

EFFECTS OF TRANSPARENCY ON PILOT TRUST AND AGREEMENT IN THE AUTONOMOUS CONSTRAINED FLIGHT PLANNER

Garrett Sadler, Henri Battiste, Nhut Ho, Lauren
Hoffmann
NVH Human Systems Integration
Canoga Park, CA
nhut.ho.51@gmail.com

Joseph Lyons
711th Human Performance Wing
Air Force Research Laboratory
Wright Patterson AFB, OH

Walter Johnson, Robert Shively
Human Systems Integration Division
NASA Ames Research Center
Moffett Field, CA

David Smith
Intelligent Systems Division
NASA Ames Research Center
Moffett Field, CA

Abstract—We performed a human-in-the-loop study to explore the role of transparency in engendering trust and reliance within highly automated systems. Specifically, we examined how transparency impacts trust in and reliance upon the Autonomous Constrained Flight Planner (ACFP), a critical automated system being developed as part of NASA's Reduced Crew Operations (RCO) Concept. The ACFP is designed to provide an enhanced ground operator, termed a *super dispatcher*, with recommended diversions for aircraft when their primary destinations are unavailable. In the current study, 12 commercial transport rated pilots who played the role of super dispatchers were given six time-pressured "all land" scenarios where they needed to use the ACFP to determine diversions for multiple aircraft. Two factors were manipulated. The primary factor was level of transparency. In low transparency scenarios the pilots were given a recommended airport and runway, plus basic information about the weather conditions, the aircraft types, and the airport and runway characteristics at that and other airports. In moderate transparency scenarios the pilots were also given a risk evaluation for the recommended airport, and for the other airports if they requested it. In the high transparency scenario additional information including the reasoning for the risk evaluations was made available to the pilots. The secondary factor was level of risk, either high or low. For high-risk aircraft, all potential diversions were rated as highly risky, with the ACFP giving the best option for a bad situation. For low-risk aircraft the ACFP found only low-risk options for the pilot. Both subjective and objective measures were collected, including rated trust, whether the pilots checked the validity of the automation recommendation, and whether the pilots eventually flew to the recommended diversion airport. Key results show that: 1) Pilots' trust increased with higher levels of transparency, 2) Pilots were more likely to verify ACFP's recommendations with low levels of transparency and when risk was high, 3) Pilots were more likely to explore other options from the ACFP in low transparency conditions and when risk was high, and 4) Pilots' decision to accept or reject ACFP's recommendations increased as a function of the transparency in the explanation. The finding that higher levels of transparency was coupled with higher levels of trust, a lower need to verify other options, and higher levels of agreement with ACFP recommendations, confirms the

importance of transparency in aiding reliance on automated recommendations. Additional analyses of qualitative data gathered from subjects through surveys and during debriefing interviews also provided the basis for new design recommendations for the ACFP.

Keywords—*trust in automation; commercial aviation; human-machine interface; automated tools; NASA*

I. INTRODUCTION

In order to enhance safety, improve system performance, and create new capabilities for the National Airspace System (NAS), the National Aeronautics and Space Administration (NASA), industry, and governmental collaborators are envisioning and developing increasingly autonomous systems (IAs) to carry out complicated tasks with minimal human intervention. For example, NASA's Unmanned Aircraft Systems Integration in the NAS (UAS in the NAS) project is developing IAs that will underpin detection and avoidance of potential UAS collision threats [1]; NASA's UAS Traffic Management (UTM) project is currently working to provide guidance and prototypes for IAs that will support the UAS self-configuration, self-optimization, and self-protection of the airspace for low-altitude unmanned vehicles [2]; and finally, NASA's Reduced Crew Operations (RCO) concept is currently working on a next-generation ground control station that will utilize IAs to assist pilots and airline dispatchers to coordinate and execute decisions in off-nominal situations [3]. The work described in the present paper utilized IA components of the RCO ground station to create a high-fidelity environment for experimental scenarios.

While IA systems provide users, organizations, and commercial interests a wide variety of benefits (e.g., increased efficiency, cost savings, reliability, and safety), a particular challenge identified by researchers concerned with the safe and effective integration of IAs into the NAS is the development and operation of these systems in ways that facilitate appropriate operator trust and reliance in automation.

Specifically, operators of systems incorporating IA components must deal with multiple factors that can lead to mistrust or distrust, and thus negatively impact appropriate reliance on these systems. For example, it is known that people often suffer from automation bias, which leads them to expect or believe that the system can do more than it really can; and then when it fails to live up to this expectation they go the other way and cease to rely on it because of mistrust/distrust [4]. Accordingly, a particular challenge identified by researchers concerned with the safe and effective integration of IAs into the NAS is the development and proper aligning of trust and reliance in order to avoid these 'hot/cold' or 'trust/mistrust' swings that impact effective use of these systems.

