A Longitudinal Field Study of Auto-GCAS Acceptance and Trust: First-Year Results and Implications

Nhut Ho, **Garrett G. Sadler**, **Lauren C. Hoffmann**, **Kevin Zemlicka**, NVH Human Systems Integration, LLC, **Joseph Lyons**, **William Ferguson**, Wright Patterson Air Force Research Lab, **Casey Richardson**, **Artemio Cacanindin**, **Samantha Cals**, Air Force Test Center, and **Mark Wilkins**, Office of the Under Secretary of Defense for Personnel and Readiness

In this paper we describe results from the first year of field study examining U.S. Air Force (USAF) F-16 pilots' trust of the Automatic Ground Collision Avoidance System (Auto-GCAS). Using semistructured interviews focusing on opinion development and evolution, system transparency and understanding, the pilot–vehicle interface, stories and reputation, usability, and the impact on behavior, we identified factors positively and negatively influencing trust with data analysis methods based in grounded theory. Overall, Auto-GCAS is an effective life-/aircraft-saving technology and is generally well received and trusted appropriately, with trust evolving based on factors including having a healthy skepticism of the system, attributing system faults to hardware problems, and having trust informed by reliable performance (e.g., lives saved). Unanticipated findings included pilots reporting reputation to not be negatively affected by system activations and an interface anticipation cue having the potential to change operational flight behavior. We discuss emergent research avenues in areas of transparency and culture, and values of conducting trust research with operators of real-world systems having high levels of autonomy.

Keywords: autonomy, human automation trust, Auto-GCAS, trust evolution, trust calibration, transparency

Address correspondence to Nhut Ho, at NVH Human Systems Integration, 20944 Sherman Way #210, Canoga Park, CA 91303, USA, nhut.ho.51@gmail.com.

Journal of Cognitive Engineering and Decision Making
2017, Volume 11, Number 3, September 2017, pp. 239–251
DOI: 10.1177/1555343417701019

INTRODUCTION

Technological advances are reshaping the modern-day battlefield as the military moves toward increased use of autonomy (i.e., systems that have both the authority and ability to execute actions in support of the mission). The Department of Defense is looking to autonomy to reduce risk to personnel, increase decision speed, and enhance mission effectiveness in austere and dangerous domains (Defense Science Board, 2016). Future autonomy may negate human limitations of fatigue, negative emotion (e.g., anger), and limited cognitive capacity and may offer greater computational power than possible within the human brain. Yet, autonomy is no panacea. Technology with high levels of decision initiative can be problematic, particularly when the system makes an error. Onnasch, Wickens, Li, and Manzey (2014) found that the impact of automation errors on human performance was most detrimental in situations where the automation had higher levels of autonomy. They attributed this effect to an inability of human operators to appropriately calibrate their trust of the technology.

Trust, the willingness of an individual to accept vulnerability in relation to another entity in a context whereby he or she has few opportunities to monitor the entity (see Mayer, Davis, & Schoorman, 1995), is the process that guides human decisions to rely or not rely on technology. Although the literature on trust of technology has burgeoned in recent years (see Hoff & Bashir, 2015, for a review), the vast majority of human–autonomy trust research is conducted in laboratory settings, threatening the external validity of the trust domain. Research is needed,

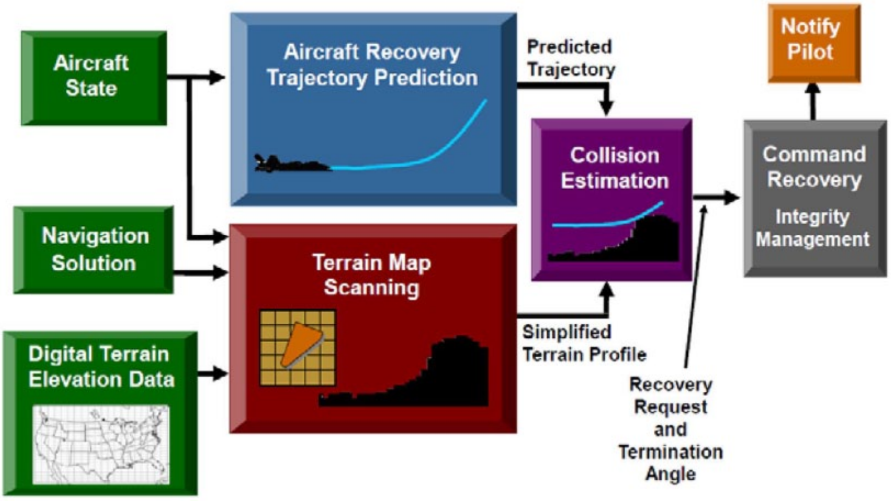

Figure 1. Automatic Ground Collision Avoidance System architecture (Burns, Harper, Barfield, Whitcomb, & Jurusik, 2011).

particularly within the military, on trust of fielded technologies used by operators in the real world. Only then can we have confidence that the extant models of human–autonomy trust are valid in the real world and, thus, have greater potential to support the military mission. It is in this spirit that we will examine within the current paper trust of a fielded technology within the U.S. Air Force (USAF), the Automatic Ground Collision Avoidance System (Auto-GCAS).

