Exploring Trust Barriers to Future Autonomy: A Qualitative Look

Joseph B. Lyons[1(✉)], Nhut T. Ho[2], Anna Lee Van Abel[1],
Lauren C. Hoffmann[2], W. Eric Fergueson[1], Garrett G. Sadler[2],
Michelle A. Grigsby[1], and Amy C. Burns[1]

[1] Air Force Research Laboratory, Wright-Patterson AFB, OH 45433, USA
{joseph.lyons.6, anna.van_abel, william.fergueson,
michelle.grigsby.1, amy.burns.3}@us.af.mil
[2] NVH Human Systems Integration, Canoga Park, CA, USA
nhut.ho.51@gmail.com, lauren.c.hoffmann@gmail.com,
garrett.g.sadler@gmail.com

Abstract. Autonomous systems dominate future Department of Defense (DoD) strategic perspectives, yet little is known regarding the trust barriers of these future systems as few exemplars exist from which to appropriately baseline reactions. Most extant DoD systems represent "automated" versus "autonomous" systems, which adds complexity to our understanding of user acceptance of autonomy. The trust literature posits several key trust antecedents to automated systems, with few field applications of these factors into the context of DoD systems. The current paper will: (1) review the trust literature as relevant to acceptance of future autonomy, (2) present the results of a qualitative analysis of trust barriers for two future DoD technologies (Automatic Air Collision Avoidance System [AACAS]; and Autonomous Wingman [AW]), and (3) discuss knowledge gaps for implementing future autonomous systems within the DoD. The study team interviewed over 160 fighter pilots from 4th Generation (e.g., F-16) and 5th Generation (e.g., F-22) fighter platforms to gauge their trust barriers to AACAS and AW. Results show that the trust barriers discussed by the pilots corresponded fairly well to the existing trust challenges identified in the literature, though some nuances were revealed that may be unique to DoD technologies/operations. Some of the key trust barriers included: concern about interference during operational requirements; the need for transparency of intent, function, status, and capabilities/limitations; concern regarding the flexibility and adaptability of the technology; cyber security/hacking potential; concern regarding the added workload associated with the technology; concern for the lack of human oversight/decision making capacity; and doubts regarding the systems' operational effectiveness. Additionally, the pilots noted several positive aspects of the proposed technologies including: added protection during last ditch evasive maneuvers; positive views of existing fielded technologies such as the Automatic Ground Collision Avoidance System; the potential for added operational capabilities; the potential to transfer risk to the robotic asset and reduce risk to pilots; and the potential for AI to participate in the entire mission process (planning-execution-debriefing). This paper will discuss the results for each technology and will discuss suggestions for implementing future autonomy into the DoD.

Keywords: Trust · Automation · Autonomy · Military

D.N. Cassenti (ed.), *Advances in Human Factors in Simulation and Modeling*,
Advances in Intelligent Systems and Computing 591, DOI 10.1007/978-3-319-60591-3_1

1 Introduction

Driverless cars, autonomous drone delivery systems, collaborative robots as teammates, robotic concierges/hosts - these concepts are no longer science fiction as they may be coming to a home or business near you very soon. The notion of autonomy dominates contemporary visions for the future. Veloso and colleagues [1] have outlined a number of potential avenues for robotic systems in supporting the human race. Robotic systems are envisioned to support the elderly in their homes with physical movement, decision making, and even companionship. Robotic systems are hoped to revolutionize transportation and delivery systems. Robotic systems can support health care and connect doctors with patients at a distance. Backbreaking factory and warehouse work could be aided through the use of intelligent exoskeletons, as could individuals who have lost mobility due to medical conditions or injuries. Therapy and rehabilitation could be supported with robots. Customer service and of course entertainment are two other domains where robotic systems will likely make a huge impact on society – and in some instances they already are. While the technology possibilities are only limited by one's imagination, there is one core element of each of the above examples that constrains the potential gains for future robotic systems, and that is the fact that all of these systems will need to, at some point, interface with humans. This has led to a burgeoning of the domain of Human-Robot Interaction (HRI) which studies numerous facets of how to improve human-robot interaction. One key challenge in this domain area is the issue of how to foster appropriate levels of trust of the robotic systems – i.e., will humans accept these technologies or reject them?