In what has become a seminal paper on the role of trust in automated systems, Lee and See [5] proposed that trust in automation is appropriate when it is properly calibrated, meaning that there is a one-to-one correspondence between the user's trust level and their understanding of the system's capability. A mismatch between user trust and perceived capability has the potential to lead to inappropriate reliance behavior in the form of misuse or disuse of the system [4]. Misuse occurs when an operator inappropriately relies on the automation due to over-trust (i.e., trust exceeds the system's actual capability). Conversely, disuse occurs when an operator's reliance is guided by under-trust (i.e., the system is more capable than the user trusts it to be). And in both cases, disuse and misuse of automation have led to unfortunate real-world outcomes [5], [6].

To help mitigate the problem of uncalibrated trust leading to inappropriate reliance, researchers have identified automation transparency—defined as a shared awareness and shared intent between a user and system—as a mechanism for engendering trust [7], [8]. Lyons proposes that ideal teaming between humans and automated systems encompasses two complementary aspects of transparency, 1) robot-to-human transparency which conveys to the human how the automation understands the world, the tasks it performs, and how it should perform them, and 2) robot-of-human transparency which conveys to the human what the automation knows about the human operator, e.g., the human's state, preferences, and task/goal awareness. Using this transparency decomposition, systems can be designed to encourage human-automation teaming and a fitting division of labor [9] between the system and user.

The current study focuses on robot-to-human transparency. Chen and colleagues [10] have developed a model for robot-to-human transparency utilizing Endsley's [11] theory of situation awareness (SA) which posits that humans acquire SA at three levels: 1) individual situation elements, 2) interpretation of the situation's elements altogether, and 3) the prediction of the near-term state based on that interpretation. Following the model, Chen et al.'s SA-based agent transparency (SAT) advocates for transparency guidelines that provide information to the user that maintain the three levels of SA. In effect, SAT aims to provide transparency which allows the user to immediately answer these questions: (Level 1 SA) What is the agent doing/attempting? (Level 2 SA) Why is the agent doing that? (Level 3 SA) What should I expect to happen next? They also note that, unlike Endsley's SA model which needs all

three SA levels to achieve optimal situation awareness, the SAT model does not assume that all three levels are needed to achieve transparency. Instead, for the SAT model these levels are simply different aspects that may be differentially needed according to the situation. For example, in a time-sensitive situation, the operator may only need to know the agent's proposed actions (level 1) and the projected outcome (level 3) to make a sufficiently informed decision. Furthermore, it may be that providing reasoning information needed to answer the question "Why?" (level 2) may be harmful because it leads to an unacceptable delay in decision making while the user assimilates this information.

The nature of the link between transparency and trust has been explored theoretically and empirically by multiple researchers [5], [7], [12]-[15]. While the relationship between transparency and trust is usually predicted in theoretical examinations, the empirical work has shown mixed results when attempting to link trust and transparency. Some studies show a significant impact of an automated system's transparency on an operator's trust [16]-[19] while other researchers report transparency level to have no effect on user trust [20], [21].

In the current study we have attempted to further our understanding of transparency as a mechanism for calibrating trust by examining how transparency impacts trust in, and reliance upon, a new automated system developed by NASA. This system, the Autonomous Constrained Flight Planner (ACFP), is being developed as part of NASA's exploration of technologies supporting reduced crew operations (RCO), an effort within NASA's Safe Autonomous Systems Operations project. The RCO effort has explored tools, technologies and concepts that can be used to reduce the number of crew members needed to fly transport aircraft. One idea has been to provide automation that will quickly and efficiently assist in rapidly making decisions for emergency diversion planning in off-nominal situations. This automation, based on the work of Meuleau, Plaunt, Smith, & Smith [22], has been incorporated into the ACFP, which is a ground-based, not aircraft based, tool utilized by a ground operator to provide enhanced support to RCO aircraft in these off-nominal situations.