Auto-GCAS is a technology that has been developed by the USAF, Lockheed, and NASA in the past 30 years to mitigate controlled flight into terrain (CFIT), the leading cause of loss of life in U.S. military aviation. Auto-GCAS avoids collisions by (1) positioning the aircraft over a map based on Digital Terrain Elevation Data (DTED) with an inertial navigation system, (2) projecting the aircraft trajectory over the DTED map, (3) generating a terrain profile from the local terrain map based on the aircraft's location and current maneuvering, (4) comparing the projected trajectory against the terrain profile to determine if an imminent threat exists, and (5) if the threat exists, activating a last-second automatic escape maneuver and returning the control to the pilot as soon as the threat is avoided (see Figure 1; Burns, Harper, Barfield, Whitcomb, & Jurusik, 2011). These functionalities were designed based on the following three ranked system requirements: (1) do no harm, which requires that Auto-GCAS not cause any harm to the pilot or the aircraft; (2) do not impede, which requires the system to be nuisance free and thus not interfere with the mission and piloting duties; and (3) avoid collision (prevent harm), which requires the system to avoid collision with terrain (Skoog, 2008). For more technical information about Auto-GCAS, see Swihart et al. (2011). The USAF has been integrating Auto-GCAS into the USAF production F-16 since 2011 and deployed it in 2014. Its deployment has been projected to save the USAF 11 F-16 aircraft, nine pilots, and $400 million over the future life of the F-16 (Wilkins & Barfield, 2016). At the time of this writing, Auto-GCAS has had four confirmed saves (Norris, 2016).

Prior research (Koltai et al., 2014; Lyons, Ho, et al., 2016; Niedober et al., 2014) was conducted in 2012 to 2014 on the trust development of Auto-GCAS engineers, experimental test pilots, and managers. It showed that the development, testing, and perspectives of experimental test pilots and engineers were closely aligned to best practices identified by research on fostering appropriate trust in automation. Researchers (e.g., Lee & See, 2004) have suggested that trust can be enhanced by showing past performance of the system, showing the process (i.e., how the system works), making the technology and its algorithms understandable, communicating or visualizing the intent of the system, and using training to assess and verify the system's reliability. Auto-GCAS exemplifies many of these recommendations: (a) It has demonstrated high reliability through extensive flight testing, shaping the trust of experimental test pilots and engineers through data; (b) the fly-up avoidance maneuver is consistent with pilots' preferred behavior (i.e., aggressive and timely wings-level and 5-g pull-up), which helps the pilot familiarize with the system and understand its behavior; (c) the intent of the system is to not only protect the pilot but also avoid interfering with the pilot (this is a unique aspect of this system in relation to prior systems and helps the pilot to understand the system's intent); and (d) the system has undergone extensive testing with experimental test pilots receiving considerable training on the system, supporting more calibrated trust over time (Ho, Sadler, Hoffmann, Lyons, & Johnson, 2017). (This publication focused on a multistakeholder trust model based on a case study of managers/leaders, engineers, and test pilots who funded, designed, built, and tested Auto-GCAS. The case study was conducted prior to Auto-GCAS fielding, whereas the current research was conducted after Auto-GCAS fielding and focuses on operational pilots who are the end users of the system.) Overall, experience with the Auto-GCAS system has moved engineers and experimental test pilots toward more optimal trust strategies, indicating the potential for appropriate trust development in operational pilots.

Because these findings were based on predeployment data with a small group of experimental test pilots, a population that is different from the general population of operational pilots who will primarily use the system, and because trust calibration is a dynamic process spanning from the time when the system is first conceived until it goes into operation and retirement, it is imperative to follow the deployment of Auto-GCAS to examine operational pilots' trust longitudinally and in the context of operation. An important advantage of long-term field studies, such as the one presented here, is the ability for data collection and analysis to reveal emergent themes or topics as they develop in real time. The field study approach allows flexibility for modifications to design, preparation, and data collection to allow such emergent themes to be adequately explored. In this paper, we will discuss the methods used for the 1st year of the field study, report the results that were obtained, and discuss the implications of our results in terms of the longitudinal nature of the research.

METHOD

Participants

Participants were pilots of the USAF Block 40/42/50/52 F-16 aircraft with Auto-GCAS. A total of 168 pilots participated in semistructured interviews, of which 158 agreed to be recorded and were used in transcription. Between April and September of 2015, interviewed pilots were stationed at eight USAF bases and two Air National Guard (ANG) bases. This selection of USAF and ANG bases provides for a broad variety of mission types (e.g., flying over water, flat and mountainous terrain) and pilot career stages (e.g., full-time active duty and part-time reserve status). Participants had a self-reported estimated average of 923 flight hours in the F-16 and an average of 45 hr flying with Auto-GCAS installed on their planes.

Interview Protocol Design

Our interview questions were developed with the assistance of subject matter experts involved in the development and fielding of Auto-GCAS and were subsequently approved through the Air Force Research Laboratory Institutional Review Board office. We utilized the semistructured interview (DeWalt & DeWalt, 2011) in order to obtain targeted data on specific topics, termed "cluster areas," while also allowing the plasticity to engage pilots when emergent themes arose. Interviews were designed to accommodate pilots' busy schedules and limited availability, with the typical interview designed to last between 15 and 30 min. The subsections that follow elaborate on the six interview question cluster areas developed for the 1st year of study.

