his or her goals.

Process is the degree to which the automation's algorithms are appropriate for the situation and able to achieve the operator's goals. Process information describes *how* the automation operates. In interpersonal relationships, this corresponds to the consistency of actions associated with adherence to a set of acceptable principles (Mayer et al., 1995). Process as a basis for trust reflects a shift away from focus on specific behaviors and toward qualities and characteristics attributed to an agent. With process, trust is in the agent and not in the specific actions of the agent. Because of this, the process basis of trust relies on dispositional attributions and inferences drawn from the performance of the agent. In this sense, process is similar to dependability and integrity. Openness, the willingness to give and receive ideas, is another element of the process basis of trust. Interestingly, consistency and openness are more important for trust in peers than for trust in supervisors or subordinates (Schindler & Thomas, 1993).

In the context of automation, the process basis of trust refers to the algorithms and operations that govern the behavior of the automation. This concept is similar to Sheridan's (1992) concept of understandability in human-automation relationships. The operator will tend to trust the automation if its algorithms can be understood and seem capable of achieving the operator's goals in the current situation.

Purpose refers to the degree to which the automation is being used within the realm of the designer's intent. Purpose describes *why* the automation was developed. Purpose corresponds to faith and benevolence and reflects the perception that the trustee has a positive orientation toward the trustor. With interpersonal relationships, the perception of such a positive orientation depends on the intentions and motives of the trustee. This can take the form of abstract, generalized value congruence (Sitkin & Roth, 1993), which can be described as whether and to what extent the trustee has a motive to lie (Hovland, Janis, & Kelly, 1953). The purpose basis of trust reflects the attribution of these characteristics to the automation. Frequently, whether or not this attribution takes place will depend on whether the designer's intent has been communicated to the operator. If so, the operator will tend to trust automation to achieve the goals it was designed to achieve.

Table 1 builds on the work of Mayer et al.

TABLE 1: Summary of the Dimensions That Describe the Basis of Trust

Study	Basis of Trust	Summary Dimension
Barber (1983)	Competence	Performance
	Persistence	Process
	Fiduciary responsibility	Purpose
Butler & Cantrell (1984)	Competence	Performance
	Integrity	Process
	Consistency	Process
	Loyalty	Purpose
	Openness	Process
Cook & Wall (1980)	Ability	Performance
	Intentions	Purpose
Deutsch (1960)	Ability	Performance
	Intentions	Purpose
Gabarro (1978)	Integrity	Process
	Motives	Purpose
	Openness	Process
	Discreetness	Process
	Functional/specific competence	Performance
	Interpersonal competence	Performance
	Business sense	Performance
	Judgment	Performance
Hovland, Janis, & Kelly (1953)	Expertise	Performance
	Motivation to lie	Purpose
Jennings (1967)	Loyalty	Purpose
	Predictability	Process
	Accessibility	Process
	Availability	Process
Kee & Knox (1970)	Competence	Performance
	Motives	Purpose
Mayer et al. (1995)	Ability	Performance
	Integrity	Process
	Benevolence	Purpose
Mishra (1996)	Competency	Performance
	Reliability	Performance
	Openness	Process
	Concern	Purpose
Moorman et al. (1993)	Integrity	Process
	Willingness to reduce uncertainty	Process
	Confidentiality	Process
	Expertise	Performance
	Tactfulness	Process
	Sincerity	Process
	Congeniality	Performance
	Timeliness	Performance
Rempel et al. (1985)	Reliability	Performance
	Dependability	Process
	Faith	Purpose
Sitkin & Roth (1993)	Context-specific reliability	Performance
	Generalized value congruence	Purpose
Zuboff (1988)	Trial and error experience	Performance
	Understanding	Process
	Leap of faith	Purpose

(1995) and shows how the many characteristics of a trustee that affect trust can be summarized by the three bases of performance, process, and purpose. As an example, Mishra (1996) combined a literature review and interviews with 33 managers to identify four dimensions of trust. These dimensions can be linked to the three bases of trust: competency (performance), reliability (performance), openness (process), and concern (purpose). The dimensions of purpose, process, and performance provide a concise set of distinctions that describe the basis of trust across a wide range of application domains. These dimensions clearly identify three types of goal-oriented information that contribute to developing an appropriate level of trust.

Trust depends not only on the observations associated with these three dimensions but also on the inferences drawn among them. Observation of performance can support inferences regarding internal mechanisms associated with the dimension of process; analysis of internal mechanisms can support inferences regarding the designer's intent associated with the dimension of purpose. Likewise, the underlying process can be inferred from knowledge of the designer's intent, and performance can be estimated from an understanding of the underlying process. If these inferences support trust, it follows that a system design that facilitates them would promote a more appropriate level of trust. If inferences are inconsistent with observations, then trust will probably suffer because of a poor correspondence with the observed agent. Similarly, if inferences between levels are inconsistent, trust will suffer because of poor coherence. For example, inconsistency between the intentions conveyed by the manager (purpose basis for trust) and the manager's actions (performance basis of trust) have a particularly negative effect on trust (Gabarro, 1978). Likewise, the effort to demonstrate reliability and encourage trust by enforcing procedural approaches may not succeed if value-oriented concerns (e.g., purpose) are ignored (Sitkin & Roth, 1993).

In conditions where trust is based on several factors, it can be quite stable or robust; in situations where trust depends on a single basis, it can be quite volatile or fragile (McKnight et al., 1998). Trust that is based on an understanding of the motives of the agent will be less fragile than trust based only on the reliability of the agent's performance (Rempel et al., 1985). These findings help explain why trust in automation can sometimes appear quite robust and at other times quite fragile (Kantowitz, Hanowski, & Kantowitz, 1997b). Both imperfect coherence between dimensions and imperfect correspondence with observations are likely to undermine trust and are critical considerations for encouraging appropriate levels of trust.

The availability of information at the different levels of detail and attributional abstraction may lead to high calibration, high resolution, and great temporal and functional specificity of trust. Designing interfaces and training to provide operators with information regarding the purpose, process, and performance of automation could enhance the appropriateness of trust. However, the mere availability of information will not ensure appropriate trust. The information must be presented in a manner that is consistent with the cognitive processes underlying the development of trust, as described in the following section.

Analytic, Analogical, and Affective Processes Governing Trust

Information that forms the basis of trust can be assimilated in three qualitatively different ways. Some argue that trust is based on an analytic process; others argue that trust depends on an analogical process of assessing category membership. Most importantly, trust also seems to depend on an affective response to the violation or confirmation of implicit expectancies. Figure 3 shows how these processes may interact to influence trust. Ultimately trust is an affective response, but it is also influenced by analytic and analogical processes. The bold lines in Figure 3 show that the affective process has a greater influence on the analytic process than the analytic has on the affective (Loewenstein, Weber, Hsee, & Welch, 2001). The roles of analytic, analogical, and affective processes depend on the evolution of the relationship between the trustor and trustee, the information available to the trustor, and the way the information is displayed.

A dominant approach to trust in the organizational sciences considers trust from a rational choice perspective (Hardin, 1992). This view

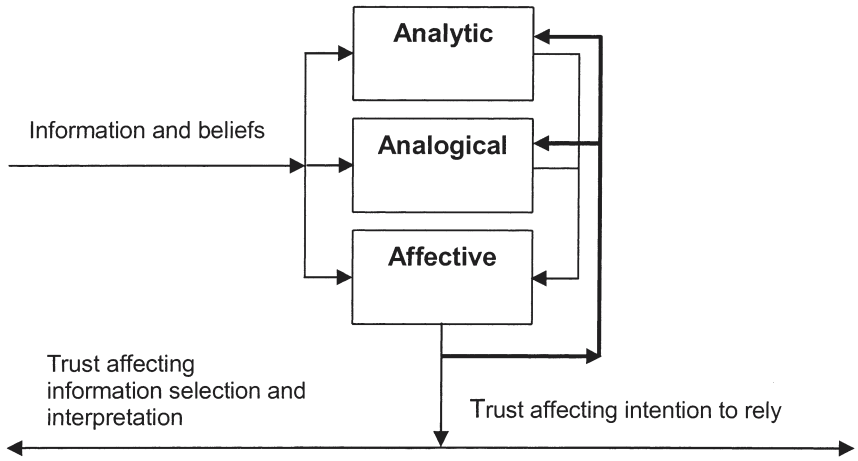

Figure 3. The interplay among the analytic, analogical, and affective process underlying trust, with the analytic and analogical processes both contributing to the affective process and the affective process also having a dominant influence on the analytic and analogical processes.

suggests that trust depends on an evaluation made under uncertainty, in which decision makers use their knowledge of the motivations and interests of the other party to maximize gains and minimize losses. Individuals are assumed to make choices based on a rationally derived assessment of costs and benefits (Lewicki & Bunker, 1996). Williamson (1993) argued that trust is primarily analytic in business and commercial relations and that it represents one aspect of rational risk analysis. Trust accumulates and dissipates based on the effect of cumulative interactions with the other agent. These interactions lead to an accumulation of knowledge that establishes trust through an expected value calculation, in which the probabilities of various outcomes are estimated with increasing experience (Holmes, 1991). Trust can also depend on an analysis of how organizational and individual constraints affect performance. For example, trust can be considered in terms of a rational argument, in which grounds for various conclusions are developed and defended according to the evidence (Cohen, 2000).

These approaches provide a useful evaluation of the normative level of trust, but they may not describe how trust actually develops. Analytic approaches to trust probably overstate cognitive capacities and understate the influence of affect and role of strategies to accommodate the limits of bounded rationality, similar to normative decision-making theories (Janis & Mann, 1977; Simon, 1957). Descriptive accounts of decision making show that expert decision makers seldom engage in conscious calculations or in an exhaustive comparison of alternatives (Crossman, 1974; Klein, 1989). The analytic process is similar to knowledge-based performance, as described by Rasmussen (1983), in which information is processed and plans are formulated and evaluated using a function-based mental model of the system. Knowledge-based or analytic processes are probably complemented by less cognitively demanding processes.

A less cognitively demanding process is for trust to develop according to analogical judgments that link levels of trust to characteristics of the agent and environmental context. These categories develop through direct observation of the trusted agent, intermediaries that convey their observations, and presumptions based on standards, category membership, and procedures (Kramer, 1999).