Trust represents one's willingness to be vulnerable to another entity in a situation where there is some risk and little ability to monitor the other [2]. The trust construct has been applied to trust of technology to represent the "attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [3, p. 54]. Thus, trust is relevant for both interactions with people as well as with intelligent agents such as robots. Chen and Barnes [4] define an agent as a technology that has autonomy, the ability to observe and act on its environment, and the ability and authority to direct actions toward goals. The current paper will use the term "future autonomy" to collectively represent the notion of robotic systems and agent-based technologies, the latter of which may have no physical embodiment. The key attribute of relevance for future autonomy in this context is that such systems will have both the capability and the authority to act in relevant operational scenarios.

Without a doubt, future autonomy is imminent, yet human trust will determine the effectiveness of said technology as humans decide whether or not to use it, and how to use it. The notion of trust calibration, or appropriate levels of trust, is the critical factor in determining the effectiveness of future autonomy. Meaning, the key challenge is in understanding when to trust and when to distrust technology. Inaccurate trust can result in catastrophic errors when humans rely on technology that is faulty or error-prone. Several accidents have been blamed on human overreliance on automated systems such as the Turkish Airlines flight 1951 in 2009 when a pilot relied on autopilot after an instrumentation failure [5]. The inverse is also problematic in that under trust can be detrimental to performance when humans fail to use a reliable technology as evidenced by the Costa Concordia cruise ship disaster that killed 32 passengers when the ship captain used manual navigation skills instead of a reliable automated navigation tool [5]. Appropriately calibrated trust is challenging because as technology gains in reliability humans tend to trust it more – appropriately so. However, the performance costs of errors are most severe in situations when a highly reliable system is given the highest level of autonomy and that system makes a mistake [6]. This is driven by the habit of humans to reduce their monitoring of highly reliable systems, which could make compensation, correction, and adaptation to novel demands more difficult when the technology fails. This paradox of automation has motivated the research community to examine the drivers (and detractors) of human trust of technology.

Trust has been a focal topic for researchers in the areas of automation [5] and robotics [7]. While a comprehensive review of this literature is beyond the scope of the current paper, the human-machine trust literature has identified a number of key trust antecedents including: performance [7], transparency [4, 8], perceived benefits of use [9], prior experiences with the system to include error types (i.e., false alarms and misses) and the timing of errors [5, 10], interactive styles (etiquette) [5], anthropomorphism [11], and individual differences such as one's perfect automation schema [12], to name a few. Yet despite the burgeoning literature on human-machine trust, little field work has been done to examine the trust barriers among **real** operators to **real** tools that have **real** consequences in the world (R3) for trust or distrust. One such study found that pilot trust of an automated safety system in fighter aircraft was driven by performance considerations (reliable performance and system behavior that does not interfere with the pilot's ability to fly and fight), transparency, perceived benefits and logical (compelling) rationale for why the technology was needed, and familiarity of system's behavior [13]. It is likely that these same dimensions will be important considerations for future autonomy. Further, as technology increases in both decision capability and authority, it is likely that decision making capability and intent of the system will be important trust considerations [14]. Thus, the antecedents of trust for systems that involve a broader range of decision options should involve more intent-based dimensions relative to mere automated systems that have both decision authority and capability but only under the confines of a narrower set of circumstances.

The Department of Defense (DoD) is very focused on technologies for autonomy. Despite the domain of autonomy being quite broad, the notion of trust in autonomy is a persistent theme throughout much of the DoD Research Doctrine [15, 16]. Yet, it is critical to contextualize the target domain when considering trust so that trust considerations have a focused technology as a trust referent. Thus, the current paper will discuss pilot reactions to two future technologies: the Automatic Air Collision Avoidance System (AACAS) and an Autonomous Wingman (AW). AACAS has already undergone flight testing and is a more mature technology than the AW technology is currently.