The ACFP compiles and utilizes information from many different sources to generate a list of diversions. The ACFP uses ATIS information at the airport, current weather, aircraft flight envelope and equipment status, aircraft location and terrain, airport/runway characteristics, and available arrivals and approaches when generating diversions. The main factors that the ACFP takes into consideration are: enroute distance and turns, enroute weather, approach weather and minimums, population density along the approach path (i.e., in case of crash), runway length and width, landing speed (given winds and required approach speed), landing crosswind, runway surface, braking action, and terrain. The weather used comes from a variety of sources, including METAR reports, radar data (storms), turbulence and icing forecasts, and PIREPS, which are all provided by the National Weather Service (NWS). The ACFP takes in all of this information and uses it to generate risk estimates for diversions to available airport runways, and then uses these estimates in turn to make a recommendation. These risk calculations and recommendations

are meant to provide substantial workload savings to operators who, otherwise, would have to gather this information and then take it into account, when utilizing their own expertise to evaluate diversion options. Clearly, trust is critical if operators are to rely on the ACFP in emergency situations, and transparency may be one of the keys to this trust. If distrusted, the operators may fail to utilize the ACFP during emergencies adding workload to an already task-saturated environment.

In this paper, we describe a high-fidelity experiment that was conducted with commercial airline pilots to examine the relationship between transparency, trust, and reliance. Pilots in the experiment interacted with the ACFP in conditions in which the transparency level and risk level of ACFP's recommendation were varied. This study follows earlier work examining trust calibration with the ACFP in a low-fidelity environment [17]. In the sections that follow, we will detail the study's overall design, the task that was carried out by pilots, details of the ground station's operations, key findings regarding pilot behavior in response to our manipulations, followed by a discussion of our results' meanings and implications for designing transparent interfaces.

II. METHOD

A. Participants

Participants were 12 commercial transport pilots who had experience with flight management systems (FMS) and glass-cockpit instrumentation. The study was conducted within the Human Systems Integration Division at NASA Ames Research Center, with pilots being recruited by the San Jose State University Research Foundation from the surrounding San Francisco Bay Area. With the exception of one participant who had 3001-5000 hours of experience, all other pilots had 10,000+ hours of flight experience as line pilots. Additionally, all pilots had previous experience making diversions from flight plans for a variety of reasons including poor weather conditions, air traffic issues, mechanical failure, and/or medical emergencies. All pilots were employed by their airlines as Captains (66.7%) or as First Officers (33.3%), and two-thirds of pilots reported having prior experience with military flying. Most pilots (75%) specified being either "somewhat familiar," "familiar," or "very familiar" with flying in the simulated geographical area in the study (Colorado, Utah, and Wyoming).

B. Experimental Design

For this study we employed a within-subjects 3 x 2 (Transparency x Risk) factorial design with three levels of Transparency (Baseline, Value, and Logic), and two levels of Risk (Low and High).

Six distinct scenarios were created, and all pilots saw each of the scenarios, for a total of six trials. Across these six trials each pilot was shown each of the 3 levels of Transparency twice, but was never shown the same Transparency level more than once for any one of the six scenarios. The Transparency level corresponded to the type of explanation for the diversion recommendation provided to the pilots. In all three levels of Transparency the pilots were able to call up and view all of the

information that the automation used to derive its recommendation. The three levels differed in that no explanation for how the automation arrived at its recommendation was provided in the Baseline condition, while the calculated success probability that ultimately drove the diversion recommendation was provided in the Value condition, and an additional explanation detailing the link between these probabilities and the information used to derive the recommendations was provided in the Logic condition. Thus we used an additive manipulation of transparency, consistent with similar methods in prior research [19].

All scenarios took place within a region roughly encompassing Colorado and parts of immediately surrounding states. The scenarios were generated by, among other things, combinations of landing conditions (e.g. runway length and surface conditions, crosswinds, etc.), en route and arrival convective weather, aircraft type, and aircraft fuel. These factors were used to generate three High Risk and three Low Risk diversions for six aircraft within each scenario. The Risk factor was a within-scenario manipulation. In particular, each scenario was designed such that the best diversion option for three of the aircraft had a probability of a successful landing between 0.8 and 1.0 (Low Risk), while the best diversion option for the other three had a probability of a successful landing between 0.4 and 0.65 (High Risk). There was never a better option available in these cases as determined by the ACFP risk algorithm. Therefore, for the High Risk aircraft, while there was a substantial risk in the recommended diversion, it was the best that the ACFP risk algorithm could find, with all other alternatives having higher risk; while for the Low Risk aircraft there might be other high and low risk alternatives, but the recommended diversion had the lowest risk.

Each pilot engaged in a total of six scenarios. Both instances of each Transparency condition were presented back-to-back, yielding three blocks. Each of the six potential orderings of the three blocks were presented to two pilots, thus providing counterbalancing for any order effects in presenting Transparency conditions. Risk was a within-scenario manipulation and counterbalancing was handled by requiring pilots to reroute their six aircraft in a prescribed order, with this order manipulated to counterbalance for Risk order effects. Across all participants and trials each of the six scenario was paired with each of the three Transparency conditions exactly 4 times.