Explicit and implicit rules frequently guide individual and collective collaborative behavior (Mishra, 1996). Sitkin and Roth (1993) illustrated the importance of rules in their distinction between trust as an institutional arrangement and interpersonal trust. Institutional trust depends on contracts, legal procedures, and formal rules, whereas interpersonal trust is based on shared values. The application of rules in

institution-based trust reflects guarantees and safety nets based on the organizational constraints of regulations and legal recourses that constrain behavior and make it more possible to trust (McKnight et al., 1998). However, when fundamental values, which frequently support interpersonal trust, are violated, legalistic approaches and applications of rules to restore trust are ineffective and can further undermine trust (Sitkin & Roth, 1993). When rules are consistent with trustworthy behavior, they can increase people's expectation of satisfactory performance during times of normal operation by pairing situations with rule-based expectations, but rules can have a negative effect when situations diverge from normal operation.

Analogical trust can also depend on intermediaries who can convey information to support judgments of trust (Burt & Knez, 1996), such as reputations and gossip that enable individuals to develop trust without any direct contact. In the context of trust in automation, response times to warnings tend to increase when false alarms occur. This effect was counteracted by gossip that suggested that the rate of false alarms was lower than it actually was (Bliss, Dunn, & Fuller, 1995). Similarly, trust can be based on presumptions of the trustworthiness of a category or role of the person rather than on the actual performance (Kramer, Brewer, & Hanna, 1996). If trust is primarily based on rules that characterize performance during normal situations, then abnormal situations or erroneous category membership might lead to the collapse of trust because the assurances associated with normal or customary operation are no longer valid. Because of this, category-based trust can be fragile, particularly when abnormal situations disrupt otherwise stable roles. For example, when roles were defined by a binding contract, trust declined substantially when the contracts were removed (Malhotra & Murnighan, 2002).

Similar effects may occur with trust in automation. Specifically, Miller (2002) suggested that computer etiquette may have an important influence on human-computer interaction. Etiquette may influence trust because category membership associated with adherence to a particular etiquette helps people to infer how the automation will perform. Intentionally developing computer etiquette could promote appropriate trust, but it could lead to inappropriate trust if people infer inappropriate category memberships and develop distorted expectations regarding the capability of the automation.

These studies suggest that trust in automation may depend on category judgments based on a variety of sources that include direct and indirect interaction with the automation. Because these analogical judgments emerge from complex interactions between people, procedures, and prolonged experience, data from laboratory experiments may need to be carefully assessed to determine how well they generalize to actual operational settings. In this way, the analogical process governing trust corresponds very closely to Rasmussen's (1983) concept of rule-based behavior, in which condition-action pairings or procedures guide behavior.

Analytic and analogical processes do not account for the affective aspects of trust, which represent the core influence of trust on behavior (Kramer, 1999). Emotional responses are critical because people not only think about trust, they also feel it (Fine & Holyfield, 1996). Emotions fluctuate over time according to the performance of the trustee, and they can signal instances where expectations do not conform to the ongoing experience. As trust is betrayed, emotions provide signals concerning the changing nature of the situation (Jones & George, 1998).

Berscheid (1994) used the concept of automatic processing to describe how behaviors that are relevant to frequently accessed trait constructs, such as trust, are likely to be perceived under attentional overload situations. Because of this, people process observations of new behavior according to old constructs without always being aware of doing so. Automatic processing plays a substantial role in attributional activities, with many aspects of causal reasoning occurring outside conscious awareness. In this way, emotions bridge the gaps in rationality and enable people to focus their limited attentional capacity (Johnson-Laird & Oately, 1992; Loewenstein et al., 2001; Simon, 1967). Because the cognitive complexity of relationships can exceed a person's capacity to form a complete mental model, people cannot perfectly predict behavior, and emotions serve to redistribute cognitive resources and manage

priorities (Johnson-Laird & Oately). Emotions can guide behavior when rules fail to apply and when cognitive resources are not available to support a calculated rational choice.

Recent neurological evidence suggests that affect may play an important role in decision making even when cognitive resources are available to support a calculated choice. This research is important because it suggests a neurologically based description for how trust might influence reliance. Damasio, Tranel, and Damasio (1990) showed that although people with brain lesions in the ventromedial sector of the prefrontal cortices retain reasoning and other cognitive abilities, their emotions and decision-making ability are critically impaired. A series of studies have demonstrated that the decision-making deficit stems from a lack of affect and not from deficits of working memory, declarative knowledge, or reasoning, as might be expected (Bechara, Damasio, Tranel, & Anderson, 1998; Bechara, Damasio, Tranel, & Damasio, 1997). The somatic marker hypothesis explains this effect by suggesting that marker signals from the physiological response to the emotional aspects of decision situations influence the processing of information and the subsequent response to similar decision situations.

Emotions help to guide people away from situations in which negative outcomes are likely (Damasio, 1996). In a simple gambling decision-making task, patients with prefrontal lesions performed much worse than did a control group of healthy participants. The patients responded to immediate prospects and failed to accommodate long-term consequences (Bechara, Damasio, Damasio, & Anderson, 1994). In a subsequent study, healthy participants showed a substantial response to a large loss, as measured by skin conductance response (SCR), a physiological measure of emotion, whereas patients with prefrontal lesions did not (Bechara et al., 1997). Interestingly, the healthy participants also demonstrated an anticipatory SCR and began to avoid risky choices before they explicitly recognized the alternative as being risky. These results parallel the work of others and strongly support the argument that emotions play an important role in decision making (Loewenstein et al., 2001; Schwarz, 2000; Sloman, 1996) and that emotional reactions may be a critical element of trust and the decision to rely on automation.

Emotion influences not only individual decision making but also cooperation and social interaction in groups (Damasio, 2003). Specifically, Rilling et al. (2002) used an iterated prisoners' dilemma game to examine the neurological basis for social cooperation and found systematic patterns of activation in the anteroventral striatum and orbital frontal cortex after cooperative outcomes. They also found substantial individual differences regarding the level of activation and that these differences are strongly associated with the propensity to engage in cooperative behavior. These patterns may reflect emotions of trust and goodwill associated with successful cooperation (Rilling et al.). Interestingly, orbital frontal cortex activation is not limited to human-human interactions. It also occurs with human-computer interactions when the computer responds to the human's behavior in a cooperative manner; however, the anteroventral striatum activation is absent during the computer interactions (Rilling et al.).

Emotion also plays an important role by supporting implicit communication between several people (Pally, 1998). Each person's tone, rhythm, and quality of speech convey information about his or her emotional state and, so, regulate the physiological emotional responses of everyone involved. People spontaneously match nonverbal cues to generate emotional attunement. Emotional, nonverbal exchanges are a critical complement to the analytic information in a verbal exchange (Pally). These results seem consistent with recent findings regarding interpersonal trust in which anonymous computerized messages were less effective than face-to-face communication because individuals judge trustworthiness from facial expressions and from hearing the way others talk (Ostrom, 1998). Less subtle cues, such as violation of etiquette rules, which include the Gricean maxims of communication (Grice, 1975), may also lead to negative affect and undermine trust.

In the context of process control, Zuboff (1988) quoted an operator's description of the discomfort that occurs when diverse cues are eliminated through computerization: "I used to listen to sounds the boiler makes and know just how it was running. I could look at the fire

in the furnace and tell by its color how it was burning. I knew what kinds of adjustments were needed by the shades of the color I saw. A lot of the men also said that there were smells that told you different things about how it was running. I feel uncomfortable being away from these sights and smells" (p. 63).

These results show that trust may depend on a wide range of cues and that the affective basis of trust can be quite sensitive to subtle elements of human-human and human-automation interactions. Promoting appropriate trust in automation may depend on presenting information about the automation in manner compatible with the analytic, analogical, and affective processes that influence trust.

GENERALIZING TRUST IN PEOPLE TO TRUST IN AUTOMATION

Trust has emerged as an important concept in describing relationships between people, and this research may provide a good foundation for describing relationships between people and automation. However, relationships between people are qualitatively different from relationships between people and automation, so we now examine empirical and theoretical considerations in generalizing trust in people to trust in automation.

Research has addressed trust in automation in a wide range of domains. For example, trust was an important explanatory variable in understanding drivers' reactions to imperfect traffic congestion information provided by an automobile navigation system (Fox & Boehm-Davis, 1998; Kantowitz, Hanowski, & Kantowitz, 1997a). Also in the driving domain, low trust in automated hazard signs combined with low self-confidence created a double-bind situation that increased driver workload (Lee, Gore, & Campbell, 1999). Trust has also helped explain reliance on augmented vision systems for target identification (Conejo & Wickens, 1997; Dzindolet, Pierce, Beck, Dawe, & Anderson, 2001) and pilots' perception of cockpit automation (Tenney, Rogers, & Pew, 1998). Trust and self-confidence have also emerged as the critical factors in investigations into human-automation mismatches in the context of machining (Case, Sinclair, & Rani, 1999) and in the control of a teleoperated robot (Dassonville, Jolly, & Desodt, 1996). Similarly, trust seems to play an important role in discrete manufacturing systems (Trentesaux, Moray, & Tahon, 1998) and continuous process control (Lee & Moray, 1994; Muir, 1989; Muir & Moray, 1996).

Recently, trust has also emerged as a useful concept to describe interaction with Internet-based applications, such as Internet shopping (Corritore, Kracher, & Wiedenbeck, 2001b; Lee & Turban, 2001) and Internet banking (Kim & Moon, 1998). Some argue that trust will become increasingly important as the metaphor for computers moves from inanimate tools to animate software agents (Castelfranchi, 1998; Castelfranchi & Falcone, 2000; Lewis, 1998; Milewski & Lewis, 1997) and as computer-supported cooperative work becomes more commonplace (Van House, Butler, & Schiff, 1998). In combination, these results show that trust may influence misuse and disuse of automation and computer technology in many domains.

Research has demonstrated the importance of trust in automation using a wide range of experimental and observational approaches. Several studies have examined trust and affect using synthetic laboratory tasks that have no direct connection to real systems (Bechara et al., 1997; Riley, 1996), but the majority have used microworlds (Inagaki, Takae, & Moray, 1999; Lee & Moray, 1994; Moray & Inagaki, 1999). A microworld is a simplified version of a real system in which the essential elements are retained and the complexities eliminated to make experimental control possible (Brehmer & Dorner, 1993). Part-task and fully immersive driving simulators are examples of microworlds and have been used to investigate the role of trust in route-planning systems (Kantowitz et al., 1997a) and road condition warning systems (Lee et al., 1999). Field studies have also revealed the importance of trust, as demonstrated by observations of the use of autopilots and flight management systems (Mosier, Skitka, & Korte, 1994), maritime navigation systems (Lee & Sanquist, 1996), and systems in process plants (Zuboff, 1988). The range of investigative approaches demonstrates that trust is not an artifact of controlled laboratory studies and shows that it plays an important role in actual work environments.