Opinion development and evolution of Auto-GCAS. To understand pilots' trust impressions, prompts facilitated explanations about initial impressions of the system, how and if their views in the system had changed over time since their initial impressions were formed, and any reported causes of change in attitude. By providing pilots' early impressions of the system, these questions allowed examination into the main research goal of this project: to determine how trust and opinion of Auto-GCAS evolves over time. An additional motivation for inquiring about initial impressions of the system is previous research showing early experience with automated systems to play a substantial role in overall opinion development and trust (Li, Hess, & Valacich, 2008).

Auto-GCAS transparency and pilot understanding. System transparency involves a shared awareness and shared intent between the user and automation (Lyons, 2013). Accordingly, questions in this cluster area involved examination of pilot understanding about why Auto-GCAS was developed, the nature of how and when it will activate, and any unintuitive, upsetting, or confusing issues that a user had experienced with the system. Additionally, the business case of Auto-GCAS was assessed, as user understanding of all the reasons a system has been implemented is a fundamental component of system transparency. However, initial interview results proved asking about why the system had been developed both appeared self-evident and had the potential to seem insensitive to those pilots who had lost a fellow airman, and this question was therefore dropped from the remaining interview protocols.

Pilot–vehicle interface. Singled out as a critical element of transparency for this system are warning indicators (i.e., chevrons; see Figure 2) on the pilot's head-up display (HUD) that converge as Auto-GCAS is about to activate and form an *X* upon activation. The Auto-GCAS user interface was a key interest in this study, as an earlier case study revealed concerns on behalf of Auto-GCAS developers about the chevrons in the display having the potential to change pilot behaviors (Koltai et al., 2014). As such, questions were asked relating to observation, preference, and opinions about the chevrons. This cluster of questions would ultimately inform the research about pilots' overall view of

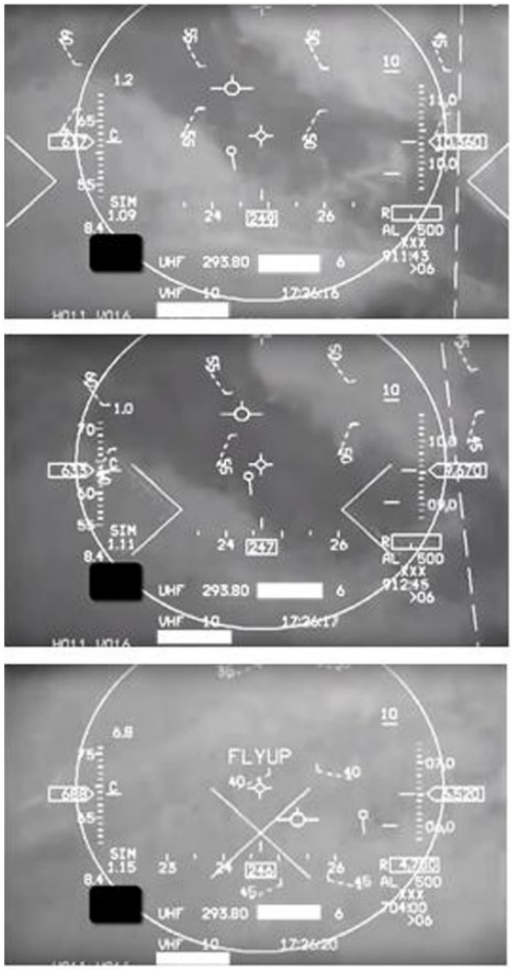

Figure 2. Chevrons on head-up display (HUD). Two opposite arrows shown in three different locations as they move toward each other: at the edge of the HUD (top picture), near the HUD's center (middle picture), and touching to form an *X* at the HUD's center (bottom picture) as Automatic Ground Collision Avoidance System activates.

the chevrons including why they did or did not prefer the display, if they believed the chevrons were altering their normal in-flight behavior, and/or how chevrons might have the potential to change it.

Stories and reputation of Auto-GCAS performance. These questions allowed us to track the flow of information about system performance, novel occurrences of unanticipated system behavior, and whether the system had an overall favorable, unfavorable, or neutral outlook at the squadron level. This information would allow researchers to determine whether there was knowledge of saves or nuisance activations at particular bases and to understand how that information flow affected Auto-GCAS acceptance and trust. An additional benefit of this group of questions allowed discovery of yet-unanticipated and anomalous issues with the system, which were reported to appropriate parties to have solutions developed and implemented.

Usability of Auto-GCAS. The avoidance of nuisance Auto-GCAS activations was identified in previous research as a critical component in the acceptance and usability of Auto-GCAS (Koltai et al., 2014). Questions about the usability were asked in order to gauge whether there were common or recurring themes relating to system activations as well as to understand how pilots felt about the usability of the system and potential interference from the system.

Impact of Auto-GCAS on operational pilots' behavior. The potential or actual impact of the introduction of Auto-GCAS on piloting behavior was a major concern for this study. Given the results of the case study, a particular focus regarding the impact of chevron warning indicators on pilot reaction was emphasized in the interview protocol. Because this study's research design involved a flexible, semistructured interview approach, pilots were also probed when their responses to any question addressed potential or actual changes in behavior while flying.

Data Collection and Analysis

Interviews were conducted on-site at USAF and ANG facilities and were coordinated carefully to fit each squadron's schedule, minimizing impact to flying and other operational duties. With the exception of participants who declined recording, interviews with participants were audio recorded and subsequently transcribed. When audio recording was declined, pen-and-paper notes were taken during the interview. At the conclusion of each day of interviews at a USAF or ANG base, our research team and subject matter experts participated in a "debriefing" session to discuss themes that were novel (i.e., had not been heard up to that point) or confirmatory (i.e., had been mentioned previously multiple times). These debriefing sessions were useful for providing an early, qualitative analysis of interview results and guiding the qualitative coding process detailed later.

