Trust Calibration for Automated Decision Aids

Project Leads

Maranda McBride, PhD, North Carolina A&T State University
Shona Morgan, PhD, North Carolina A&T State University

Statement of Problem

Given the variety of complex situations that arise in the context of homeland security where uncertainty and vulnerability persist, it is essential that measures be taken to enhance the safety of U.S. citizens. Kaplan (2007) describes various new technological devices likely to be developed to meet the needs of this “war on terror” era. Such devices include data mining technology, communication systems, hazard detection devices, command and control systems, screening technologies, and biometric identification systems. All of these devices will rely on some form of automation and are designed to expedite the decision making process; thus, it is suitable to refer to them as automated decision aids (ADAs). ADAs of this type are generally classified as level 3, 4, or 5 automation since they are typically designed to assist with the assessment of information and not execute certain courses of action without human intervention (Parasuraman, Sheridan, & Wickens, 2000).

The intent of ADAs is to enable users to make timely decisions by providing pertinent information in a more efficient manner than a human being working alone can achieve. Human decision making supported by the use of ADAs is expected to result in higher quality decisions (McGuirl & Sarter, 2006); however, in order for ADAs to be effective, decision makers (DMs) must consider the information provided by such systems to be trustworthy and reliable.
Technological devices that are trustworthy and reliable have a greater likelihood of being accepted by users than those that are not.

Unfortunately, no matter how robust the design, it is likely that ADA software is going to fall short of expectations at some time. Such shortfalls often occur when users misunderstand the capabilities of the ADA or when the ADA has become the subject of hacking or some other form of technological sabotage. In such cases, DMs begin to view the ADA as ineffective and develop a level of distrust in the system. In some instances, DMs will even choose to reject the automated capabilities at their disposal and rely solely on their own abilities, which is likely to lead to time-consuming, inefficient, and unproductive processes.

While occasional automation failures may result in system “disuse” (Parasuraman & Riley, 1997; Yeh & Wickens, 2001), positive experiences with ADAs may cause some DMs to rely too heavily on automated systems. These individuals can become overly confident in the capabilities of the system and judge periodic verification and validation of the information being provided by the ADA unnecessary. This, in turn, can lead to “misuse” of ADAs whereby DMs rely on the system to perform tasks that exceed its capabilities or for which it was not designed (Parasuraman & Riley, 1997).

Cases of disuse and misuse often occur when a user’s trust in a system is poorly calibrated. “Calibration” is a term used to describe the process by which automated system users learn to adjust their behavior based upon the specific characteristics (e.g., performance) of the system. When trust is miscalibrated, the perceived and actual performance of the system are not in proper alignment with one another (McGuirl & Sarter, 2006).

Certainly there will continue to be a pressing need for new technologies designed to assist with decision making tasks, especially in high-risk fields such as homeland security. Therefore, it is imperative that DMs’ trust be calibrated so that they effectively use the ADAs at their disposal. DMs with properly calibrated trust are essential in order to prevent many of the adverse consequences associated with both automation disuse and misuse. Thus one of the primary research questions to be investigated is, What are the most effective methods of ADA trust calibration?

---

**Background**

Studies over the past 20 years have provided compelling theoretical and experimental evidence supporting the importance of trust in human-machine systems (Lee & Moray, 1992, 1994; Muir, 1987, 1994; Muir & Moray, 1996). Rapid increases in ADA development have identified two extreme polarities in their use. On one end, some DMs refuse to use ADAs to improve their performance on certain tasks while on the opposite end other DMs tend to demonstrate an exceedingly high level of dependence on these systems to perform tasks. These behaviors are due in part to the degree of trust that users have in the capabilities of
ADAs, whereby a lack of trust may lead to the disuse of ADAs, while excessive trust may lead to their misuse (Lee, 2008; Lee & See, 2004; Parasuraman & Riley, 1997).

Several researchers purport that many of the influences on trust between humans are applicable to trust between humans and machines (Atoyan, Duquet, & Robert, 2006; Muir & Moray, 1996; Nass & Lee, 2001; Nass, Moon, Fogg, Reeves, & Dryer, 1995; Nickerson & Reilly, 2004; Parasuraman, Molloy, & Singh, 1993; Reeves & Nass, 1996; Sheridan, 1975; Sheridan & Hennessy, 1984). For example, just as people generally limit their interaction with untrustworthy people, they also have a tendency to avoid interacting with or relying on machines (e.g., ADAs) that they do not trust (Lee & See, 2004; Muir, 1987). In a two-party relationship, a person’s trust is directly influenced by past experiences or interactions with the other party (Rempel, Holmes, & Zanna, 1985). When it comes to the development of human-machine trust, experiences are used to determine the level of faith that users place in a machine’s ability to perform its tasks and whether or not it is safe to rely on the machine. For example, in a study of civil aviation pilots conducted by Dusire and Falzon (1999), trust was strongly correlated with control actions based on the information provided by automated devices. Studies of air traffic controller trust in conflict probe automation conducted by Masalonis et al. (1998) resulted in similar findings.