The Automatic Air Collision Avoidance systems is part of the Air Force Research Laboratory's Integrated Collision Avoidance Program which seeks to integrate the already-fielded Automation Ground Collision Avoidance System (AGCAS) with AACAS. AACAS was designed to mitigate mid-air collisions among fighters by calculating future aircraft trajectories of cooperative and non-cooperative aircraft and using a collision avoidance algorithm to determine if an automatic maneuver is required to avoid aircraft collision [17, 18]. Like prior systems such as AGCAS, AACAS must avoid interference with the pilots [18] which by avoiding nuisance the pilots should view the system as more trustworthy and be more likely to trust it [13]. The AW is more of a future concept, but would involve the notion of a robotic aircraft that serves as a subordinate to the flight lead. The AW would handle its own flight maneuvers but would be under the direct control of the flight lead to use as needed. Unlike current Remotely Piloted Aircraft (RPAs) the AW would not be remotely piloted but rather would able to respond to higher-level commands from the flight lead. Relative to AACAS, AW would be expected to be capable of handling a broader range of activities, whereas AACAS has one action – avoiding collision with other aircraft.

2 Method

2.1 Participants

The participants were operational F-16 (N = 131) and F-22 (N = 35) pilots at operational Air Force bases. Nine different F-16 units were visited, 4 of which were outside of the Continental United States. Two F-22 units were visited both of which were in the Continental United States. All of the pilots had, at a minimum, completed basic flight training and were operational pilots within the Air Force. The F-16 pilots had an average of 836 flight hrs. and the F-22 pilots averaged 372 h.

2.2 Procedure

Semi-structured interviews were conducted in person at the F-16/F-22 units. The current data were collected as part of a larger set of interviews centered on trust of ground collision avoidance systems. All pilots were first given an informed consent document which discussed the study objectives. Following consent, the pilots were administered a structured interview focused on attitudes and experiences of ground collision avoidance technologies that are already fielded on the F-16 and F-22. Following this set of questions, the pilots were given written descriptions of the two future technologies (AACAS and AW). After the pilots read the descriptions, a few questions were asked relating to their attitudes toward these systems. The current paper focuses on a subset of those data, and in particular, on responses to two questions: (1) In your opinion, what would be the biggest trust barrier with the AACAS system?, and (2) In your opinion, what would be the biggest trust barrier with an autonomous wingman? Responses were recorded by digital recorders (based on approval of the pilots) for later transcription and analysis. The entire interview lasted on average between 20–30 min. Data were coded with NVivo version 11 qualitative analysis software package. Note that each pilot was asked to provide the "biggest" trust barrier but they could provide multiple trust barriers for each technology.

3 Results

The relevant data clusters are reported in Tables 1 and 2 below. As shown in Table 1, the primary trust barriers reported by F-16 pilots for AACAS involved performance-related issues (e.g., reliability, connectivity issues, and concern about interference). The

Table 1. Clusters and frequencies for F-16 pilots. Note: AACAS = Automatic Air Collision Avoidance System.

AACAS	Cluster	Frequency
	Too conservative	3
	Tactical disadvantage	3
	Concern about close formation flight	41
	Interference concerns	44
	Digital problems	44
	Reliability	45
Autonomous Wingman	Changing training	2
	Replacing pilots	5
	Accountability concerns	8
	Being hacked	9
	Adaptation (concerns about the system's ability to adapt)	16
	Hitting me	18
	Added communications requirement	34
	Lack of a thinking human	42
	Reliability	45
	Workload concerns	47

Table 2. Clusters and frequencies for F-22 pilots. Note: AACAS = Automatic Air Collision Avoidance System.