Although substantial research suggests that

trust is a valid construct in describing human-automation interaction, several fundamental differences between human-human and human-automation trust deserve consideration in extrapolating from the research on human-human trust. A fundamental challenge in extrapolating from this research to trust in automation is that automation lacks intentionality. This becomes particularly challenging for the dimension of purpose, in which trust depends on features such as loyalty, benevolence, and value congruence. These features reflect the intentionality of the trustee and are the most fundamental and central elements of trust between people. Although automation does not exhibit intentionality or truly autonomous behavior, it is designed with a purpose and thus embodies the intentionality of the designers (Rasmussen, Pejterson, & Goodstein, 1994). In addition, people may attribute intentionality and impute motivation to automation as it becomes increasingly sophisticated and takes on human characteristics, such as speech communication (Barley, 1988; Nass & Lee, 2001; Nass & Moon, 2000).

Another important difference between interpersonal trust and trust in automation is that trust between people is often part of a social exchange relationship. Often trust is defined in a repeated-decision situation in which it emerges from the interaction between two parties (Sato, 1999). In this situation, a person may act in a trustworthy manner to elicit a favorable response from another person (Mayer et al., 1995). There is symmetry to interpersonal trust, in which the trustor and trustee are each aware of the other's behavior, intents, and trust (Deutsch, 1960). How one is perceived by the other influences behavior. There is no such symmetry in the trust between people and machines, and this leads people to trust and respond to automation differently.

Lewandowsky et al. (2000) found that delegation to human collaborators was slightly different from delegation to automation. In their study, participants operated a semiautomatic pasteurization plant in which control of the pump could be delegated. In one condition operators could delegate to an automatic controller, and in another condition they could delegate control to what they thought was another operator (in reality, the pump was controlled by the same automatic controller). Interestingly, they found that delegation to the human, but not to automation, depends on people's assessment of how others perceive them: People are more likely to delegate if they perceive their own trustworthiness to be low (Lewandowsky et al.). They also found the degree of trust to be more strongly related to the decision to delegate to the automation, as compared with the decision to delegate to the human. One explanation for these results is that operators may perceive the ultimate responsibility in a human-automation partnership to lie with the operator, whereas operators in a human-human partnership may perceive the ultimate responsibility as being shared (Lewandowsky et al.). Similarly, people are more likely to disuse automated aids than they are human aids, even though self-reports suggest a preference for automated aids (Dzindolet, Pierce, Beck, & Dawe, 2002).

These results show that fundamental differences in the symmetry of the exchange relationship, such as the ultimate responsibility for system control, may lead to a different profile of factors influencing human-human delegation and human-automation delegation.

Another important difference between interpersonal trust and trust in automation is the attribution process. Interpersonal trust evolves, particularly in romantic relationships. It often begins with a basis in performance or reliability, then progresses to a process or dependability basis, and finally evolves to a purpose or faith basis (Rempel et al., 1985). Muir (1987, 1994) has argued that trust in automation develops in a similar series of stages. However, trust in automation can also follow an opposite pattern, in which faith is important early in the interaction, followed by dependability, and then by predictability (Muir & Moray, 1996). When the purpose of the automation was conveyed by a written description, operators initially had a high level of purpose-level trust (Muir & Moray). Similarly, people seemed to trust at the level of faith when they perceived a specialist television set (e.g., one that delivers only news content) as providing better content than a generalist television set (e.g., one that can provide a variety of content; Nass & Moon, 2000).

These results contradict Muir's (1987) predictions, but this can be resolved if the three

dimensions of trust are considered as different levels of attributional abstraction, rather than as stages of development. Attributions of trust can be derived from a direct observation of system behavior (performance), an understanding of the underlying mechanisms (process), or from the intended use of the system (purpose). Early in the relationship with automation there may be little history of performance, but there may be a clear statement regarding the purpose of the automation. Thus, early in the relationship, trust can depend on purpose and not performance. The specific evolution depends on the type of information provided by the human-computer interface and the documentation and training. Trust may first be based on faith or purpose and then, as experience increases, operators may develop a feeling for the automation's dependability and predictability (Hoc, 2000). This is not a fundamental difference compared with trust between people, but it reflects how people are often thrust into relationships with automation in a way that forces trust to develop in a way different from that in many relationships between people.

A similar phenomenon occurs with temporary groups, in which trust must develop rapidly. In these situations of "swift trust," the role of dimensions underlying trust is not the same as their role in more routine interpersonal relationships (Meyerson, Weick, & Kramer, 1996). Likewise, with virtual teams trust does not progress as it does with traditional teams, in which different types of trust emerge in stages; instead, all types of trust emerge at the beginning of the relationship (Coutu, 1998). Interestingly, in a situation in which operators delegated control to automation and to what the participants thought were human collaborators, trust evolved in a similar manner, and in both cases trust dropped quickly when the performance of the trustee declined (Lewandowsky et al., 2000). Like trust between people, trust in automation develops according to the information available, rather than following a fixed series of stages.

Substantial evidence suggests that trust in automation is a meaningful and useful construct to understand reliance on automation; however, the differences in symmetry, lack of intentionality, and differences in the evolution of trust suggest that care must be exercised in extrapolating findings from human-human trust to human-automation trust. For this reason, research that considers the unique elements of human-automation trust and reliance is needed, and we turn to this next.

A DYNAMIC MODEL OF TRUST AND RELIANCE ON AUTOMATION

Figure 4 expands on the framework used to integrate studies regarding trust between humans and addresses the factors affecting trust and the role of trust in mediating reliance on automation. It complements the framework of Dzindolet et al. (2002), which focuses on the role of cognitive, social, and motivational processes that combine to influence reliance. The framework of Dzindolet et al. (2002) is particularly useful in describing how the motivational processes associated with expected effort complement cognitive processes associated with automation biases to explain reliance, issues that are distributed across the upper part of Figure 4.

As with the distinctions made by Bisantz and Seong (2001) and Riley (1994), this framework shows that trust and its effect on behavior depend on a dynamic interaction among the operator, context, automation, and interface. Trust combines with other attitudes (e.g., subjective workload, effort to engage, perceived risk, and self-confidence) to form the intention to rely on the automation. For example, exploratory behavior, in which people knowingly compromise system performance to learn how it behaves, could influence intervention or delegation (Lee & Moray, 1994). Once the intention is formed, factors such as time constraints and configuration errors may affect whether or not the person actually relies on the automation. Just as the decision to rely on the automation depends on the context, so too does the performance of that automation: Environmental variability, such as weather conditions, and a history of inadequate maintenance can degrade the performance of the automation and make reliance inappropriate.

Three critical elements of this framework are the closed-loop dynamics of trust and reliance, the importance of the context on trust and on mediating the effect of trust on reliance,

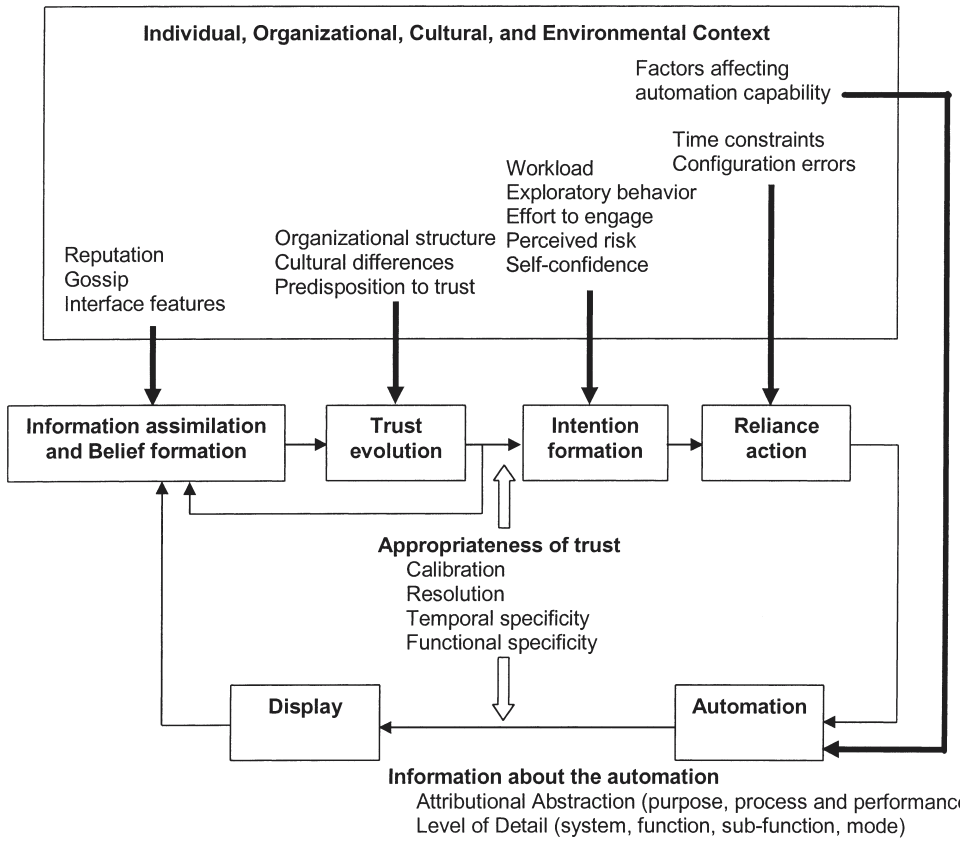

Figure 4. A conceptual model of the dynamic process that governs trust and its effect on reliance.

and the role of information display on developing appropriate trust. The research addressing trust between people provides a basis for understanding how these three factors might influence trust in automation.

The Dynamics of Trust and Reliance

Figure 4 shows that trust and its effect on reliance are part of a closed-loop process in which the dynamic interaction with the automation influences trust and trust influences the interaction with the automation. If the system is not trusted, it is unlikely to be used; if it is not used, the operator may have limited information regarding its capabilities, and it is unlikely that trust will grow. Because trust is largely based on observation of the behavior of the automation, in most cases automation must be relied upon for trust to grow (Muir & Moray, 1996). Relying on automation provides operators with an opportunity to observe how the automation works and, thus, to develop greater trust. Also, highly trusted automation may be monitored less frequently (Muir & Moray, 1996), and so trust may continue to grow even if occasional faults occur.