Fan, Hyams, and Kuchar (1998) explored the issue of trust as a human’s willingness to accept direction from an automated system. In their study of the use of an in-flight replanning aid, they determined that pilots were more willing to follow the direction of the automated system if it was accompanied by supplemental information that could be used to validate the decision. This study demonstrated that during the initial stages of trust development, users often rely on additional sources to substantiate information provided by the system; however, the need for verification diminishes as confidence in the system increases.

Maes (1994) presented a model of an informal “testing” approach to trust. In this model, as the user spends more time with an ADA, its actions become more predictable and the degree of trust increases. This supports the theory that trust is developed over time through mutually satisfying interactions between two parties. However, if trust is ever breached, confidence may be eroded and trust will need to be regained. In the case of human-machine systems, the two most common means of regaining trust is through consistently good performance over time (Kantowitz, Hanowski, & Kantowitz, 1997) or by making the same error consistently so that the user can predict and acclimate to it (Muir & Moray, 1996).

Sheridan (1988) identified seven ADA design characteristics humans use to assign trust: reliability, robustness, familiarity, understandability, usefulness, explication of intention, and dependability. Ease of use is another design attribute which impacts trust and acceptance of ADAs. One method that interface designers are increasingly using to improve ease of use is the integration of anthropomorphic attributes. These attributes create more “natural” interactions intended to elicit user trust and increase system acceptance (Marsh & Meech, 2000).
The Technology Acceptance Model (TAM) incorporates ease of use along with two other trust characteristics—perceived usefulness and behavioral intention—to determine whether or not a user will accept an automated system (Davis, Bagozzi, & Warshaw, 1989; Davis, 1989; Mathieson, Peacock, & Chin, 2001). This model has been employed by several researchers investigating the use of information technology (e.g., Adams, Nelson, & Todd, 1992; Bahmanziari, Pearson, & Crosby, 2003; Hu, Lin, & Chen, 2005; Taylor & Todd, 1995). The list of system design characteristics believed to impact user trust also includes system integrity, level of security (Jian, Bisantz, & Drury, 2000), and the level of automation (Parasuraman, Sheridan, & Wickens, 2000). Based upon the characteristics listed above, distrust of an ADA may be the result of inconsistent or unstable system performance such as failures, errors (de Vries, Midden, & Bouwhuis, 2003; Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Johnson, 2004; Lee & Moray, 1992; Moray, Inagaki, & Itoh, 2000; Wiegmann, Rich, & Zhang, 2001), or poor user feedback; a lack of understanding of the ADA software capabilities; or even poor interaction between the user and ADA interface (Friedman & Kahn, 1997).

Researchers (e.g., Atoyan, Duquet, & Robert, 2006; Lee & See, 2004; Merritt & Ilgen, 2008; Nickerson & Reilly, 2004; van Dongen & van Maanen, 2005) have also suggested that system characteristics are not the only things that impact a user’s trust; certain user characteristics also affect people’s assignment of trust to automated systems. For instance, it has been found that extroverts are typically more willing to trust other people than introverts (Gaines et al., 1997; Omodei & McLennan, 2000; Shikishima, Hiraishi, & Ando, 2006). They also tend to demonstrate a stronger propensity to trust overall than do introverts. As a result, extroverts’ initial trust in a system may be greater than introverts’; however, extroverts’ trust level is also likely to decline more rapidly if their expectations are not met (Merritt & Ilgen, 2008).

Another individual influence which was identified by Tseng and Fogg (1999) is the user’s level of expertise. Typically, users who are highly competent in a task area are less likely to blindly trust in an ADA than are novices who tend to have a greater need for the information provided by the ADA and are more likely to accept it. Other individual user differences which may affect trust in ADAs and in turn affect their use include the user’s age (Wiegmann, McCarley, Kramer, & Wickens, 2006), level of self-confidence (de Vries et al., 2003; Lee & Moray, 1994), and perceived usefulness of the system (Dzindolet, Pierce, Beck, & Dawe, 2002).