AACAS	Cluster	Frequency
	Reliability	4
	Digital problems	3
	Close formation flight concerns	5
	Cause tactical disadvantage	8
	Interference	11
Autonomous Wingman	Replace pilots	1
	Reliability	2
	Being hacked	2
	Hitting me	2
	Changing training	3
	Accountability concerns	3
	Added communications requirement	4
	Adaptation (concerns about the system's ability to adapt)	6
	Lack of thinking human	8
	Workload concerns	11

primary trust barriers reported by F-16 pilots for the AW included: workload concerns, reliability, and the lack of a human decision maker. As shown in Table 2, the primary trust barriers reported by F-22 pilots for AACAS involve performance issues: concern about interference and that the system could create a tactical disadvantage in combat. The primary trust barriers reported by F-22 pilots for AW involve concern about increased workload, the lack of a human decision maker, and the AW's ability to adapt to novel constraints.

4 Discussion

The present paper examined trust barriers among operational pilots in relation to two forms of future autonomy within the Air Force, namely the AACAS and the AW technologies. While both 4th and 5th Gen fighter pilots served as the samples, the responses were fairly consistent between both sets of pilots; therefore, the data will be discussed across both samples rather than by a specific platform type (i.e., F-16/F-22). For AACAS, the primary concerns for pilots revolved around performance issues. Like prior fielded automated systems on fighter aircraft [13, 17, 18] pilots were very concerned about interference. Pilots did not want AACAS preventing them from getting close enough to other aircraft for training, battle damage checks, or most importantly, during Basic Flight Maneuvers (BFM). These concerns about interference were largely in preventing the pilot from engaging in a maneuver that was desired, essentially demonstrating concerns about false alarms. There were also concerns about the system causing harm by maneuvering one aircraft into another during the execution of an automated avoidance action in close formation. Pilots also reported concerns about the reliability of the system in general (given the complexity of the data links and algorithms required for the system), as well as concerns about the data linkages between cooperative and non-cooperative aircraft. In this case, cooperative aircraft would be those with a similar AACAS system and sensing capability, and non-cooperative would be those without AACAS. Given the speeds and tactical requirements of operating a fighter aircraft, these concerns are logical as pilots need to maintain a tactical edge on the battlefield. Pilots want a system that is both highly reliable, but not prone to nuisance activations (e.g., false alarms – activating when an activation was not necessary). Consistent with the literature on trust of automation [3, 4, 6, 7], performance and reliability are significant drivers of trust of technologies like AACAS. Additionally, pilots reported that they could see value in AACAS if the reliability of the system was very high and interference could be eliminated/minimized. This value was noted mostly as a "last ditch" maneuver to avoid an otherwise imminent collision.

The reported trust barriers for AW were a bit broader than those for AACAS and this may be reflected for two reasons: (1) AACAS is a more mature technology relative to AW and as such the trust concerns from pilots of AW may be driven by an overall uncertainty associated with AW, and (2) AW is intended to operate within a broader array of situations which thus creates greater complexity for trust evaluations. The pilots' top concern was related to the expectation that the AW would add to an already high-workload environment. Operational fighter pilots operate at a high ops tempo and the flight requirements, communication requirements, and operational requirements create a high workload situation. Adding the complexity of communicating with and "leading" an AW raises concerns that pilots do not want the added workload. Like AACAS, reliability was also an issue for pilots when considering the AW.