C. Task/Apparatus

This study used a simulation environment designed to support research within the NASA Ames Human Systems Integration Division addressing issues in air-ground integration for proposed future air traffic management systems. The current study built upon capabilities that were designed to enable exploration of concepts for the ground support of single-piloted aircraft, an effort which was part of the NASA's Reduced Crew Operations (RCO) project [23]. In the current study an operator at a simulated advanced ground station monitored and produced diversions for aircraft. Six principal components of the whole prototype ground

station were utilized; this subset of the functionalities included: the Autonomous Constrained Flight Planner (ACFP) recommender system (see Fig. 1), an Aircraft Control List (ACL), Automatic Terminal Information Service (ATIS) broadcasts, a Traffic Situation Display (TSD), Federal Aviation Administration (FAA)-issued approach plates and airport charts, and pop-up windows that contained evaluations of specific diversions provided by the ACFP.

Researchers in NASA's RCO project designed the ACFP as a tool to support flight path monitoring and re-routing by incorporating an Emergency Landing Planner (ELP) algorithm previously developed at NASA [22] and most recently utilized in a precursor study to the present experiment [17]. The ACFP provided automated diversion recommendations and evaluations during the study's landing scenarios by evaluating the risks of going to airport runways within 200 nautical miles (370.4 kilometers) of an aircraft. The top recommendation and route, plus any transparency information were then presented on a pop-up window on the TSD. The system would also give risk, route, and transparency information for any other airport runway upon request by the pilot. The TSD provided pilots with a visual display of the geographic area including convective weather cells and turbulence boxes, as well as icons representing the locations of available airports. Additionally, the TSD displayed information related to each aircraft's current state, including altitude, heading, indicated airspeed, and location. By using the ACL, pilots were able to shift focus between the six simulated aircraft in the TSD and see the type of the selected aircraft (e.g., Airbus A320, Boeing 747, etc.). By accessing a menu in the TSD, local airport weather conditions were also available to participants in the form of the ATIS broadcast for the corresponding landing site. Pilots were able to look up the approach plate for each available approach, which provided a schematic diagram of the approach as well as the ceiling and visibility minimums for the approach.

Each scenario required pilots to land all aircraft under their control (6 per trial). Following examples set forth in [17], the ACFP window presented information to the pilot using one of the three hierarchical levels of transparency. In the Baseline Transparency condition (Fig. 2, top), participants were provided with a recommendation from ACFP displaying the recommended landing site. This included the airport and runway number, runway length (in feet), approach name/type, and the distance to the landing site (in nautical miles). In the Value Transparency condition (Fig. 2, middle), pilots were

provided with a "risk statement" in addition to the information presented in the Baseline Transparency condition. This statement provided the evaluation from ACFP of the probability of success for landing on the first attempt (e.g., "There is a 55% chance that you will be able to successfully complete the approach and landing under current conditions"). The success probabilities refer to the chance of having to perform a "go-around" or missed approach. The Logic Transparency condition (Fig. 2, bottom) included information from the Baseline and Value conditions, as well as additional statements to explain the rationale behind the ACFP recommendation. The statements contained descriptions of relevant factors along phases of flight that led to the determination for the recommendation, such as information about the enroute, approach, or landing phase.

D. Measures

Trust was measured using a 7-item scale to gauge pilot's intentions to be vulnerable to, that is 'trust', the ACFP [17]. Pilots rated their agreement with the items using a 7-point Likert scale. Trust measures were taken after each Transparency condition and the scale evidenced high reliability with alphas ranging from 0.88-0.92. Example items included: "I think using the [ACFP] will lead to positive outcomes," "I would feel comfortable relying on the recommendations of the [ACFP] in the future," and "when the task was hard I felt like I could depend on the [ACFP]."

Behavioral data was recorded and stored in timestamped system log files. A sample of the type of data recorded includes: when pilots toggled focus on an aircraft, when ACFP queries were made, when and what diversions were executed, requests for airport ATIS data, and when and what approach charts were viewed. Three main types of behavioral/performance measures were derived from the recorded data.

1) *Verification*: Did the pilot look at the ATIS data or approach plates upon which the ACFP based its recommendations?

2) *Exploration*: Did the pilot request ACFP evaluations and routings or ATIS data for any other airport than the one initially recommended by the ACFP?

3) *Agreement*: Did the pilot execute the diversion recommended by the ACFP?