In the present study, we employed semistructured interviews to facilitate the use of grounded theory as an inductive, hypothesis-generating method, and surveys facilitated data for deductive, qualitative analysis. Grounded theory is a method of qualitative analysis that is ideal for discovering emergent themes in a novel area of research by developing categories and themes that emerge "organically" from interview transcript data (Bernard, 2011). We applied the grounded theory approach to interview data by reviewing transcripts for the development of an initial "codebook," attaching codes from the generated codebook to transcript passages, all while continually keeping running notes (known as memos) about the process and any intercode relationships. Throughout this process, the codebook was iteratively revised and improved to reflect categorical relationships that arose during the data coding process.

Interview transcripts were analyzed using the NVivo 11 qualitative analysis software package. Previous research (Bernard, 2011) has indicated after conducting between 25 and 30 interviews, few novel categories tend to arise. Using this finding as a guideline, a representative sample of 30 interview transcripts was taken from the full set of interviews for development of a codebook. After this codebook was assembled, we had a three-person team code the remainder of interview transcripts and meet weekly to discuss results, ensure intercoder reliability, reassess code definitions, and discuss possible code relationships. The findings presented next are the result of the qualitative analysis produced from this coding process.

RESULTS

Given the richness and extensiveness of results obtained, major themes that were revealed are reported here, including percentages of how many participants from transcribed interviews adhered to each major theme, although many responses were provided without direct questions

TABLE 1: Summary of Elements of Auto-GCAS Acceptance and Trust

Question Cluster	Primary Themes	Secondary Themes
Opinion development and evolution of Auto-GCAS	1. System well received 1a. Deployment overdue 1b. Loss of life 2. Healthy skepticism 3. No change in trust 4. Trust despite faults 5. Trust informed by saves	Backup tool Distrust of increased automation
Auto-GCAS transparency and pilot understanding	1. DTED knowledge 2. General knowledge gap 3. Reported no confusion 4. Issues loading DTED data 5. On/off, over-water confusion 6. Save lives and aircraft	Distance error/ calculations for a fly-up
Chevrons	1. Chevrons defaulted on 1a. Provide feedback/SA 2. Chevrons cause no change in behavior 3. Chevrons potential conservative change	Chevrons distracting Chevrons potential aggressive change
Stories and reputation	1. Favorable reputation 1a. Favorable reputation despite limitations 2. Positive story 3. Negative story 4. Third-person activation stories	First-person activation stories
Usability of Auto-GCAS	1. No nuisance 2. Potential nuisance in combat	Concerns about dangerous activations
Impact of Auto-GCAS on behavior	1. No influence 2. Learning opportunity 3. Influence situation dependent	Negative influence if recurring

Note. Auto-GCAS = Automatic Ground Collision Avoidance System; DTED = Digital Terrain Elevation Data; SA = situation awareness.

from the interview script. Also included are samples of source quotes demonstrating key points assessed after interview coding, providing examples of the major results found. Table 1 summarizes major themes found throughout the interview process.

Opinion Development and Evolution of Auto-GCAS

To understand how pilots trust Auto-GCAS, they were first asked about their initial perceptions and attitudes toward the system. Data from this question indicated that Auto-GCAS was well received for the majority of interviewed pilots (59%), as two common themes included pilots thinking that a system capable of preventing CFIT in cases of G-force-induced loss of consciousness (GLOC) or spatial disorientation (SD) was long overdue for the F-16 platform (25%), often coupled with the mention of fellow pilots who had lost their lives (23%), as these quotes illustrate:

> I thought it was great. Because it was long overdue and I have a dude, a friend, who controlled flight into terrain while we were deployed and it would have saved his life.

> Absolutely. I've known enough people, and I think pilots, most of us, have known

> enough people that Auto-GCAS would have saved lives. Friends of ours.

About 20% of pilots discussed the value of the system as a backup tool; several pilots expressed that Auto-GCAS has reduced anxiety they experience due to the risk of GLOC and SD. Responses indicated a majority of pilots have a high level of confidence that Auto-GCAS would protect them in an emergency situation. This theme often came up in conjunction with assurances that the system would not alter their flight behaviors, ameliorating some initial design concerns that it might cause pilots to fly more aggressively.

> I would say—it's in the background. I completely trust it. I don't know if there's like trust is the right word. It's kind of there. It isn't like changing anything that I do.

Additionally, about 20% of pilots were initially skeptical of Auto-GCAS. The majority of these comments fell under the theme of "healthy skepticism" to indicate that the system was generally well received, with questions remaining regarding implementation, the possibility of erroneous activations, or uneasiness about a perceived loss of autonomy in the cockpit:

> My initial impression was probably like that with most things, is it sounds like a good idea, but the underlying question would be what's the implementation going to be? Because there are many good ideas for the jet that are not implemented correctly, which then makes them nearly utter failures.

A minority of pilots (8%) characterized their initial perceptions of Auto-GCAS as negative due to an overall distrust of increased automation in the F-16 platform, due to concerns that it would lead to skill degradation over time, thus compromising the safety of all F-16 pilots:

> Initially, yeah, probably a little distrustful. Like every pilot's afraid of the bugs in the software taking over the jet.