In contrast to AACAS, when pilots considered the AW, they reported concerns about the lack of a human decision maker in the cockpit. Given the time-sensitive and dangerous domains that military personnel are faced with, these concerns are logical. Specifically, there are concerns that the system would make the wrong decision when faced with a difficult situation. Herein, it would be useful to highlight the intent-based transparency of the AW to the pilots, as called for in general by [14]. By using intent-based transparency methods the pilots and AW would have more opportunities to establish shared intent, which is crucial in dynamic, morally contentious situations. Shared intent allows two or more entities to establish predictable behaviors/reactions to novel constraints. Shared intent is important in this context due to the fact that pilots reported concerns about accountability for the AW. This is also important because pilots also noted that they have concerns about the AW's ability to adapt to novel demands. The pilots seemed to want the AW to be able to think and respond "like a human," however that may not be the best approach for this human-machine team. A more fruitful approach may involve leveraging the strengths of the AW and building a flight lead-AW relationship in a way that maximizes the strengths and minimize the weakness of each partner. This heterogeneous, but synergistic approach could maximize the effectiveness of the human-autonomy team. Further, the pilots noted concerns about potential hacking of the AW, and the potential for the AW to physically "hurt" the pilot by running into her/him. Thus, while performance-related concerns were definitely present for AW, similar to AACAS, the pilots seemed to also consider intent-based issues in relation to AW. The identification of these trust barriers are important for researchers and designers to consider in the development and fielding of future autonomy. Like AACAS, the pilots reported a number of potent benefits of a system like an AW to include: risk reduction for pilots (i.e., fewer pilots in harm's way), using the AW to engage particularly risky targets or in very risky situations, using the AW to carry additional assets such as weapons and sensors, using the AW to jam surface-to-air missile batteries (i.e., to protect the pilot).

The next section presents a series of recommendations for military organizations seeking to field future autonomy systems. First, performance-related issues will be a paramount concern among operators. Thus, military organizations are encouraged to use videos as a means to "show" the performance and reliability of the system. There are two potential ways in which videos can be incorporated. Videos of operational performance should be shown to highlight both positive and negative exemplars of the system's performance. The positive videos should boost trust, as demonstrated by a prior field study examining trust of the AGCAS system [19]. Yet, care must be taken to avoid situations of over trust as videos have the potential to generate high trust among individuals with little system experience which could negatively impact trust calibration. The videos serve as operational evidence of the system's performance. While videos of negative system performance may cause a decrease in trust they are important for sharing stories among operators and will help the operators to understand the limits of the system. After all, a decrease in trust can be beneficial if it leads to a more accurate calibration of one's trust. The second type of video might include test videos which show the system in scripted scenarios that test the limits of the system. Such videos would be impossible (and unethical) to create in actual operations, so testing seems like the right opportunity for such videos. Anecdotally for the present study, following the interviews, most of the pilots had an opportunity to discuss AACAS with a subject matter expert (SME) on the system and when that SME showed the pilots a successful test video of the AACAS in a close proximity high-speed pass, the effects on pilot trust were virtually spontaneous. In this case, "seeing is truly believing."

Understanding the intent of the systems also seems to be an important theme emerging from this research. Using intent-based transparency methods should help to foster shared intent between the human and the system [14]. This shared intent, should in turn support predictability for how the system will behave in novel situations. If one understands the rules that govern the system's behavior (i.e., goals, goal priorities, interactive styles, rules of engagement) then the system's reaction to novel demands should be more predictable, at least more understandable. Intent-based transparency could be established through education and joint human-machine training. The educational aspects could focus on the background and purpose of the system, why it was designed, how the system sets and prioritizes goals in changing contexts, and the rationale for decision making processes. More importantly, the human should engage in joint human-machine training to experience how the system reacts to novel demands. Herein, the design of the scenario should be done in such a way as to stress the boundaries of the situation to maximize the range of potential decision options. Again, the interest is in building an understanding of the behavioral rules used to govern the system's behavior, and in establishing some predictability of how the system executes those rules in various conditions. In this sense, having experience with a system reacting to the same or very similar circumstances is less variable than exposure to a smaller subset of encounters with novel stimuli that stress the system's range of behavioral flexibility.

The current research is not without limitations. One limitation is that the study was limited to military personnel and military technologies. Future autonomy in the commercial sector could be perceived differently than military technologies. Further, non-military personnel may be more or less accepting of future autonomy, relative to military personnel. For instance, autonomous cars are beginning to hit the market and recent accidents have been blamed on overreliance on the technology. Military fighter pilots may be more prone to be skeptical of new technologies. A second related limitation is that only military technologies have been considered in this study. Technologies that are available on the commercial market may be perceived differently than military technologies. However, both AACAS and AW fit the criteria for "R3" in that they are **real** technologies, with **real** operators that have the potential for **real** consequences in the world. Finally, the current study involved qualitative data, future research on this topic might include experimental studies to pinpoint the impact of different trust factors on trust intentions and trust-based behavior.