Fig. 1. Labeled image of the ground control station used by participants.

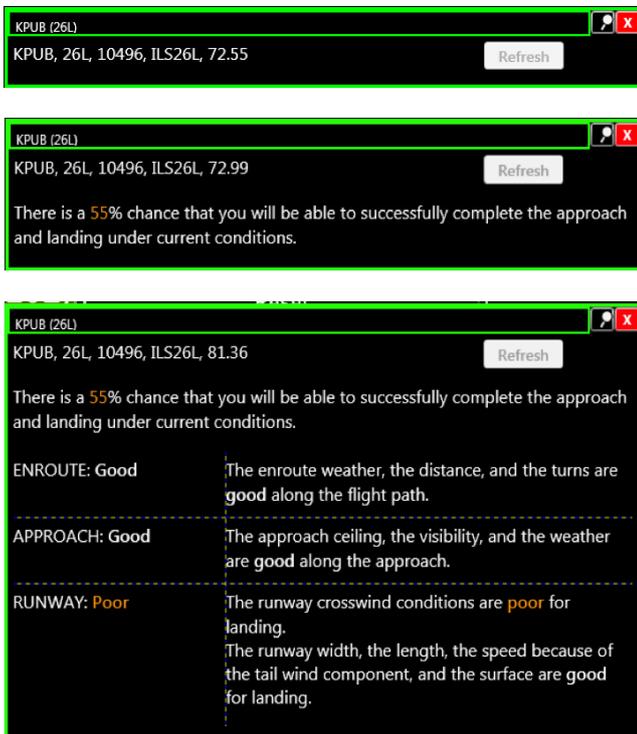


Fig. 2. Screen capture of ACFP query result in the Baseline (top), Value (middle), and Logic (bottom) transparency conditions.

III. RESULTS

A. Trust

Pilots' trust increased with higher levels of Transparency. As also reported by Lyons and colleagues [18] in a separate

examination of data from this study, there was a statistically significant effect of Transparency on trust, $F(2, 22) = 4.39$, $p < 0.05$, with highest trust in the Logic condition and lowest trust in the Baseline condition (Figure 3).

For each pilot there were six replications of the six conditions defined by the 3 (Transparency - Baseline, Value, Logic) x 2 (Risk - Low or High) design. These replications were converted to probabilities for the three behavioral measures: verification, exploration, and agreement. This was done by simply taking the number of times the pilots verified a recommendation, explored new options, or agreed with a recommendation, and dividing this total by six. Initial examination of this data revealed it to have substantially non-normal distributions due to a strong positive skew in the probabilities (e.g., many probabilities of 1.0). When tests of normality were conducted on the behavioral measures for each of the six conditions, the tests rejected normality ($p < .05$) in two of six conditions for the verification data, in three of six conditions for the exploration data, and in five of six conditions for the agreement data. Based on this we decided to utilize non-parametric tests in our analyses.

B. Verification

Pilots were more likely to verify recommendations from the ACFP that had low levels of Transparency, and also when the recommendation had a High Risk (see Fig. 4, top). In the Low Transparency Baseline scenarios pilots were more likely to verify the ACFP recommendation (62%-65% of the time) compared to High Transparency Value and Logic scenarios (28% - 44% of the time). When a Friedman paired sample test was used to evaluate this transparency effect it was found to be statistically significant ($p = 0.015$). Pilots were also more likely to verify ACFP recommendations when they were High Risk (49%) versus Low Risk (43%), but, a Wilcoxon signed rank test did not find this effect significant ($p > 0.10$). A closer examination of the means showed that an effect of risk might have been confined to the two higher levels of Transparency, but Wilcoxon signed rank tests that examined the effects of risk within only the Value and the Logic conditions failed to confirm this ($p > 0.10$).

C. Exploration

Pilots were less prone to explore non-recommended diversion options for Low Risk aircraft within the two higher Transparency conditions (Value and Logic), than for Low Risk aircraft in the Baseline. On the other hand, for High Risk aircraft the tendency to explore other options was high, and approximately equal, in all three Transparency conditions (Figure 4, middle). A Friedman paired sample test showed no statistically significant main effect of Transparency ($p > 0.10$), but a Wilcoxon paired sample test did show a significant main effect of Risk ($p = 0.006$). However, there is a clear indication of an interaction where the effect of Risk is confined to the High Transparency condition. This was confirmed with separate follow-up Friedman tests examining the effect of Transparency for the High and Low Risk aircraft. These found that there was no effect of Transparency for the High Risk aircraft (Baseline = 73%, Value = 76% Logic = 78%, $p > 0.10$), but a significant effect of Transparency for the Low Risk aircraft (Baseline = 74%, Value = 38%, Logic = 37%,

$p=0.018$). As with verification, higher levels of Transparency at the lower risk level led pilots to take the recommendations with less follow-up checking. However, unlike the tendency for verification to decrease at the higher levels of Transparency when risk was high, the pilots continued to explore other options across all levels of Transparency for planes given a higher risk recommendation.