The type of automation may have a large effect on the ability to observe the automation and on the dynamics of trust. Four types of automation have important implications for the dynamics of trust: information acquisition, information analysis, decision selection, and action implementation (Parasuraman, Sheridan, & Wickens, 2000). These types of automation vary greatly regarding the observability of the automation when it is not being relied upon. For example, it is possible for operators to observe the behavior of information acquisition automation even when they are not relying on it, as in automation that cues potential targets (Yeh & Wickens, 2001). Information acquisition automation even makes it possible for users to dynamically

balance their attention between the information from the automation and the raw data (Wickens, Gempler, & Morphew, 2000). In contrast, it is not possible for operators to observe action implementation automation unless they are relying on it, as in automation that controls a pump (Lee & Moray, 1994). This difficulty in observing the automation after operators adopt manual control may be one reason trust fails to recover fully when the reliability of the automation improves (Lee & Moray, 1992). The relative difficulty in observing action implementation automation may also contribute to the highly nonlinear patterns of reliance observed with action implementation automation (Farrell & Lewandowsky, 2000).

Another way to distinguish types of automation is whether the automation requires reliance or compliance (Meyer, 2001). With reliance-oriented automation, the automation indicates when the system is operating normally, and the operators act when the automation fails to provide this indication. The automation provides an implicit indicator of the need to act. With compliance-oriented automation (e.g., a hazard warning), the automation indicates an abnormal situation, and the operators act in response to this indicator. The automation provides an explicit indicator of the need to act. Meyer (2001) demonstrated the importance of this distinction by showing that people were more cautious with compliance-oriented automation and that integrating compliance-oriented information with a monitored signal further increased their caution.

The dynamic interplay among changes in trust, the effect of these changes on reliance, and the resulting feedback can generate substantial nonlinear effects that can lead people to adopt either fully automatic or fully manual control (Lee & Moray, 1994). The nonlinear cascade of effects noted in the dynamics of interpersonal trust (Ostrom, 1998) and the high degree of volatility in the level of trust in some relationships suggest that a similar phenomenon occurs between people (McKnight et al., 1998). For example, initial level of trust is important because it can frame subsequent interactions, such as in resolving the ambiguity of E-mail messages (Coutu, 1998). A similar effect is seen in which the initial expectations of perfect automation performance can lead to disuse when people find the automation to be imperfect (Dzindolet et al., 2002). This nonlinear behavior may also account for the large individual differences noted in human-automation trust (Lee & Moray, 1994; Moray, Inagaki, & Itoh, 2000).

A small difference in the predisposition to trust may have a substantial effect when it influences an initial decision to engage the automation. Such minor variations in the initial level of trust might lead one person to engage the automation and another to adopt manual control. Reliance on the automation may then lead to a substantial increase in trust for the first operator, whereas the trust of the second operator might decline. Characterizing the dynamics of the feedback loop shown in Figure 4 is essential to understanding why trust sometimes appears volatile and at other times stable. This closed-loop framework also highlights the importance of temporal specificity in characterizing the appropriateness of trust. The higher the temporal specificity, the more likely that trust will correspond to the current capability of the automation. Poor temporal specificity will act as a time delay, leading to an unstable pattern of reliance and increasing the volatility of trust. The volatility of trust may depend on the characteristics of the operator, the automation, and the context in which they interact through a closed-loop feedback process, rather than on any one of these three factors alone.

Many studies have considered the effect of faults on trust in the automation. This effect is best understood by considering the development and erosion of trust as a dynamic process. Trust declines and then recovers with acute failures, and it declines with chronic failures until operators understand the fault and learn to accommodate it (Itoh, Abe, & Tanaka, 1999; Lee & Moray, 1992). Chronic failures represent a permanent decline or deficiency in the capability of the automation (e.g., a software bug), whereas acute failures represent a transient failure, after which the capability of the automation returns to its prefault capability (e.g., a maintenance-related failure that is repaired). The effect of faults on trust is not instantaneous; faults cause trust to decline over time. Likewise, the recovery after faults is not instantaneous but occurs over time (Lee & Moray, 1992).

A time-series analysis of this response suggests that a first-order differential equation can account for the change of trust over time. This means that a fault that affects operators' trust in an automated system will influence trust over time according to an exponential function. The largest effect will be seen immediately, with a residual effect distributed over time. The time-series analysis identifies the time constant of trust, which corresponds to the temporal specificity of trust and determines how quickly the trust changes to reflect changes in the capabilities of the automation. This time-series equation accounted for between 60% and 86% of the variance in the overall patterns of operators' reliance on automation (Lee & Moray, 1994). This analysis also shows that trust and reliance has inertia, which may be a critical element in understanding reliance on automation (Farrell & Lewandowsky, 2000; Lee & Moray).

This dynamic approach to decision making, as defined by the decision to rely on the automation, differs from most approaches. Most theories and analytic techniques applied to reliance and decision making adopt a static approach (i.e., one that does not consider the passage of time). Static approaches that have been used to identify miscalibration of self-confidence in decision making address only cumulative experience, rather than the evolving experience and continuous recalibration that are critical for appropriate reliance on automation (Lichtenstein, Fischhoff, & Phillips, 1982; Wright, Rowe, Bolger, & Gammack, 1994). For example, traditional decision-making theories describe the outcome of the decision process but not the vacillation that accompanies most decision making (Busemeyer & Townsend, 1993).

A dynamical approach may better reflect the factors influencing reliance and their effect over time. A dynamical approach also capitalizes on the power of differential equations to represent continuous interaction among system components over time. Such an approach supports analyses that can define attractor fields, stability metrics, and phase transition points (van Gelder & Port, 1995). These concepts have successfully described motor control (Kelso, 1995; Shaw & Kinsella-Shaw, 1988) but have been applied only recently to cognitive and social behavior (Beer, 2001; Thelen, Schoner, Scheier, & Smith, 2001). This approach may be useful in understanding trust. Specifically, a dynamical approach can describe new types of trust miscalibration – for example, hysteresis and enhanced contrast. *Hysteresis* constitutes a delayed response that depends on the direction of the change. *Enhanced contrast* is an accelerated response that also depends on the direction of the change. Hence a dynamical approach may lead to new analysis methods to describe the development and erosion of trust.

Context, Reliability, Reliance, and Trust

Figure 4 shows that reliance depends on many factors in addition to trust, particularly those affecting the capability of the automation, and that trust depends on the context. Identifying how these factors interact with trust to affect reliance can help to reconcile conflicting literature and support more robust design guidance. For example, the level of trust combines with other attitudes and expectations, such as subjective workload and self-confidence, to determine the intention to rely on the automation. A variety of system and human performance constraints also affect how the intention to rely on automation translates into actual reliance. The demands of configuring and engaging automation may make reliance inappropriate even when trust is high and the automation is very capable (Kirlik, 1993). For example, the operator may intend to use the automation but not have sufficient time to engage it. Because of this, some propose that automation should have the final authority for decisions and actions that must be allocated to automation in time-critical situations (Moray et al., 2000). Understanding the role of trust in the decision to rely on automation requires a consideration of the operating context that goes beyond trust alone.

A particularly important variable that interacts with trust to influence reliance is self-confidence. Self-confidence is a particularly critical factor in decision making in general (Bandura, 1982; Gist & Mitchell, 1992) and in mediating the effect of trust on reliance in particular (Lee & Moray, 1994). When operators' self-confidence is high and trust in the system is low, they are more inclined to rely on manual control. The opposite is also true: Low self-confidence is related to a greater inclination to rely on the automatic

controller (Lee & Moray, 1994). A study of a vehicle navigation aid found that system errors cause trust and reliance to decline more for those in a familiar city, whose self-confidence was high, as compared with those in an unfamiliar city, whose self-confidence was low (Kantowitz et al., 1997a). In a comparison of students and pilots, students tended to have greater self-confidence and were less inclined to allocate tasks to automation, which they tended to distrust. These results were attributed to the pilots' greater familiarity with the automation (Riley, 1996). A similar effect was noted in a situation where people had greater confidence in their own ability than in the automation, even though the automation performance was based on their own behavior and had exactly the same level of performance (Liu, Fuld, & Wickens, 1993). Biases in self-confidence can have a substantial effect on the appropriate reliance on automation.

The multitasking demands of a situation can also interact with trust to influence reliance. A situation in which an operator has highly reliable automation, combined with the responsibility for multiple tasks in addition to monitoring the automation, can lead to overtrust in automation and undermine the detection of automation failures (Parasuraman, Molloy, & Singh, 1993). In contrast, when the operators' only task was to monitor the automation, they detected almost all failures (Thackray & Touchstone, 1989). Because these findings might be attributed to inexperienced participants or an unrealistically high failure rate, the experiment was repeated using highly experienced pilots (Parasuraman, Mouloua, & Molloy, 1994) and infrequent faults (Molloy & Parasuraman, 1996). The findings were the same. A similar pattern of results occurs when highly reliable systems appear to reinforce the perception that other cues are redundant and unnecessary and that the task can be completely delegated to the automation (Mosier, Skitka, Heers, & Burdick, 1998).

Excessive trust and a cycle of greater trust leading to less vigilant monitoring and greater trust may explain why monitoring of automation in a multitask environment is imperfect. Another explanation is that operators adopt a eutactic monitoring pattern that can be considered optimal because it balances the costs of monitoring and the probability of failures (Moray, 2003). Similar to probability matching in signal detection theory, it may be appropriate to rely on imperfect automation even though it will lead to some incorrect outcomes (Wickens et al., 2000).

One approach to reduce overtrust in the automation and increase detection of failures is through adaptive automation, which shifts between manual and automatic control according to the capabilities of the person and the situation. In one study, participants monitored an automated engine status display while performing a tracking and fuel management task. This multitask flight simulation included adaptive automation that returned the engine status monitoring task to the person for 10 min in one condition, and in another condition the monitoring task was automated during the whole experiment. The 10 min in manual control substantially improved subsequent detection of automation failures (Parasuraman, Mouloua, & Molloy, 1996). Passive monitoring combined with the responsibility for other tasks seems to increase reliance on the automation.