The next question within the cluster focusing on opinion development and trust evolution asked pilots to describe how their trust had changed after flying with Auto-GCAS. About 37% of pilots reported that their trust level had not changed after flying with the system on their jets, and 17% of pilots had trust in the system despite faults. Some had heard about or personally experienced some of the system limitations or initial problems with the system:

> Yeah, yeah. I think we're, we are very trusting in our systems until it proves us wrong. But we've had a couple [unwarranted] fly-ups with this and stuff. But other than that, we haven't had any major issues. Fortunately we haven't had to test it because we haven't had any incidents. But I think everybody in the squadron has faith in it.

About 21% of pilots reported their trust in the system had increased due to positive stories or saves:

> [My trust has] moved in the positive direction. Just because I know, I've heard stories of auto fly-ups that have been legitimate saves. So hearing those kinds of stories is like, okay, this is probably a good thing.
>
> I just had not seen the system work. But I know folks who have had saves from the system since it's been incorporated. So that's why I raised my level of trust in the system.

Auto-GCAS Transparency and Pilot Understanding

The next cluster of questions focused on Auto-GCAS transparency and pilot understanding, and pilots were asked what they believe Auto-GCAS does. Most pilots (62%) correctly understood that the system uses DTED, but about 13% of pilots incorrectly believed the system to use a distance-based algorithm:

> It reads DTED data that we put on our DTCs [data transfer cartridges] and then

> it's constantly calculating, predicting your flight path and setting.

> I don't specifically know the parameters, so I'm not even going to try. It'll calculate a pull-up for you based on your altitude to the ground, based on your configuration and it will recover, whether it's a dive or a climb at an unusual attitude.

Most pilots thought the system was developed to save lives (53%) or to save machines and lives (16%):

> Well, we have a lot of GLOCs and ground impacts, controlled flight into terrain or GLOC. So it was developed to save the pilots from those.

When asked how well the system performs or whether there were any aspects of the system they found to be confusing, about 25% reported having "no confusion." About 28% of pilots expressed confusion about the loading of the DTED and understanding whether the system was on or off at any given time, and about 14% reported not understanding whether or not the system was working while over water:

> The one thing I find unintuitive is so when we're flying over the water . . . the indications that we get in the HUD are similar to the GCAS not working. It'll flash and say, "No GCAS." And that to me is confusing. Because obviously we're out over the water. We don't need DTED. But the GCAS is still working, it just doesn't need DTED. So that's kind of confusing.

> The only thing I'd say is that we have issues when we load the DTS [Digital Terrain System] in the JMPS [Joint Mission Planning System] file. I'd say more often times than not, it either gets corrupted or it dumps. So we won't have it anyways. A lot of time it'll just be flashing lights [and show] "No Auto-GCAS" in the HUD for the whole flight because either it's dumped the file or whatever it's corrupted.

Chevrons

The next theme focused on questions relating to the chevrons in the HUD, which serve as a visual indication of an approaching fly-up. Approximately 88% of pilots had observed the chevrons during flight. Those who had observed the chevrons were asked if they had a preference for them being turned on or off during flight, and about 61% reported wanting the chevrons defaulted on, with about 46% liking the feedback they provided indicating the system was working.

> The reason I would prefer to have them is I would rather know, I would rather have one more visual cue that a fly-up is coming rather than just being surprised by it.

When asked whether they thought the chevrons changed how people fly, about 26% said the chevrons did not change how pilots fly, and 8% thought they should not change how people fly, but some indicated the chevrons either had the potential to cause a distraction while flying (9%), caused or had the potential to cause more conservative flying (16%), or had the potential to cause more aggressive flying (8%).

> I don't think they would change the way we fly. Good or bad, I don't think so.

Stories and Reputation

The next question cluster focused on stories, reputation, and trusted information sources for Auto-GCAS. When asked what the reputation of Auto-GCAS was among the pilots in their squadron, about 57% said the system had a favorable reputation. Whether firsthand or secondhand accounts, about 29% had a positive story of the system, and about 23% had a negative story.

> Yeah. We've seen a couple of videos of it already saving guys' lives, down range mostly. Employed in less than ideal conditions.

> The only negative I've heard is these unwarranted fly-ups. But those were really kind of few and far between.

When asked what information sources they trusted most when considering Auto-GCAS, and what sources they would use to seek out information, 47% said they trusted formal sources, such as briefs or academics, and 30% reported informal channels, such as word of mouth. Additionally, 46% of pilots would trust formal subject matter experts, such as experimental, developmental, and operational test pilots, when seeking out information about the system:

> It's one of those systems where you can't really test it yourself. Like I can't point at the ground and see if it's going to save me. But the test guys have tested those parameters and its performance and all that. Like they have the experience. I think that's probably the best source.

Pilots were asked if they had experienced a system activation and, if so, to briefly describe it. About 10% of pilots described activations resulting from a strafe maneuver, although some reported other unwarranted and warranted activations.

> I don't know if you guys have already heard my story? Heard that there was someone in the squadron? Well, that was me, who had the save for that. So I think it's performed fairly well overall.

Usability of Auto-GCAS

Pilots were asked if Auto-GCAS had ever interfered with their ability to perform a mission and, if so, to describe the situation. Most pilots (78%) said that Auto-GCAS had been no nuisance, although about 18% thought that the system might have the potential to become a nuisance or danger, as these quotes illustrate:

> Not yet. I mean, maybe it will one day, but it hasn't yet.