D. Agreement

The probability that pilots would chose to land on the same runway as that recommended by the ACFP increased as a function of Risk, and as a function of Transparency for the two higher levels of Transparency (Value and Logic, see Fig. 4, bottom), although there was no evidence of an increase in the Baseline condition. A Wilcoxon paired sample test found the main effect of Risk to be statistically significant ($p=0.046$), while a Friedman paired sample test found no statistically significant effect of Transparency ($p > 0.10$).

IV. DISCUSSION

This study envisioned a future concept of operations in which a ground operator is charged with a hybrid duty composed of present day airline dispatcher and pilot first officer roles. In this study we examined this concept by placing pilots in a high-fidelity environment and requiring them to execute diversion decisions under time pressure for multiple aircraft. In a real-world setting, the ground operator would be in communication with the pilot in the air as well as with air traffic controllers (ATCs) in order to coordinate decisions. Because this study had transport pilots (i.e., not dispatchers) performing the role of a ground operator who has executive authority (i.e., makes diversion decisions without coordinating with pilots or ATCs), the simulation described here did not fully examine the trust and behaviors of a hypothetical ground operator, a civil aviation position that does not yet exist. As such, this study was only a first look at possible ground operator behavior for subcomponents of the ground operator's tasks and responsibilities. Despite these differences from the proposed real-world position, this is a critical extension of prior trust-transparency research because of the use of actual operators, a high-fidelity simulation, and ecologically-valid constraints. Further, this study extends prior trust-transparency work [17], [18] with the inclusion of reliance metrics. Notably, the impact of reliance was evaluated as a function of both the level of risk associated with a decision as well as the level of transparency associated with the automation recommendation. The remainder of this discussion will consider the findings as they related to pilot verification and exploration, or *Information Seeking*; pilot decisions on where to divert an aircraft, or *Flight Decisions*; and *ACFP Design*.

A. Information Seeking

The Baseline conditions showed no trends associated with Risk, with pilots verifying 61%-65% of the time, and exploring 73%-74% of the time. This is probably not surprising since no risk evaluation was provided for Baseline, but it does suggest that the pilots were insensitive to the risk that the ACFP was identifying (but not displaying). Otherwise, one could expect the tendency to explore or verify to have been affected. On the other hand there were clear effects associated with Risk in the

Value and Logic conditions. For the Low Risk aircraft, the tendency to explore and verify were much less than in the Baseline condition, but otherwise similar, with pilots verifying 28%-34% of the time, and exploring 37%-38% of the time. The decrease shows that the pilots invested much less time and effort if they were shown a low risk rating by the ACFP. This in and of itself is also not that surprising, particularly given that they were under time pressure and knew that they would have High Risk aircraft to which they would have to attend. The lack of any significant differences between the Value and Logic conditions indicates that the Logic statements played little or no role in the pilots' decisions to verify these risk statements, or explore for better options

When we turn our attention to how pilots dealt with High Risk aircraft in the Value and Logic scenarios we see a slightly different picture. Verification behavior does go up, 41%-44% (again no real difference between the two), but remains substantially below the 65% rate for the High Risk aircraft in the Baseline scenario. It seems that pilots still had a fair degree of trust in the risk estimate rendered by the ACFP, and again the impact of the Logic statements played little or no role in pilots' decisions to verify. On the other hand, exploration behavior for High Risk aircraft in the Value and Logic scenarios is much higher (76%-78%) and not statistically different from that found in the Baseline scenario (73%). This change in exploration behavior as a function of aircraft Risk within both the Value and Logic scenarios, when compared with the lack of such an effect for verification behavior within these scenarios, highlights an important distinction between verification and exploration. That is, it seems that despite the apparent trust that the pilots had in the ACFP to make a decent estimation of risk, this did not translate into trust that the ACFP could find the best alternative destination. In particular, one post-hoc hypothesis supported by this data is that if the pilots had trust in the initial risk estimate, and risk was low, then they could just go on to accept that recommendation; but if risk level was high, then they felt compelled to look for a better option. But this brings up the question of why this should be the case. Further research is indicated.