Faults with automation and environmental situations that cause automation to perform poorly are fundamental factors affecting trust and the disuse of automation. Not surprisingly, many studies have shown that automation faults cause trust to decline. Importantly, this effect is often specific to the automatic controller with the fault. A fault in the automation did not cause trust to decline in other similar, but functionally independent, automatic controllers (Lee & Moray, 1992; Muir & Moray, 1996). A multitrait-multimethod analysis also showed that trust was specific to the particular automatic controller and independent of self-confidence (Lee & Moray, 1994). These results indicate that the functional specificity of trust is not limited to the system as a whole but can be linked to particular controllers. The exact degree of functional specificity depends on the information presented to operators and on their goals in controlling the system.

In a similar study, participants rated their trust in a pasteurization system on three subjective scales corresponding to the purpose, process, and product dimensions of trust. Changes in the level of system performance accounted for a greater percentage of variance in trust ratings associated with the performance dimension

than in the ratings associated with process or purpose. The rating that corresponded to the process dimension was most influenced by the presence or absence of a fault, as compared with the process or purpose dimensions (Lee & Moray, 1992). These results show that trust reflects the effects of faults in a way that is consistent with the theoretical expectations outlined in Figure 4.

Trust is a nonlinear function of automation performance and the dynamic interaction between the operator and the automation. It tends to be conditioned by the worst behaviors of an automated system (Muir & Moray, 1996), which is also the case with people: Negative interactions have a greater influence than do positive interactions (Kramer, 1999). In addition, initial experience has a lasting effect on trust: An initially low level of reliability leads to lower trust and reliance when reliability subsequently improves. Likewise, trust is more resilient if automation reliability starts high and declines than if it starts low and increases (Fox & Boehm-Davis, 1998). This effect may reflect nonlinear feedback mechanisms in which a distrusted system is not used and so becomes even less trusted and less used.

Beyond the effect of initial levels of trust, trust decreases with decreasing reliability, but some evidence suggests that below a certain level of reliability, trust declines quite rapidly. The absolute level of this drop-off seems to be highly system and context dependent, with estimates ranging from 90% (Moray et al., 2000) and 70% (Kantowitz et al., 1997a) to 60% (Fox, 1996). Another consistent finding is that the effect of an automation fault on trust depends as much on its predictability as on its magnitude. A small fault with unpredictable results affected trust more than did a large fault of constant error (Moray et al.; Muir & Moray, 1996). Trust can develop when a systematic fault occurs for which a control strategy can be developed. Consistent with this result, reliance on automation follows the level of trust, and people rely on faulty automation if they have prior knowledge of the fault (Riley, 1994). Similarly, a discrepancy between the operator's expectations and the behavior of the automation can undermine trust even when the automation performs well (Rasmussen et al., 1994). No single level of reliability can be identified that will lead to distrust and disuse; instead, trust depends on the timing, consequence, and expectations associated with failures of the automation.

The environmental context not only influences trust and reliance, it can also influence the performance of the automation: The automation may perform well in certain circumstances and not in others. For this reason, appropriateness of trust often depends on how sensitive people are to the influence of the environment on the performance of the automation. Precisely resolving differences in the context can lead to more appropriate trust (Cohen, 2000). For example, Masalonis (2000) described how training enhanced the appropriateness of trust in the context of situation-specific reliability of decision aids for air traffic controllers.

A similar benefit was found in a situation in which aircraft predictor information was only partially reliable. Knowing that the predictor was not completely reliable helped pilots to calibrate their trust and adopt an appropriate allocation of attention between the raw data and the predictor information (Wickens et al., 2000). Similarly, Bisantz and Seong (2001) showed that failures attributable to different causes, such as intentional sabotage versus hardware or software failures, have different effects on trust. Thus, for some systems, it may be useful to discriminate between perturbations driven by enemy intent and undirected variations. These results suggest that trust is more than a simple reflection of the performance of the automation; appropriate trust depends on the operators' understanding of how the context affects the capability of the automation.

Effects of Display Content and Format on Trust

Trust evolves through analytic-, analogical-, and affect-based interpretations of information regarding the capabilities of the automation. Because direct observation of the automation is often impractical or impossible, perception of the automation-related information is usually mediated by a display. This suggests that the appropriateness of trust – that is, the match between the trust and the capabilities of the automation – depends on the content and format of the display. Information regarding the capabilities of

the automation can be defined along the dimensions of detail and abstraction. The dimension of *abstraction* refers to information regarding the performance, process, and purpose of the automation. The dimension of *detail* influences the functional specificity of the trust: whether trust is focused on the mode of the automatic controller, the automatic controller as a whole, or a group of automatic controllers. Organizing this information in the display in a way that supports analytic-, analogical-, and affect-based assimilation of this information may be an important means of guiding appropriate expectations regarding the automation. If the information is not available in the display or if it is formatted improperly, trust may not develop appropriately.

The content and format of the interface have a powerful effect on trust. Substantial research in this area has focused on trust in Internet-based interactions (for a review, see Corritore, Kracher, & Wiedenbeck, 2003a). In many cases, trust and credibility depend on surface features of the interface that have no obvious link to the true capabilities of the system (Briggs, Burford, & Dracup, 1998; Tseng & Fogg, 1999). In an on-line survey of more than 1400 people, Fogg, Marshall, Laraki, et al. (2001) found that for Web sites, credibility depends heavily on "real-world feel," which is defined by factors such as speed of response, listing a physical address, and including photos of the organization. Similarly, a formal photograph of the author enhanced trustworthiness of a research article, whereas an informal photograph decreased trust (Fogg, Marshall, Kameda, et al., 2001). Visual design factors of the interface, such as cool colors and a balanced layout, can also induce trust (Kim & Moon, 1998). Similarly, trusted Web sites tend to be text based, use empty space as a structural element, have strictly structured grouping, and use real photographs (Karvonen & Parkkinen, 2001).

These results show that trust tends to increase when information is displayed in a way that provides concrete details that are consistent and clearly organized. However, caution is needed when adding photographs and other features to enhance trust. These features might inadvertently degrade trust because they can undermine usability and can be considered manipulative by some (Riegelsberger & Sasse, 2002).

A similar pattern of results appears in studies of automation that supports target detection. Increasing image realism increased trust and led to greater reliance on the cueing information (Yeh & Wickens, 2001). Similarly, the tendency of pilots to blindly follow the advice of the system increased when the aid included detailed pictures as a guide (Ockerman, 1999). Just as highly realistic images can increase trust, degraded imagery can decrease trust, as was shown in a target cueing situation (MacMillan, Entin, & Serfaty, 1994). Adjusting image quality and adding information to the interface regarding the capability of the automation can promote appropriate trust. In a signal detection task, the reliability of the sources was coded with different levels of luminance, leading participants to weigh reliable sources more than unreliable ones (Montgomery & Sorkin, 1996). Likewise, increasing warning urgency increased compliance even when the participants were subject to a high rate of false alarms (Bliss et al., 1995). Although reliance and trust depend on the supporting information provided by the system, the amount of information must be tailored to the available decision time (Entin, Entin, & Serfaty, 1996). These results suggest that the particular interface form – in particular, the emphasis on concrete realistic representations – can increase the level of trust.

Similarly, trust between people builds rapidly with face-to-face communication but not with text-only communication. Interestingly, trust established in face-to-face communication transferred to subsequent text-only communication (Zheng, Bos, Olson, & Olson, 2001). Cassell and Bickmore (2000) suggested that creating a computer that is a conversational partner will induce people to trust the system by providing the same social cues that people use in face-to-face conversation. To maximize trust, designers should select speech parameters (e.g., words per minute, intonation, and frequency range) that create a personality consistent with the user and the content being presented (Nass & Lee, 2001). In addition, the speech should be consistent. A study of a speech interface showed that people were more trusting of a system that used synthetic speech consistently, as compared with one that used a combination of synthetic and human speech (Gong, Nass, Simard, & Takhteyev, 2001).

These results suggest that it may be possible to use degrees of synthetic speech to calibrate trust in a manner similar to the way degraded images can calibrate trust. However, care must be taken in adopting speech interactions because subtle differences may have important effects on trust and reliance. People may comply with command-based messages in a way that they might not with informational messages (Crocoll & Curry, 1990; Lee et al., 1999). More generally, using speech to create a conversational partner, as suggested by Cassell and Bickmore (2000), may lead people to attribute human characteristics to the automation in such a way as to induce false expectations that could lead to inappropriate trust.

In summary, the effect of subtle cues that are often associated with concrete examples and physical interactions may reflect a more general finding that specific instances and salient images invoke affective responses more powerfully than do abstract data (Finucane, Alhakami, Slovic, & Johnson, 2000; Loewenstein et al., 2001). Situations described in terms of frequency counts, vivid images, and specific instances lead to different assessments of risk as compared with when data are presented as context-free percentages and probabilities (Slovic, Finucane, Peters, & McGregor, in press). Affect-oriented display characteristics that influence risk judgments may have a similar power to influence trust.

IMPLICATIONS FOR CREATING TRUSTABLE AUTOMATION

Design errors, maintenance problems, and unanticipated variability make completely reliable and trustworthy automation unachievable. For this reason, creating highly trustable automation is important. Trustable automation supports adaptive reliance on the automation through highly calibrated, high-resolution, and high-specificity trust. Appropriate trust can lead to performance of the joint human-automation system that is superior to the performance of either the human or the automation alone (Sorkin & Woods, 1985; Wickens et al., 2000). The conceptual model in Figure 4 provides a theoretical basis to identify tentative design, evaluation, and training guidance and directions for future research that can lead to more trustable automation.

Make Automation Trustable

Appropriate trust and reliance depend on how well the capabilities of the automation are conveyed to the user. This can be done by making the algorithms of the automation simpler or by revealing their operation more clearly. Specific design, evaluation, and training considerations include the following:

- Design for appropriate trust, not greater trust.
- Show the past performance of the automation.
- Show the process and algorithms of the automation by revealing intermediate results in a way that is comprehensible to the operators.
- Simplify the algorithms and operation of the automation to make it more understandable.
- Show the purpose of the automation, design basis, and range of applications in a way that relates to the users' goals.
- Train operators regarding its expected reliability, the mechanisms governing its behavior, and its intended use.
- Carefully evaluate any anthropomorphizing of the automation, such as using speech to create a synthetic conversational partner, to ensure appropriate trust.