> I'm below the terrain and if I'd rolled out at any time, with the delay, I'm kind of like weaving through canyons. So I was worried that if it rolled out to recover me and it drove straight too long, it might end up hitting something out in front of me when I'm like turning around.

Impact of Auto-GCAS on Behavior

The next question cluster focused on the impact of Auto-GCAS on operational behavior. Pilots were asked if they felt like having an activation would influence their reputation as a pilot or if it would influence their view of a fellow pilot. The widespread response was that it would have no influence, as 59% of pilots reported. Commons themes for why activations would have no influence on reputation included pilots having a strong camaraderie for their fellow airmen and recognizing the benefit of an activation potentially saving a life, as these quotes demonstrate:

> It shouldn't. Any more than any other, I guess, mistake or breakdown in your crosscheck would. That kind of goes with the safety culture of the unit. I mean, nobody is invulnerable to mistakes. So if they happen you debrief to them to make certain they don't happen again. Or at least find a way to do something different next time. But I would say no, it should not affect that pilot's reputation. And it should not prevent them from using the Auto-GCAS.

Additionally, many pilots (30%) reported that having activations and discussing them as a unit would serve as a learning opportunity:

> We all lose SA [situation awareness] at some point. I'd rather lose SA at 20,000 feet than lose it at 500 and it saves me. Lesson learned.

> No, I think in that scenario, it's like, "Hey, good! Well, at least you didn't hit the ground. Now let's talk about why you almost hit the ground."

About 22% of pilots also discussed how a pilot's reputation may be influenced negatively after an activation, depending on the situation or if it was a recurring event:

> If you do that and something got screwed up and you recover from it, now you tell everybody else and we all learn from it. Then it's good. You got lucky, it saved you

whatever on that. But if you continually do it, something's wrong.

DISCUSSION

In this section, we provide explanations of results and their significance, relate the findings to previous research in human automation trust, discuss the limitations of our approach, and make suggestions for future research. The most significant finding of the field study 1st-year results was that Auto-GCAS is an effective life- and aircraft-saving technology and is generally well received and trusted appropriately by the operational pilot community. As a newly deployed system in the military combat environment, rich with operational complexities and challenges, there are a multitude of real-world factors that can positively and negatively influence pilots' trust in it. Key factors that positively influence pilot trust in Auto-GCAS include (a) the design to not interfere with the pilot's ability to perform a mission and operate only in the background; (b) that it "works as advertised," as evidenced by its ability to reliably activate and perform a collision avoidance recovery as expected by pilots; (c) that pilots have a knowledgeable understanding of the return on investment of its business case (i.e., the cost of losing a pilot or aircraft to CFIT far exceeds the cost of its development and implementation); and (d) that operational pilots place great trust in the experimental test pilots in testing and vetting the system prior to fielding. These factors align well with the best practices identified in our case study and by prior research on fostering appropriate trust in automation.

Coexisting with these factors are a number of other factors that negatively affect pilot trust, including unintuitive and confusing Auto-GCAS displays, system errors due to corrupted DTED, and concerns with unwarranted activations (especially during low-angle strafe [LAS] missions). These issues will continue to affect pilot trust in the future and present operational risks; thus, we will monitor and track the likelihood and consequences of these risks in the next few years of the study. Overall, after 1 year of implementation, the benefits of the system (most notably, the pilots and aircraft that the system has saved) far exceed the risks, and the findings on pilot trust evolution (i.e., healthy skepticism, trust informed by saves, attribution of system faults to hardware problems) suggest that the operational pilot community is developing appropriate trust in the system.

Although the above findings corroborate with the conclusion of our case study, there were two interesting and unanticipated findings. First, prior to the field study, the experimental test pilots in our case study commented that the HUD chevrons, designed initially as a test safety feature, could be misused as a terrain-following tool and may encourage aggressive flying behavior. However, some operational pilots in the field study reported that they tend to fly more conservatively because of the chevrons, particularly during LAS passes, which are low-altitude attacks with a short range to the target, having a planned dive angle of 15° or less with a minimum recovery altitude of 75 feet above ground. Although a majority of the pilot participants liked the chevrons as an anticipatory activation cue, and as a tool that enhances their SA of the aircraft's proximity to the ground, it may not be possible to use these results to quantify and perform a trade study of the benefits and undesirable unintended consequences of the chevrons. A viable alternative approach may be to conduct a well-controlled experiment examining the effects of the chevrons in realistic scenarios relevant to actual USAF missions where LAS passes are required, as chevrons appear in the HUD more often during LAS. Such an experiment would generate useful insights or guidelines for designing cues that help promote the system transparency and facilitate appropriate user trust development, thereby helping the operational community.

Second, the finding that the majority of participants did not think an Auto-GCAS activation would influence the reputation of pilots was unanticipated. Prior to the deployment of the system, there were concerns among Auto-GCAS developers (engineers, experimental pilots, and managers/leaders) and the test community that the USAF policy of requiring pilots to report activations may negatively affect their standing in their squadron. Although not all activations are reported, pilots used the activations as valuable opportunities to share lessons learned about their understanding of the newly deployed system and to develop local policies that promote best practices about safety. This finding revealed an important dimension of fighter pilot culture: In high-pace and high-workload military combat environments, camaraderie and brotherhood bond pilots as a cohesive squadron. When there is a warranted activation, they place more value in a fellow pilot's life being saved than how the activation may affect the pilot's reputation. Whether these results would vary if all activations were formally reported and documented is presently unknown.