B. Flight Decisions

Finally, we need to turn our attention to how the Transparency and Risk factors affected whether the pilots actually took the ACFP diversion recommendation. This data was the most difficult to analyze because of overall high rates agreement, ranging between 73% and 95%. For example Low Risk aircraft in the Baseline, Value and Logic scenarios, generated particularly skewed data with, respectively, 5, 8, and 7 out of 12 pilots agreeing with the recommended alternative 100% of the time; while the numbers for the High Risk data were 5, 6, and 2 out of 12. Despite this, the non-parametric analysis of the agreement data showed a significant effect of Risk on decision making, with pilots' diversion choices agreeing with the recommended diversions less often for the High Risk aircraft (73%-86%) than for the Low Risk aircraft (79%-95%). While we found no significant impact of Transparency level, the data suggest that this effect of Risk was confined to the Value and Logic scenarios, an interaction that our non-parametric analysis was ill-equipped to detect. An inspection of Figure 4 (bottom) shows this interaction,

revealing no effect of Risk on agreement in the Baseline scenarios (both 21%), and with agreement going down as risk goes up for the Value and Logic scenarios. This pattern can be understood in light of the pilots' information seeking behavior, where verification and exploration behaviors increased as risk increased. Specifically, a post-hoc hypothesis suggested here is that pilots' confidence in the solutions being presented by the ACFP decreased with higher risk and this led to a higher rejection of the recommendations. This of course would not be represented in the Baseline condition due to the already noted insensitivity of pilots to diversion risk in the absence of an ACFP risk evaluation.

C. ACFP Design

The design for transparency of the ACFP, specifically in the Logic condition, and in overall ground station components, represents a practical implementation of the SA-based Agent Transparency concept advocated by [10]: Level 1 SA is provided by the TSD, ACL; Level 2 SA is provided by the reasoning statements returned from ACFP for the enroute, approach, and landing phases of flight; and Level 3 SA is provided by the probability statement for successful landing on the first approach. The automation's reasoning or diversions additionally represents a partial implementation of Goritzlehner et al.'s [21] recommendation that proper automation transparency involves the system revealing its inner workings.

From semi-structured interviews conducted following the simulation, pilots in the experiment provided feedback and suggestions for improvement on the ACFP interface design. Most common among pilots' comments was the desire for reduction of workload by reducing the number of menus one has to access in order to obtain information required to execute decisions. In order to properly evaluate a recommendation given by ACFP, pilots would have to access menus to retrieve ATIS broadcasts in the local area and review approach plates for ACFP's suggestion as well as possible alternatives. Many

pilots expressed a desire that ACFP coordinate with relevant components of the ground station so that 1) all ATIS broadcasts/METAR data for local airports be displayed continually and 2) that the airport and approach plate charts for the ACFP recommended landing site automatically be returned alongside the ACFP output. Additionally, during pilot testing of our experimental scenarios prior to running our study, two subject matter experts indicated that they would like ACFP to be customizable so that it weighted certain factors (e.g., crosswinds, approach weather, airport facilities, etc.) more highly than others according to either personal preferences or situational need. Pilots were also asked about their decision-making strategies when executing diversion decisions during actual flight. Pilots primarily reported that they do not have a generalized strategy for diversion planning in emergency situations due to the fact that diversion decisions must be situation-specific. Nevertheless, top factors reported to be important in an emergency include distance to the landing site and whether the diversion airport has services and facilities for their airline.

D. Conclusion

Our finding that transparency played a significant role in trust in and agreement with ACFP diversion recommendations allows us to build upon design elements for displays that encourage appropriate behavior while lightening the workload of an aviation professional partly responsible for the safety of multiple aircraft.

One particular noteworthy finding is that pilots seemed to accept risk evaluations, but for some reason this did not translate into an acceptance that the automation had found the best alternative, particularly when the risk was high. Clearly there was something more complex going on in their evaluations. Reliance, and by implication trust, therefore must be seen as a more nuanced attribute.

This aspect of the system can be improved to better inform the user of the logic and decision making model employed by the ACFP, as is evidenced by one pilot suggesting, during his post-simulation interview, a decision model that is effectively equivalent to the one actually employed by the ACFP.

ACKNOWLEDGMENT

The authors would like to thank Mr. Dominic Wong and Mr. Christian Plaunt for their help with the ACFP software. Support for this work was jointly provided by the NASA Safe Autonomous Systems Operation Program (SASO) and the United States Air Force Research Laboratory, through an interagency agreement (NASA SAA2-402995-1/USAF Alliance No. 15-160-HPW-02MOA).