Given the impact of trust on technology adoption, the increase in the amount of research on methods of trust calibration is not surprising. One component of trust calibration deemed significant by researchers is that users understand the conditions under which their systems perform optimally (Cohen, Parasuraman, & Freeman, 1998). Given this understanding, the appropriate degree of trust can develop. For example, one study conducted by Sorkin, Kantowitz, and Kantowitz (1988) found that providing information on the diagnostic capabilities of a binary alarm system helped users allocate the appropriate level of trust in a system.
McGuirl and Sarter (2006) discovered that providing dynamic system confidence information improved trust calibration for pilots engaged in flight tasks and helped them properly allocate tasks. The discovery of individual differences in how humans allocate trust to automated systems prompted Merritt and Ilgen (2008) to suggest that users first be assessed for introversion/extroversion characteristics and then assigned to computer-based training which caters specifically to their personality type to help establish the proper level of historical-based trust. These methods are just a few that have been investigated to date.

Several other factors that affect the efficient use of ADAs relate more to the implementation process than to user trust. Although this document does not focus on the implementation of new technology, the authors would be remiss if they did not discuss key barriers to implementation; however, the following review of the literature related to these factors falls short of being comprehensive.

Klein and Sorra (1996) describe several implementation barriers, including what they refer to as implementation climate and innovation-values fit. Implementation climate involves organizational policies and practices such as the availability of training (Edmondson, Bohmer, & Pisano, 2001; Fleischer, Liker, & Arnsdorf, 1988), support services (Mathieson et al., 2001; Rousseau, 1989), user incentive programs (Klein, Hall, & Laliberte, 1990; Lawler & Mohrman, 1987), and budgetary constraints (Nord & Tucker, 1987). Creating a climate conducive to the implementation of a new technology may require the development of organizational innovations such as those described in Johns (1993), which come with their own set of implementation barriers. Innovation-values fit is similar to the usefulness characteristic suggested by Sheridan (1988), which refers to a system’s perceived utility. Without proper climate or fit, attempts to implement innovative systems are likely to fail, making trust calibration irrelevant.

Other impediments to successful implementation and use include security and privacy concerns (Stewart, Mohamed, & Marosszeky, 2004), lack of technical acumen (Mathieson et al., 2001; Stewart et al., 2004), fear brought on by uncertainty (Stewart et al., 2004), and resistance to administrative changes (Edmondson et al., 2001; Levitt & March, 1988; Nelson & Winter, 1982). Each of these elements jeopardizes successful implementation and efficient use of technology such as ADAs.

Synthesis

The fast pace of technological advances has made it apparent that user acceptance plays a vital role in how well cutting edge information systems, such as ADAs, live up to their potential. Trust has proven to be a key component in user acceptance. If the barriers to successful implementation can be overcome, proper calibration of trust will ultimately lessen the occurrences of disuse and misuse of ADAs. Additionally, it is believed that building a
person’s trust in the capabilities of the system will increase acceptance (Bahmanziari et al., 2003).

Many factors affect trust, including both system characteristics and individual or user characteristics. This review of the literature revealed various calibration methods that have already been tested. Much of the emphasis has been on the need to develop robust systems that meet the requirements and expectations of the user; however, user characteristics are also attracting attention. Supplementary analysis of the impact of user characteristics might lead to administrative innovations such as training, recruitment, and even organizational changes to improve technology adoption (Damanpour & Evan, 1984; Johns, 1993).

Future Directions

The literature review did not reveal any studies involving the degree of trust that homeland security operators have in Department of Homeland Security (DHS)-specific ADAs. Therefore, a series of empirical studies should be conducted to evaluate the capabilities of DHS ADAs currently in use and identify areas where operator trust is not properly calibrated based upon system performance. This will provide some of the preliminary data needed to identify which of the factors listed in this document should be considered when developing trust calibration methods for homeland security applications. Future research should also consider which calibration methods are most effective and how to address the multifaceted nature of trust development and its impact on ADA use.

An additional concern is that although many calibration techniques require extensive system use to achieve results, often DMs are not allowed sufficient time to acclimate to a new system before it is implemented. Research suggests that user training may be instrumental to trust calibration; however, will this training be adequate to develop enough initial trust to use the system effectively? Further research is needed to investigate these issues and to determine the most effective trust calibration methods for critical homeland security-related decision making tasks.
Contact Information

Maranda McBride, PhD
North Carolina A&T State University
1601 E. Market Street
Greensboro, NC 27411
Phone: (336) 334-7656 ext. 7025
E-mail: mcbride@ncat.edu

Shona Morgan, PhD
North Carolina A&T State University
1601 E. Market Street
Greensboro, NC 27411
Phone: (336) 334-7656 ext. 4034
E-mail: smorgan@ncat.edu

Maranda McBride, PhD, is an Associate Professor of Management at North Carolina Agricultural &Technical State University (NCA&T). Her research interests include human-computer interaction, decision support display design, perception, and usability.

Shona Morgan, PhD, is an Associate Professor of Management at NCA&T. Her research interests include optimization problems and supply chain management.

References