These unanticipated findings highlight the value of conducting research on trust with technologies used by operators in the real world, because it allows us to discover unknown factors affecting trust development. We may not have conjectured that the HUD's chevrons have the potential to cause both aggressive and conservative flying behavior without following the system's implementation in the real world. Likewise, we may not have hypothesized that pilot culture of camaraderie and brotherhood would trump any negative influences associated with having an activation. These discoveries can be used to drive deeper and better-informed inquiry into how operators develop trust with real systems. As in the case of the chevrons, our findings reveal how factors such as display may affect pilot behavior based on the situation, and in positive or negative ways. Lyons (2013) identified displays/interfaces to be one of the two means to promote system transparency (i.e., shared awareness and shared intent between the user and automation). There is now emerging research on how to build in transparency (via interfaces) for highly autonomous safety systems (such as Auto-GCAS) in a way that supports operator appropriate trust development (Chen & Barnes, 2014; Lyons, Sadler, et al., 2016; Sadler et al., 2016).

Our finding about the significance of pilot camaraderie suggests that the squadron culture has a strong influence on how trust in the system is developed. Of the trust models that include culture as an important factor in the trust calibration process (e.g., Lee & See, 2004), these studies derived and synthesized models from theoretical and well-controlled laboratory experiments. There is a need for gathering data having high external validity to validate and substantiate these models. Results from the 1st year of data collection suggest that the field study methodology is a viable and valuable approach to research human–automation trust. To achieve a sound field study of a system and operators within a complex combat environment, it was required that researchers obtain and leverage organizational support for the project (e.g., from operational stakeholders). By establishing close working relationships with USAF stakeholders, we were able to periodically feed results through organizational channels, which led to rapid solutions and system improvements and sustained project interest and support, particularly through the dissemination of success stories among squadrons.

A key limitation of our study was that the participants were new to Auto-GCAS and had on average 50 hr of flight time with the system. The findings presented are good indicators of pilots' initial trust after 1 year, as their trust and opinions in the system should continue to evolve within the next several years. By continuing to track this evolution, we expect to arrive at a time when trust and opinions become stabilized.

ACKNOWLEDGMENTS

This research was supported by Grant No. FA8650-09-D-6939, 0034, from the U.S. Department of Defense, Air Force Research Laboratory.

REFERENCES

Bernard, H. R. (2011). *Research methods in anthropology: Qualitative and quantitative approaches*. Lanham, MD: AltaMira.

Burns, A., Harper, D., Barfield, A., Whitcomb, S., & Jurusik, B. (2011, October). *Auto GCAS for analog flight control system*. Paper presented at the 30th Digital Avionics Systems Conference, Seattle, WA.

Chen, J. Y. C., & Barnes, M. J. (2014). Human-agent teaming for multi-robot control: A review of the human factors issues. *IEEE Transactions on Human–Machine Systems*, *44*, 13–29.

Defense Science Board. (2016). *Report of the defense science board summer study on autonomy*. Task force report by the Office of the Undersecretary of Defense for Acquisition, Technology, and Logistics, Washington, DC.

DeWalt, K. M., & DeWalt, B. R. (2011). *Participant observation: A guide for fieldworkers*. Lanham, MD: Rowman & Littlefield.

Ho, N., Sadler, G. G., Hoffmann, L. C., Lyons, J. B., & Johnson, W. W. (2017). Trust of a military automated system in an operational context. *Manuscript submitted for publication.*

Hoff, K., & Bashir, M. (2015). Trust in automation. *Human Factors*, *57*, 407–434.

Koltai, K., Ho, N., Masequesmay, G., Niedober, D., Skoog, M., Cacanindin, A., . . . Lyons, J. (2014). *Influence of cultural, organizational, and automation capability on human automation trust: A case study of Auto-GCAS experimental test pilots*. Paper presented at the International Conference on Human Computer Interaction in Aerospace, Santa Clara, CA. doi:20140011822.

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

Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. *Journal of Strategic Information Systems*, *17*, 39–71.

Lyons, J. B. (2013). Being transparent about transparency: A model for human–robot interaction. In D. Sofge, G. J. Kruijff, & W. F. Lawless (Eds.), *Trust and autonomous systems: Papers from the AAAI spring symposium* (Technical Report SS-13-07). Menlo Park, CA: AAAI Press.

Lyons, J. B., Ho, N., Koltai, K., Masequesmay, G., Skoog, M., Cacanindin, A., & Johnson, W. (2016). A trust-based analysis of an Air Force collision avoidance system: Test pilots. *Ergonomics in Design*, *24*(1), 9–12.

Lyons, J. B., Sadler, G. G., Koltai, K., Battiste, H., Ho, N. T., Hoffmann, L. C., . . . Shively, R. (2016). Shaping trust through transparent design: Theoretical and experimental guidelines. In P. Savage-Knepshield & J. Chen (Eds.), *Advances in intelligent systems and computing: Vol. 499. Advances in human factors in robots and unmanned systems* (127–136). Cham, Switzerland: Springer.

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

Niedober, D., Ho, N., Koltai, K., Masequesmay, G., Skoog, M., Cacanindin, A., . . . Lyons, J. (2014). Influence of cultural, organizational and automation factors on human-automation trust: A case study of Auto-GCAS engineers and developmental history. *International Conference on Human–Computer Interaction*, *8512*, 473–484.