References

1. Veloso, M., Aisen, M., Howard, A., Jenkins, C., Mutlu, B., Scassellati, B.: WTEC Panel Report on Human-Robot Interaction Japan, South Korea, and China. World Technology Evaluation Center, Inc., Arlington (2012)

2. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrated model of organizational trust. Acad. Manag. Rev. **20**, 709–734 (1995)
3. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors **46**, 50–80 (2004)
4. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multirobot control: a review of the human factors issues. IEEE Trans. Hum.-Mach. Syst. **44**(1), 13–29 (2014)
5. Hoff, K.A., Bashir, M.: Trust in automation: integrating empirical evidence on factors that influence trust. Hum. Factors **57**, 407–434 (2015)
6. Onnasch, L., Wickens, C.D., Li, H., Manzey, D.: Human performance consequences of stages and levels of automation: an integrated meta-analysis. Hum. Factors **56**, 476–488 (2014)
7. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., de Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. Hum. Factors **53**(5), 517–527 (2011)
8. Lyons, J.B., Saddler, G.G., Koltai, K., Battiste, H., Ho, N.T., Hoffmann, L.C., Smith, D., Johnson, W.W., Shively, R.: Shaping trust through transparent design: theoretical and experimental guidelines. In: Savage-Knepshield, P., Chen, J. (eds.) Advances in Human Factors in Robotics and Unmanned Systems, pp. 127–136. Springer, Cham (2017)
9. Li, X., Hess, T.J., Valacich, J.S.: Why do we trust new technology? A study of initial trust formation with organizational information systems. J. Strateg. Inf. Syst. **17**, 39–71 (2008)
10. Guznov, S., Lyons, J.B., Nelson, A., Wooley, M.: The effects of automation error types on operators trust and reliance. In: Proceedings of HCI International, Toronto, CA (2016)
11. Pak, R., Fink, N., Price, M., Bass, B., Sturre, L.: Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. Ergonomics **55**(9), 1–14 (2012)
12. Merritt, S.M., Unnerstall, J.L., Lee, D., Huber, K.: Measuring individual differences in the perfect automation schema. Hum. Factors **57**, 740–753 (2015)
13. Lyons, J.B., Ho, N.T., Koltai, K., Masequesmay, G., Skoog, M., Cacanindin, A., Johnson, W.W.: A trust-based analysis of an air force collision avoidance system: test pilots. Ergon. Des. **24**, 9–12 (2016)
14. Lyons, J.B.: Being transparent about transparency: a model for human-robot interaction. In: Sofge, D., Kruijff, G.J., Lawless, W.F. (eds.) Trust and Autonomous Systems: Papers from the AAAI Spring Symposium (Technical Report SS-13-07). AAAI Press, Menlo Park (2013)
15. Defense Science Board (DSB) Task Force on the Role of Autonomy in Department of Defense (DoD) Systems. Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. Washington, DC (2012)
16. Defense Science Board (DSB) Summer Study on Autonomy. Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. Washington, DC (2016)
17. Wadley, J., Jones, S.E., Stoner, D.E., Griffin, E.M., Swihart, D.E., Hobbs, K.L., Burns, A.C., Bier, J.M.: Development of an automatic air collision avoidance system for fighter aircraft. In: AIAA Infotech@Aerospace Conference, Guidance, Navigation, and Control and Co-located Conferences. Boston, MA (2013)
18. Jones, S.E., Petry, A.K., Eger, C.A., Turner, R.M., Griffin, E.M.: Automatic integrated collision system. In: 17th Australian Aerospace Congress. Melbourne, AU (2017)
19. Ho, N.T., Sadler, G.G., Hoffmann, L.C., Lyons, J.B., Fergueson, W.E., Wilkins, M.: A longitudinal field study of auto-GCAS acceptance and trust: first year results and implications. J. Cognit. Eng. Decis. Mak. (in press)