Substantial research issues face the design of trustable automation. First, little work has addressed how interface features influence affect. Recent neurological evidence suggests the existence of structurally separate pathways for affective and analytic responses, and this difference may imply equally separate interface considerations. For example, music and prosody may be particularly adept at influencing affective responses (Panksepp & Bernatzky, 2002), and so sonification (Brewster, 1997) rather than visualization may be an effective way to calibrate trust. Another important research issue is the trade-off between trustworthy automation and trustable automation. Trustworthy automation is automation that performs efficiently and reliably. Achieving this performance sometimes requires very complex algorithms that can be extremely hard to understand. To the extent that system performance depends on appropriate trust, there may be some circumstances in which making automation simpler but less capable outweighs the benefits of making it more complex and trustworthy but less

trustable. The trade-off between trustworthy and trustable automation and the degree to which interface design and training can mitigate this trade-off merit investigation.

Relate Context to Capability of the Automation

Appropriate trust depends on the operator's understanding of how the context affects the capability of the automation. Specific design, evaluation, and training considerations include the following:

- Reveal the context and support the assessment and classification of the situations relative to the capability of the automation.
- Show how past performance depends on situations.
- Consider how the context influences the relevance of trust; the final authority for some time-critical decisions should be allocated to automation.
- Evaluate the appropriateness of trust according to calibration, resolution, and specificity. Specificity and resolution are particularly critical when the performance of the automation is highly context dependent.
- Training to promote appropriate trust should address how situations interact with the characteristics of the automation to affect its capability.

Little research has addressed the challenges of promoting appropriate trust in the face of a dynamic context that influences its capability. The concept of ecological interface design has led to a substantial research effort that has focused on revealing the physical and intentional constraints of the system being controlled (Rasmussen & Vicente, 1989; Vicente, 2002; Vicente & Rasmussen, 1992). This approach could be viewed as conveying the context to the operator, but no research has considered how to relate this representation to one that describes the capabilities of the automation. Several researchers have acknowledged the importance of context in developing appropriate trust (Cohen et al., 1999; Masalonis, 2000), but few have considered the challenge of integrating ecological interface design with interface design for automation (Furukawa & Inagaki, 1999, 2001). Important issues regarding this integration include the links among the three levels of attributional abstraction (purpose, process, and performance) and the dimension of detail, which describe the characteristics of the automation and the abstraction hierarchy. Another important consideration is the link between the interface implications of skill-, rule-, and knowledge-based performance and those of the analytic, analogical, and affective processes that govern trust.

Consider Cultural, Organizational, and Team Interactions

Trust and reliance depend on organizational and cultural factors that go beyond the individual operator working with automation. Extrapolation from the relatively little research in human-automation interactions and the considerable research base on organization and interpersonal trust suggests the following design, evaluation, and training considerations:

- Consider the organizational context and the indirect exposure to automation that it facilitates.
- Consider individual and cultural differences in evaluations because they may influence reliance in ways that are not directly related to the characteristics of the automation.
- Cultural differences regarding expectations of the automation can be a powerful force that can lead to misuse and disuse unless addressed by appropriate training.
- Gossip and informal discussion of automation capabilities need to be considered in training and retraining so that the operators' understanding of the automation reflects its true capabilities.

Very little research has considered how individual and cultural differences influence trust and reliance. Of the few studies that have considered these issues, the findings suggest that individual and cultural differences can influence human-automation interactions in unexpected ways and merit further investigation, particularly in the context of extrapolating experimental data. Another important consideration is the role of team and organizational structure on the diffusion of trust among coworkers. Understanding how communication networks combine with the frequency of direct interaction with the automation and its reliability could have important implications for supporting appropriate trust. Research has not yet considered the evolution of trust in multiperson groups that share responsibility for managing automation. In this situation, people must develop appropriate trust not only in the automation but also in the other people who might assume manual control. Even more challenging is the development

of appropriate meta trust in the other people. *Meta trust* refers to the trust a person has that the other person's trust in the automation is appropriate. Multiperson, computer-mediated management of automation represents an unexplored area with considerable implications for many emerging systems.

CONCLUSION

"Emotion is critical for the appropriate direction of attention since it provides an automated signal about the organism's past experience with given objects and thus provides a basis for assigning or withholding attention relative to a given object" (Damasio, 1999, p. 273).

Trust influences reliance on automation; however, it does not determine reliance. As with other psychological constructs that explain behavior, it is important to identify the boundary conditions within which trust makes a difference. For trust, the boundary conditions include situations in which uncertainty and complexity make an exhaustive evaluation of options impractical. Trust is likely to influence reliance on complex, imperfect automation in dynamic environments that require the person to adapt to unanticipated circumstances. More generally, trust seems to be an example of how affect can guide behavior when rules fail to apply and when cognitive resources are not available to support a calculated rational choice (Barley, 1988). The conceptual model developed in this paper identifies some of the factors that interact with trust to influence reliance. By helping to define boundary conditions regarding the influence of trust, this model may lead to more precise hypotheses, more revealing measures of trust, and design approaches that promote more appropriate trust.

Trust is one example of the important influence of affect and emotions on human-technology interaction. Emotional response to technology is not only important for acceptance, it can also make a fundamental contribution to safety and performance. This is not a new realization; as Neisser stated, "human thinking begins in an intimate association with emotions and feelings which is never entirely lost" (as quoted in Simon, 1967, p. 29). However, little systematic research has addressed how these considerations should influence design and modeling of human-technology interaction. Trust is one example of how affect-related considerations should influence the design of complex, high-consequence systems.

Because automation and computer technology are growing increasingly complex, the importance of affect and trust is likely to grow. Computer technology allows people to develop relationships and collaborate with little or no direct contact, but these new opportunities produce complex situations that require appropriate trust. As computer technology grows more pervasive, trust is also likely to become a critical factor in consumer products, such as home automation, personal robots, and automotive automation. Designing trustable technology may be a critical factor in the success of the next generation of automation and computer technology.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant No. 0117494. This manuscript was greatly improved by the comments on earlier versions from Amy Bisantz, John Hajdukiewicz, Jenna Hetland, Greg Jamieson, Robin Mitchell, Raja Parasuraman, Chris Miller, Michelle Reyes, Kim Vicente, Christopher Wickens, and an anonymous reviewer. Special thanks to Neville Moray for his inspiration and insight.

REFERENCES

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Upper Saddle River, NJ: Prentice Hall.

Baba, M. L. (1999). Dangerous liaisons: Trust, distrust, and information technology in American work organizations. *Human Organization, 58,* 331–346.

Baba, M. L., Falkenburg, D. R., & Hill, D. H. (1996). Technology management and American culture: Implications for business process redesign. *Research Technology Management, 39*(6), 44–54.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37,* 122–147.

Barber, B. (1983). *The logic and limits of trust.* New Brunswick, NJ: Rutgers University Press.

Barley, S. R. (1988). The social construction of a machine: Ritual, superstition, magical thinking and other pragmatic responses to running a CT scanner. In M. Lock & D. R. Gordon (Eds.), *Biomedicine examined* (pp. 497–539). Dordrecht, Netherlands: Kluwer Academic.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50,* 7–15.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience, 18,* 428–437.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*(5304), 1293–1295.

Beer, R. D. (2001). Dynamical approaches in cognitive science. *Trends in Cognitive Sciences, 4,* 91–99.

Berscheid, E. (1994). Interpersonal relationships. *Annual Review of Psychology, 45,* 79–129.

Bhattacharya, R., Devinney, T. M., & Pillutla, M. M. (1998). A formal model of trust based on outcomes. *Academy of Management Review, 23,* 459–472.

Bisantz, A. M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics, 28*(2), 85–97.

Bliss, J., Dunn, M., & Fuller, B. S. (1995). Reversal of the cry-wolf effect – An investigation of two methods to increase alarm response rates. *Perceptual and Motor Skills, 80,* 1231–1242.

Brehmer, B., & Dorner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field-study. *Computers in Human Behavior, 9,* 171–184.

Brewster, S. A. (1997). Using non-speech sound to overcome information overload. *Displays, 17,* 179–189.

Briggs, P., Burford, B., & Dracup, C. (1998). Modeling self-confidence in users of a computer-based system showing unrepresentative design. *International Journal of Human-Computer Studies, 49,* 717–742.

Burt, R. S., & Knez, M. (1996). Trust and third-party gossip. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 68–89). Thousand Oaks, CA: Sage.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognitive approach to decision making in an uncertain environment. *Psychological Review, 100,* 432–459.

Butler, J. K., & Cantrell, R. S. (1984). A behavioral decision theory approach to modeling trust in superiors and subordinates. *Psychological Reports, 55,* 19–28.

Case, K., Sinclair, M. A., & Rani, M. R. A. (1999). An experimental investigation of human mismatches in machining. *Proceedings of the Institution of Mechanical Engineers Part B – Journal of Engineering Manufacture, 213,* 197–201.

Cassell, J., & Bickmore, T. (2000). External manifestations of trustworthiness in the interface. *Communications of the ACM, 43*(12), 50–56.

Castelfranchi, C. (1998). Modelling social action for AI agents. *Artificial Intelligence, 103,* 157–182.

Castelfranchi, C., & Falcone, R. (2000). Trust and control: A dialectic link. *Applied Artificial Intelligence, 14,* 799–823.

Cohen, M. S. (2000). *Training for trust in decision aids, with applications to the rotorcraft's pilot's associate.* Presented at the International Conference on Human Performance, Situation Awareness, and Automation, Savannah, GA. Abstract retrieved March 8, 2004, from http://www.cog-tech.com/papers/Trust/Adaptive%20Automation%20abstract%20final.pdf

Cohen, M. S., Parasuraman, R., & Freeman, J. (1999). *Trust in decision aids: A model and its training implications* (Tech. Report USAATCOM TR 97-D-4). Arlington, VA: Cognitive Technologies.

Conejo, R. A., & Wickens, C. D. (1997). *The effects of highlighting validity and feature type on air-to-ground target acquisition performance* (ARL-97-11/NAWC-ONR-97-1). Savoy, IL: Aviation Research Laboratory Institute for Aviation.

Cook, J., & Wall, T. (1980). New work attitude measures of trust, organizational commitment and personal need non-fulfillment. *Journal of Occupational Psychology, 53,* 39–52.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2001a, March–April). *Trust in the online environment.* Workshop presented at the CHI 2001 Conference on Human Factors in Computing Systems, Seattle, WA.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2001b). Trust in the online environment. In M. J. Smith, G. Salvendy, D. Harris, & R. J. Koubek (Eds.), *Usability evaluation and interface design: Cognitive engineering, intelligent agents and virtual reality* (pp. 1548-1552). Mahwah, NJ: Erlbaum.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003a). On-line trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies, 58,* 737–758.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (Eds.). (2003b). Trust and technology [Special issue]. *International Journal of Human-Computer Studies, 58*(6).