Norris, G. (2016). Auto-GCAS saves unconscious F-16 pilot-declassified USAF footage. *Aviation Week & Space Technology*. Retrieved from http://aviationweek.com/technology/auto-gcas-saves-unconscious-f-16-pilot-declassified-usaf-foo tage

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, *56*, 476–488.

Sadler, G. G., Ho, N. T., Battiste, H., Hoffmann, L. C., Lyons, J. B., Johnson, W., . . . Smith, D. (2016, September). *Effects of transparency on pilot trust and agreement in the autonomous constrained flight planner*. Paper presented at the IEEE/AIAA Digital Avionics Systems Conference, Sacramento, CA.

Skoog, M. (2008). *Digital terrain data compression and rendering for automatic ground collision avoidance systems*. Edwards AFB, CA: NASA, Dryden Flight Research Center.

Swihart, D. E., Barfield, A. F., Griffin, E. M., Lehmann, R. C., Whitcomb, S. C., Skoog, M. A., . . . Prosser, K. E. (2011). Design, integration and flight test of an autonomous ground collision avoidance system. *Gyroscopy and Navigation*, *2*(2), 84–91.

Wilkins, M., & Barfield, A. (2016). *Fighter/attack automatic collision avoidance systems business case*. Washington, DC: Office of Personnel Risk and Resiliency, Pentagon.

Nhut Ho is president of NVH Human Systems Integration, LLC, a company that specializes in developing theory and methods to effectively integrate humans and technologies. His interests and projects include autonomy, human trust in artificial intelligence–based systems, modeling, and principles for the design of transparent human–automation teamwork systems. He also conducts case and field studies of operators' acceptance and trust in fielded, complex systems. He obtained his MS and PhD from Massachusetts Institute of Technology.

Garrett G. Sadler is a human factors researcher for NVH Human Systems Integration. With NVH, he studies the role human trust plays in human–machine interaction, automation transparency, and interface design. Collaborating with NASA and Air Force Research Laboratory, he conducts research utilizing a diverse range of both qualitative and quantitative methodologies. He received a BS in mathematics ("pure" math emphasis) in 2008 and an MA in anthropology (cultural) in 2014.

Lauren C. Hoffmann is a human factors researcher for NVH Human Systems Integration. With NVH, she works in collaboration with the Human Autonomy Teaming Laboratory at NASA Ames Research Center and the Human Performance Wing within the Air Force Research Laboratory. She earned her master's degree in experimental psychology in 2015 and her bachelor's in psychology in 2010. She studies cognitive psychology and uses applied statistics to inform human factors engineering research in human–machine interaction.

Kevin Zemlicka is an anthropologist whose research has included studies of U.S. Air Force pilots, Ugandan Pentecostalism in the United States, and public access and usage of beaches and urban parks. He completed his MA in anthropology in 2012 at California State University, Northridge.

Joseph Lyons is the technical advisor for the Human Trust and Interaction Branch within the 711th Human Performance Wing at Wright-Patterson Air Force Base, Ohio. He received his PhD in industrial/organizational psychology from Wright State University in Dayton, Ohio, in 2005. His research interests include human–machine trust, interpersonal trust, leadership, and social influence. From 2011 to 2013 he served as a program officer for the Air Force Office of Scientific Research.

William Fergueson is a behavioral scientist for the 711th Human Performance Wing at Wright-Patterson Air Force Base, Ohio. He enlisted as a backshop avionics maintainer for the B-1 bomber in 2007. He commissioned after receiving his undergraduate degree in psychology from the University of Oklahoma in Norman in 2013. He has spent the past 2 years studying trust in automation with the Automatic Ground Collision Avoidance System study.

Casey Richardson is an F-15 and F-16 experimental test pilot at Eglin Air Force Base, Florida. He flies with the 40th Flight Test Squadron and is also division chief of requirements and plans at the Air Force SEEK EAGLE office. He is a graduate of the U.S. Air Force Academy with a bachelor's in physics. He is also a graduate of U.S. Air Force test pilot school with an MS in flight test engineering.

Artemio Cacanindin is acting director of Flight Systems Integration, 412th Test Engineering Group, Edwards Air Force Base, California. He holds BS and MS degrees in aeronautics and astronautics from the University of Washington. He has over 30 years of flight dynamics and avionics flight test engineering experience on aircraft including the F-15, F-16, B-2, and F-35. His organization conducts all subsystems, structures, performance, flying qualities, airdrop, propulsion, reliability/maintainability, and human factors testing at Edwards Air Force Base.

Samantha Cals is the lead human factors engineer for the 416th Flight Test Squadron at Edwards Air Force Base. Her role focuses on measuring the usability of newly implemented technologies as well as pilot's workload and situation awareness within various aircraft platforms for the U.S. Air Force and its foreign partners. Her main research interests include human–automation teaming, system usability, transparency, and human–automation interaction issues, such as acceptance, compliance, reliance, and trust in automation.

Mark Wilkins serves as a senior aviation safety analyst for the Office of the Secretary of Defense for Personnel and Readiness/Personnel Risk and Resiliency. He advises senior Department of Defense officials on all aspects of military aviation, focusing on emerging aviation safety technologies. He is a retired U.S. Air Force colonel and command pilot with over 3,500 hr flight time, mostly in the F-15C. He is a graduate of the Industrial College of the Armed Forces.