Institute for Homeland Security Solutions
Applied research • Focused results

March 2010

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

Adams, D. H., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly, 16*(2), 227–247.

Atoyan, H., Duquet, J.-R., & Robert, J.-M. (2006). *Trust in new decision aid systems*. Paper presented at the Human-Computer Interaction (IHM) Conference.

Bahmanziari, T., Pearson, J. M., & Crosby, L. (2003). Is trust important in technology adoption? A policy capturing approach. *Journal of Computer Information Systems, 43*(4), 46–54.

Cohen, M., Parasuraman, R., & Freeman, J. (1998). *Trust in decision aids: A model and its training implications*.

Damanpour, F., & Evan, W. (1984). Organizational innovation and performance: The problem of "organizational lag." *Administrative Science Quarterly, 29*(3), 392–409.

Davis, F. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Davis, F., Bagozzi, R., & Warshaw, P. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003.

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies, 58*, 719–735.

Dusire, S., & Falzon, P. (1999). *Trust and pilot-controller interaction.* Paper presented at the Tenth International Symposium on Aviation Psychology, Columbus, OH.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*(6), 697–718.

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*(1), 79–94.

Edmondson, A., Bohmer, R., & Pisano, G. (2001). Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative Science Quarterly, 46*(4), 685–716.

Fan, T., Hyams, D., & Kuchar, J. K. (1998). *Preliminary study of in-flight replanning performed on the flight deck*. National Center of Excellence for Aviation Operations Research (NEXTOR) Research Report RR-98-04.

Fleischer, M., Liker, J., & Arnsdorf, D. (1988). *Effective use of computer-aided design and computer-aided engineering in manufacturing*. Ann Arbor, MI: Industrial Technology Institute.

Friedman, B., & Kahn, P. H. J. (1997). Human agency and responsible computing: Implications for computer system design. In B. Friedman (Ed.), *Human values and the design of computer technology* (pp. 221–235). New York: Cambridge University Press.

Gaines, S., Panter, A., Lyde, M., Steers, W., Rusbult, C., Cox, C., 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*(3), 610–623.

Hu, P. J.-H., Lin, C.,& Chen, H. (2005). User acceptance of intelligence and security informatics technology: A study of COPLINK. *Journal of the American Society for Information Science and Technology, 56*(3), 235–244.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53–71.

Johns, G. (1993). Constraints on the adoption of psychology-based personnel practices: Lessons from organizational innovation. *Personnel Psychology, 46*, 569–592.

Johnson, J. D. (2004). *Type of automation failure: The effects on trust and reliance in automation* (unpublished master's thesis). Georgia Institute of Technology, Atlanta.

Kantowitz, B., Hanowski, R., & Kantowitz, S. (1997). Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors, 39*(2), 164–176.

Kaplan, E. (2007). *Homeland security technologies: Backgrounder.* Council on Foreign Relations. Last accessed on March 4, 2010, from http://www.cfr.org/publication/14827/homeland_security_technologies.html

Klein, K. J., Hall, R. J., & Laliberte, M. (1990). Training and the organizational consequences of technological change: A case study of computer-aided design and drafting. In U. E. Gattiker & L. Larwood (Eds.), *End-User Training (Technological Innovation and Human Resources)* (pp. 7–36). New York: de Gruyter.

Klein, K. J., & Sorra, J. S. (1996). The challenge of innovation implementation. *Academy of Management Review, 21*(4), 1055–1080.

Lawler, E., & Mohrman, S. (1987). Quality circles: After the honeymoon. *Organizational Dynamics, 15*(4), 42–54.

Lee, J. (2008). Review of a pivotal human factors article: "Humans and automation: Use, misuse, disuse, abuse." *Human Factors, 50*(3), 404–410.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35*(10), 1243–1270.

Lee, J., & Moray, N. (1994). Trust, self confidence, and operator's adaptation to automation. *International Journal of Human-Computer Studies, 40(1)*, 153–184.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80.

Levitt, B., & March, J. (1988). Organizational learning. *Annual Review of Sociology, 14*(1), 319–338.

Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM, 37*(7), 30–40.

Marsh, S., & Meech, J. (2000). *Trust in design.* Paper presented at the Conference on Human Factors in Computing, The Hague, The Netherlands.

Masalonis, A. J., Duley, J. A., Galster, S. M., Castano, D. J., Metzger, U., & Parasuraman, R. (1998). *Air traffic controller trust in a conflict probe during Free Flights.* Paper presented at the 42nd Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA.

Mathieson, K., Peacock, E., & Chin, W. (2001). Extending the technology acceptance model: The influence of perceived user resources. *ACM SIGMIS Database, 32*(3), 86–112.

McGuirl, J. M., & Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors, 48*(4), 656–665.

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors, 50*(2), 194–210.

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*(1), 44–58.

Muir, B. M. (1987). Trust between human and machines, and the design of decision aids. *International Journal of Man-Machine Studies, 27*(5–6), 527–539.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37*(11), 1905–1922.

Muir, B. M., & Moray, N. (1996). Trust in automation, Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39*(3), 429–460.

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*(3), 171–181.

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*(2), 223–239.

Nelson, R., & Winter, S. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Belknap Press of Harvard University Press.

Nickerson, J. V., & Reilly, R. R. (2004). *A model for investigating the effects of machine autonomy on human behavior*. Paper presented at the 37th Hawaii International Conference on System Sciences.

Nord, W. R., & Tucker, S. (1987). *Implementing routine and radical innovations*. Lexington, MA: Lexington Books.

Omodei, M., & McLennan, J. (2000). Conceptualizing and measuring global interpersonal mistrust-trust. *The Journal of Social Psychology, 140*(3), 279–294.

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

Parasuraman, R., & Riley, V. (1997). Human and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253.

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 30*(3), 286–297.

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.

Rousseau, D. M. (1989). Managing the change to an automated office: Lessons from five case studies. *Information Technology & People, 4*, 31–52.

Sheridan, T. B. (1975). *Considerations in modeling the human supervisory controller.* Paper presented at the IFAC 6th World Congress, Laxenburg, Austria.

Sheridan, T. B. (1988). *Trustworthiness of command and control systems.* Paper presented at the 2nd IFAC/IFIP/IFORS/IEA Symposium.

Sheridan, T. B., & Hennessy, R. T. (1984). *Research and modeling of supervisory control behavior*. Washington, DC: National Academy.

Shikishima, C., Hiraishi, K., & Ando, J. (2006). Genetic and environmental influences on general trust: A test of a theory of trust with behavioral genetic and evolutionary psychological approaches. *Japanese Journal of Social Psychology, 22*(1), 48–57.

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30*(4), 445–459.

Stewart, R., Mohamed, S., & Marosszeky, M. (2004). An empirical investigation into the link between information technology implementation barriers and coping strategies in the Australian construction industry. *Construction Innovation, 4*(3), 155–171.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research, 6*(2), 144–176.

Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology—Users want to trust, and generally do. But that trust is undermined, often forever, when the system delivers erroneous information. *Communications of the ACM, 42*(5), 39–44.

van Dongen, K., & van Maanen, P. P. (2005). *Designing for dynamic task allocation*. Paper presented at the 7th International Conference on Naturalistic Decision Making.

Wiegmann, D. A., McCarley, J. S., Kramer, A. F., & Wickens, C. D. (2006). Age and automation interact to influence performance of a simulated luggage screening task. *Aviation, Space, and Environmental Medicine, 77*(8), 825–831.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science, 2*(4), 352–367.

Yeh, M., & Wickens, C. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors, 43*(3), 355.