Couch, L. L., & Jones, W. H. (1997). Measuring levels of trust. *Journal of Research in Personality, 31*(3), 319–336.

Coutu, D. L. (1998). Trust in virtual teams. *Harvard Business Review, 76,* 20–21.

Crocoll, W. M., & Coury, B. G. (1990). Status or recommendation: Selecting the type of information for decision aiding. In *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 1524–1528). Santa Monica, CA: Human Factors and Ergonomics Society.

Crossman, E. R. F. W. (1974). Automation and skill. In E. Edwards & F. P. Lees (Eds.), *The human operator in process control* (pp. 1–24). London: Taylor & Francis.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences, 351,* 1413–1420.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness.* New York: Harcourt.

Damasio, A. R. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain.* New York: Harcourt.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social-stimuli. *Behavioural Brain Research, 41,* 81–94.

Dassonville, I., Jolly, D., & Desodt, A. M. (1996). Trust between man and machine in a teleoperation system. *Reliability Engineering and System Safety, 53,* 319–325.

Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution, 2,* 265–279.

Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations, 13,* 123–139.

Doney, P. M., Cannon, J. P., & Mullen, M. R. (1998). Understanding the influence of national culture on the development of trust. *Academy of Management Review, 23,* 601–620.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors, 44,* 79–94.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology, 13,* 147–164.

Entin, E. B., Entin, E. E., & Serfaty, D. (1996). Optimizing aided target-recognition performance. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 233–237). Santa Monica, CA: Human Factors and Ergonomics Society.

Farrell, S., & Lewandowsky, S. (2000). A connectionist model of complacency and adaptive recovery under automation. *Journal of Experimental Psychology – Learning, Memory, and Cognition, 26,* 395–410.

Fine, G. A., & Holyfield, L. (1996). Secrecy, trust, and dangerous leisure: Generating group cohesion in voluntary organizations. *Social Psychology Quarterly, 59,* 22–38.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13,* 1–17.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior.* Reading, MA: Addison-Wesley.

Fogg, B., Marshall, J., Kameda, T., Solomon, J., Rangnekar, A., Boyd, J., et al. (2001). Web credibility research: A method for online experiments and early study results. In *CHI 2001 Conference on Human Factors in Computing Systems* (pp. 293–294). New York: Association for Computing Machinery.

Fogg, B., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., et al. (2001). *What makes Web sites credible? A report on a large quantitative study.* In *CHI 2001 Conference on Human Factors in Computing Systems* (pp. 61–68). New York: Association for Computing Machinery.

Fox, J. M. (1996). Effects of information accuracy on user trust and compliance. In *CHI 1996 Conference on Human Factors in Computing Systems* (pp. 35-36). New York: Association for Computing Machinery.

Fox, J. M., & Boehm-Davis, D. A. (1998). Effects of age and congestion information accuracy of advanced traveler information on user trust and compliance. *Transportation Research Record, 1621,* 43–49.

Furukawa, H., & Inagaki, T. (1999). Situation-adaptive interface based on abstraction hierarchies with an updating mechanism for maintaining situation awareness of plant operators. In *Proceedings of the 1999 IEEE International Conference on*

Systems, Man, and Cybernetics (pp. 693–698). Piscataway, NJ: IEEE.

Furukawa, H., & Inagaki, T. (2001). A graphical interface design for supporting human-machine cooperation: Representing automation's intentions by means-ends relations. In *Proceedings of the 8th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems* (pp. 573–578). Laxenburg, Austria: International Federation of Automatic Control.

Gabarro, J. J. (1978). The development of trust influence and expectations. In A. G. Athos & J. J. Gabarro (Eds.), *Interpersonal behavior: Communication and understanding in relationships* (pp. 290–230). Englewood Cliffs, NJ: Prentice Hall.

Gaines, S. O., Panter, A. T., Lyde, M. D., Steers, W. N., Rusbult, C. E., Cox, C. L., et al. (1997). Evaluating the circumplexity of interpersonal traits and the manifestation of interpersonal traits in interpersonal trust. *Journal of Personality and Social Psychology, 73,* 610–623.

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy – A theoretical analysis of its determinants and malleability. *Academy of Management Review, 17,* 183–211.

Gong, L., Nass, C., Simard, C., & Takhteyev, Y. (2001). When non-human is better than semi-human: Consistency in speech interfaces. In M. J. Smith, G. Salvendy, D. Harris, & R. Koubek (Eds.), *Usability evaluation and interface design: Cognitive engineering, intelligent agents, and virtual reality* (pp. 1558–1562). Mahwah, NJ: Erlbaum.

Grice, H. P. (Ed.). (1975). *Logic and conversation: Vol. 3. Speech acts.* New York: Academic.

Gurtman, M. B. (1992). Trust, distrust, and interpersonal problems: A circumplex analysis. *Journal of Personality and Social Psychology, 62,* 989–1002.

Halprin, S., Johnson, E., & Thornburry, J. (1973). Cognitive reliability in manned systems. *IEEE Transactions on Reliability, R-22,* 165-169.

Hardin, R. (1992). The street-level epistemology of trust. *Politics and Society, 21,* 505–529.

Hoc, J. M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics, 43,* 833–843.

Holmes, J. G. (1991). Trust and the appraisal process in close relationships. In W. H. Jones & D. Perlman (Eds.), *Advances in personal relationships* (Vol. 2, pp. 57-104). London: Jessica Kingsley.

Hovland, C. I., Janis, I. L., & Kelly, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change.* New Haven, CT: Yale University Press.

Huang, H., Keser, C., Leland, J., & Shachat, J. (2002). *Trust, the Internet and the digital divide* (RC22511). Yorktown Heights, NY: IBM Research Division.

Inagaki, T., Takae, Y., & Moray, N. (1999). Decision supported information for takeoff safety in human-centered automation: An experimental investigation of time-fragile characteristics. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* (Vol. 1, pp. 1101–1106). Piscataway, NJ: IEEE.

Itoh, M., Abe, G., & Tanaka, K. (1999). Trust in and use of automation: Their dependence on occurrence patterns of malfunctions. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* (Vol. 3, pp. 715–720). Piscataway, NJ: IEEE.

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment.* New York: Free.

Jennings, E. E. (1967). *The mobile manager: A study of the new generation of top executives.* Ann Arbor, MI: Bureau of Industrial Relations, University of Michigan.

Johns, J. L. (1996). A concept analysis of trust. *Journal of Advanced Nursing, 24,* 76–83.

Johnson-Laird, P. N., & Oately, K. (1992). Basic emotions, rationality, and folk theory. *Cognition and Emotion, 6,* 201–223.

Jones, G. R., & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review, 23,* 531–546.

Kantowitz, B. H., Hanowski, R. J., & Kantowitz, S. C. (1997a). Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors, 39,* 164–176.

Kantowitz, B. H., Hanowski, R. J., & Kantowitz, S. C. (1997b). Driver reliability requirements for traffic advisory information. In Y. I. Noy (Ed.), *Ergonomics and safety of intelligent driver interfaces* (pp. 1–22). Mahwah, NJ: Erlbaum.

Karvonen, K. (2001). Designing trust for a universal audience: A multicultural study on the formation of trust in the internet in the Nordic countries. In C. Stephanidis (Ed.), *First International Conference on Universal Access in Human-Computer Interaction* (Vol. 3, pp. 1078–1082). Mahwah, NJ: Erlbaum.

Karvonen, K., & Parkkinen, J. (2001). Signs of trust: A semiotic study of trust formation in the Web. In M. J. Smith, G. Salvendy, D. Harris, & R. J. Koubek (Eds.), *First International Conference on Universal Access in Human-Computer Interaction* (Vol. 1, pp. 1076–1080). Mahwah, NJ: Erlbaum.

Kee, H. W., & Knox, R. E. (1970). Conceptual and methodological considerations in the study of trust and suspicion. *Conflict Resolution, 14,* 357–366.

Kelso, J. A. S. (1995). *Dynamic patterns: Self-organization of brain and behavior.* Cambridge, MA: MIT Press.

Kikuchi, M., Wantanabe, Y., & Yamasishi, T. (1996). Judgment accuracy of other's trustworthiness and general trust: An experimental study. *Japanese Journal of Experimental Social Psychology, 37,* 23–36.

Kim, J., & Moon, J. Y. (1998). Designing towards emotional usability in customer interfaces – Trustworthiness of cyber-banking system interfaces. *Interacting With Computers, 10,* 1–29.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an "aid" can (and should) go unused. *Human Factors, 35,* 221–242.

Klein, G. A. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in man-machine system research* (Vol. 5, pp. 47–92). Greenwich, CT: JAI.

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology, 50,* 569–598.

Kramer, R. M., Brewer, M. B., & Hanna, B. A. (1996). Collective trust and collective action: The decision to trust as a social decision. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 357–389). Thousand Oaks, CA: Sage.

Kramer, R. M., & Tyler, T. R. (Eds.). (1996). *Trust in organizations: Frontiers of theory and research.* Thousand Oaks, CA: Sage.

Lee, J. D., Gore, B. F., & Campbell, J. L. (1999). Display alternatives for in-vehicle warning and sign information: Message style, location, and modality. *Transportation Human Factors, 1,* 347–377.

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35,* 1243–1270.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40,* 153–184.

Lee, J. D., & Sanquist, T. F. (1996). Maritime automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance* (pp. 365–384). Mahwah, NJ: Erlbaum.

Lee, J. D., & Sanquist, T. F. (2000). Augmenting the operator function model with cognitive operations: Assessing the cognitive demands of technological innovation in ship navigation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 30,* 273–285.

Lee, M. K. O., & Turban, E. (2001). A trust model for consumer Internet shopping. *International Journal of Electronic Commerce, 6*(1), 75–91.

Lewandowsky, S., Mundy, M., & Tan, G. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology – Applied, 6,* 104–123.

Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 114-139). Thousand Oaks, CA: Sage.

Lewis, M. (1998). Designing for human-agent interaction. *AI Magazine, 19*(2), 67–78.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York: Cambridge University Press.

Liu, Y., Fuld, R., & Wickens, C. D. (1993). Monitoring behavior in manual and automated scheduling systems. *International Journal of Man-Machine Studies, 39,* 1015–1029.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127,* 267–286.

MacMillan, J., Entin, E. B., & Serfaty, D. (1994). Operator reliance on automated support for target acquisition. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 1285–1289). Santa Monica, CA: Human Factors and Ergonomics Society.

Malhotra, D., & Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly, 47,* 534–559.

Masalonis, A. J. (2000). *Effects of situation-specific reliability on trust and usage of automated decision aids.* Unpublished Ph.D. dissertation, Catholic University of America, Washington, DC.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20,* 709–734.

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review, 23,* 473–490.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors, 43,* 563–572.

Meyerson, D., Weick, K. E., & Kramer, R. M. (1996). Swift trust and temporary groups. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 166–195). Thousand Oaks, CA: Sage.

Milewski, A. E., & Lewis, S. H. (1997). Delegating to software agents. *International Journal of Human-Computer Studies, 46,* 485–500.

Miller, C. A. (2002). Definitions and dimensions of etiquette. In C. Miller (Ed.), *Etiquette for human-computer work* (Tech. Report FS-02-02, pp. 1–7). Menlo Park, CA: American Association for Artificial Intelligence.

Mishra, A. K. (1996). Organizational response to crisis. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 261–287). Thousand Oaks, CA: Sage.

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors, 38,* 311–322.

Montgomery, D. A., & Sorkin, R. D. (1996). Observer sensitivity to element reliability in a multielement visual display. *Human Factors, 38,* 484–494.

Moorman, C., Deshpande, R., & Zaltman, G. (1993). Factors affecting trust in market-research relationships. *Journal of Marketing, 57*(1), 81–101.

Moray, N. (2003). Monitoring, complacency, scepticism and eutactic behaviour. *International Journal of Industrial Ergonomics, 31,* 175–178.

Moray, N., & Inagaki, T. (1999). Laboratory studies of trust between humans and machines in automated systems. *Transactions of the Institute of Measurement and Control, 21,* 203–211.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology – Applied, 6,* 44–58.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing, 58*(3), 20–38.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology, 8,* 47–63.

Mosier, K. L., Skitka, L. J., & Korte, K. J. (1994). Cognitive and social issues in flight crew/automation interaction. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 191–197). Hillsdale, NJ: Erlbaum.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aides. *International Journal of Man-Machine Studies, 27,* 527–539.

Muir, B. M. (1989). *Operators' trust in and use of automatic controllers in a supervisory process control task.* Unpublished Ph.D. dissertation, University of Toronto, Toronto, Canada.

Muir, B. M. (1994). Trust in automation: 1. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37,* 1905–1922.

Muir, B. M., & Moray, N. (1996). Trust in automation: 2. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39,* 429–460.

Muller, G. (1996). Secure communication – Trust in technology or trust with technology? *Interdisciplinary Science Reviews, 21,* 336–347.

Nass, C., & Lee, K. N. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology – Applied, 7,* 171–181.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56,* 81–103.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies, 43,* 223–239.

National Transportation Safety Board. (1997). *Marine accident report – Grounding of the Panamanian passenger ship* Royal Majesty *on Rose and Crown Shoal near Nantucket, Massachusetts, June 10, 1995* (NTSB/MAR97/01). Washington, DC: Author.

Norman, D. A., Ortony, A., & Russell, D. M. (2003). Affect and machine design: Lessons for the development of autonomous machines. *IBM Systems Journal, 42,* 38–44.

Nyhan, R. C. (2000). Changing the paradigm – Trust and its role in public sector organizations. *American Review of Public Administration, 30,* 87–109.

Ockerman, J. J. (1999). Over-reliance issues with task-guidance systems. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 1192–1196). Santa Monica, CA: Human Factors and Ergonomics Society.

Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action. *American Political Science Review, 92,* 1–22.

Pally, R. (1998). Emotional processing: The mind-body connection. *International Journal of Psycho-Analysis, 79,* 349–362.

Panksepp, J., & Bernatzky, G. (2002). Emotional sounds and the brain: The neuro-affective foundations of musical appreciation. *Behavioural Processes, 60,* 133–155.

Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance consequences of automation-induced "complacency." *International Journal of Aviation Psychology, 3,* 1–23.

Parasuraman, R., Mouloua, M., & Molloy, R. (1994). Monitoring automation failures in human-machine systems. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 45–49). Hillsdale, NJ: Erlbaum.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38,* 665–679.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (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,* 286–297.

Parasuraman, S., Singh, I. L., Molloy, R., & Parasuraman, R. (1992). Automation-related complacency: A source of vulnerability in contemporary organizations. *IFIP Transactions A – Computer Science and Technology, 13,* 426–432.

Picard, R. W. (1997). *Affective computing.* Cambridge, MA: MIT Press.

Rasmussen, J. (1983). Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13,* 257–266.

Rasmussen, J., Pejterson, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering.* New York: Wiley.

Rasmussen, J., & Vicente, K. J. (1989). Coping with human errors through system design: Implications for ecological interface design. *International Journal of Man-Machine Studies, 31,* 517–534.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* New York: Cambridge University Press.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49*(1), 95–112.

Riegelsberger, J., & Sasse, M. A. (2002). Face it – Photos don't make a Web site trustworthy. In *CHI 2002 Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 742–743). New York: Association for Computing Machinery.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation theory and applications* (pp. 19–35). Mahwah, NJ: Erlbaum.

Riley, V. A. (1994). *Human use of automation.* Unpublished Ph.D. dissertation, University of Minnesota, Minneapolis, MN.
Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron, 35,* 395–405.
Ring, P. S., & Vandeven, A. H. (1992). Structuring cooperative relationships between organizations. *Strategic Management Journal, 13,* 483–498.
Ross, W., & LaCroix, J. (1996). Multiple meanings of trust in negotiation theory and research: A literature review and integrative model. *International Journal of Conflict Management, 7,* 314–360.
Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35,* 651–665.
Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26,* 443–452.
Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist, 35,* 1–7.
Rousseau, D., Sitkin, S., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23,* 393–404.
Sato, Y. (1999). Trust and communication. *Sociological Theory and Methods, 13,* 155–168.
Schindler, P. L., & Thomas, C. C. (1993). The structure of interpersonal trust in the workplace. *Psychological Reports, 73,* 563–573.
Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition and Emotion, 14,* 433–440.
Shaw, R., & Kinsella-Shaw, J. (1988). Ecological mechanics: A physical geometry for intentional constraints. *Human Movement Science, 7,* 155–200.
Sheridan, T. B. (1975). Considerations in modeling the human supervisory controller. In *Proceedings of the IFAC 6th World Congress* (pp. 1-6). Laxenburg, Austria: International Federation of Automatic Control.
Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control.* Cambridge, MA: MIT Press.
Sheridan, T. B., & Hennessy, R. T. (1984). *Research and modeling of supervisory control behavior.* Washington, DC: National Academy.
Simon, H. (1957). *Models of man.* New York: Wiley.
Simon, H. (1967). Motivational and emotional controls of cognition. *Psychological Review, 74,* 29–39.
Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Individual differences in monitoring failures of automation. *Journal of General Psychology, 120,* 357–373.
Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science, 4,* 367–392.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22.
Slovic, P., Finucane, M. L., Peters, E., & McGregor, D. G. (in press). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis.*
Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human-Computer Interaction, 1,* 49–75.
Sparaco, P. (1995, January 30). Airbus seeks to keep pilot, new technology in harmony. *Aviation Week and Space Technology,* pp. 62–63.
Stack, L. (1978). Trust. In H. London & J. E. Exner, Jr. (Eds.), *Dimensions of personality* (pp. 561–599). New York: Wiley.
Tan, H. H., & Tan, C. S. (2000). Toward the differentiation of trust in supervisor and trust in organization. *Genetic, Social, and General Psychological Monographs, 126,* 241–260.
Tenney, Y. J., Rogers, W. H., & Pew, R. W. (1998). Pilot opinions on cockpit automation issues. *International Journal of Aviation Psychology, 8,* 103–120.
Thackray, R. I., & Touchstone, R. M. (1989). Detection efficiency on an air traffic control monitoring task with and without computer aiding. *Aviation, Space, and Environmental Medicine, 60,* 744–748.
Thelen, E., Schoner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perservative reaching. *Behavior and Brain Sciences, 24,* 1–86.
Trentesaux, D., Moray, N., & Tahon, C. (1998). Integration of the human operator into responsive discrete production management systems. *European Journal of Operational Research, 109,* 342–361.
Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. *Communications of the ACM, 42*(5), 39–44.
van Gelder, T., & Port, R. E. (1995). It's about time: An overview of the dynamical approach to cognition. In R. F. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 1–43). Cambridge, MA: MIT Press.
Van House, N. A., Butler, M. H., & Schiff, L. R. (1998). Cooperating knowledge, work, and trust: Sharing environmental planning data sets. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 335–343). New York: Association for Computing Machinery.
Vicente, K. J. (1999). *Cognitive work analysis: Towards safe, productive, and healthy computer-based work.* Mahwah, NJ: Erlbaum.
Vicente, K. J. (2002). Ecological interface design: Progress and challenges. *Human Factors, 44,* 62–78.
Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-22,* 589–606.
Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors, 2,* 99–126.
Wicks, A. C., Berman, S. L., & Jones, T. M. (1999). The structure of optimal trust: Moral and strategic. *Academy of Management Review, 24,* 99–116.
Williamson, O. (1993). Calculativeness, trust and economic organization. *Journal of Law and Economics, 36,* 453–486.
Wright, G., Rowe, G., Bolger, F., & Gammack, J. (1994). Coherence, calibration, and expertise in judgmental probability forecasting. *Organizational Behavior and Human Decision Processes, 57,* 1–25.
Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18,* 129–166.
Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors, 43,* 355–365.
Zheng, J., Bos, N., Olson, J. S., & Olson, G. M. (2001). Trust without touch: Jump-start trust with social chat. In *CHI 2001 Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 293–294). New York: Association for Computing Machinery.
Zuboff, S. (1988). *In the age of smart machines: The future of work technology and power.* New York: Basic Books.

John D. Lee is an associate professor of industrial engineering and director of the Cognitive Systems Laboratory at the University of Iowa. He received his Ph.D. in mechanical engineering in 1992 from the University of Illinois at Urbana-Champaign.

Katrina A. See is a human-systems analyst with Aptima in Woburn, Massachusetts. She received an M.S. in industrial engineering, with a focus on trust in automation, from the University of Iowa in 2002.

Date received: August 5, 2002
Date accepted: September 4, 2003