REFERENCES

- [1] C. R. Rorie, L. Fern, and J. Shively, "The impact of suggestive maneuver guidance on UAS pilot performing the detect and avoid function," in *AIAA Infotech @ Aerospace*, ed: Amer. Inst. of Aeronautics and Astronautics, 2016.
- [2] Nat. Aeronautics and Space Admin. (2015, December 1). *Unmanned Aircraft System (UAS) Traffic Management (UTM)* [Online]. Available: <https://utm.arc.nasa.gov/index.shtml>
- [3] S. V. Ligda, U. Fischer, K. Mosier, M. Matessa, V. Battiste, and W. W. Johnson, "Effectiveness of advanced collaboration tools on crew communication in reduced crew operations," in *Eng. Psychology and Cognitive Ergonom.: 12th Int. Conf.*, Los Angeles, CA, 2015, pp. 416-427.
- [4] R. Parasuraman and V. Riley, "Humans and automation: use, misuse, disuse, abuse," *Human Factors*, vol. 39, pp. 230-253, 1997.
- [5] J. D. Lee and K. A. See, "Trust in automation: designing for appropriate reliance," *Human Factors*, vol. 46, pp. 50-80, 2004.
- [6] J. B. Lyons and C. K. Stokes, "Human-human reliance in the context of automation," *Human Factors*, vol. 54, pp. 112-121, 2012.
- [7] J. Y. Chen and M. J. Barnes, "Human-agent teaming for multirobot control: a review of human factors issues," in *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 1, pp. 13-29, Feb. 2014.
- [8] J. B. Lyons, "Being transparent about transparency: a model for human-robot interaction," in *2013 Assoc. for the Advancement of Artificial Intell. Spring Symp.*, Palo Alto, CA, 2013, pp. 48-53.
- [9] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," in *IEEE Trans. Syst. Man Cybern. A, Syst. Humans*, vol. 30, no. 3, pp. 286-297, May 2000.
- [10] J. Y. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia, and M. Barnes, "Situation awareness-based agent transparency," Army Research Lab., Aberdeen, MD, Rep. ARL-TR-6905, Apr. 2014.
- [11] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, pp. 32-64, 1995.
- [12] J. Y. Chen, M. J. Barnes, and M. Harper-Sciarini, "Supervisory control of multiple robots: human-performance issues and user-interface design," in *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, vol. 41, no. 4, pp. 435-454, Jul. 2011.
- [13] K. A. Hoff and M. Bashir, "Trust in automation: integrating empirical evidence on factors that influence trust," *Human Factors*, vol. 57, pp. 407-434, 2014.
- [14] B. M. Muir, "Trust between humans and machines, and the design of decision aids," *Int. J. Man-Machine Studies*, vol. 27, pp. 527-539, 1987.
- [15] F. M. Verberne, J. Ham, and C. J. Midden, "Trust in smart systems: sharing driving goals and giving information to increase trustworthiness

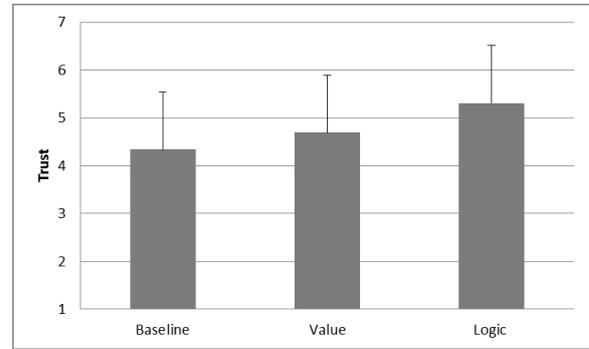


Fig. 3. Means for trust by condition (taken from [18]).

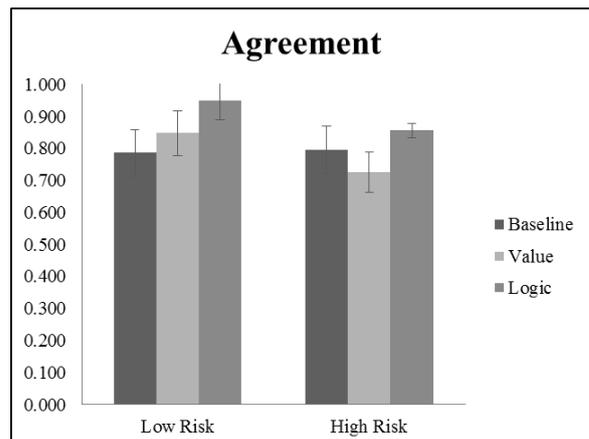
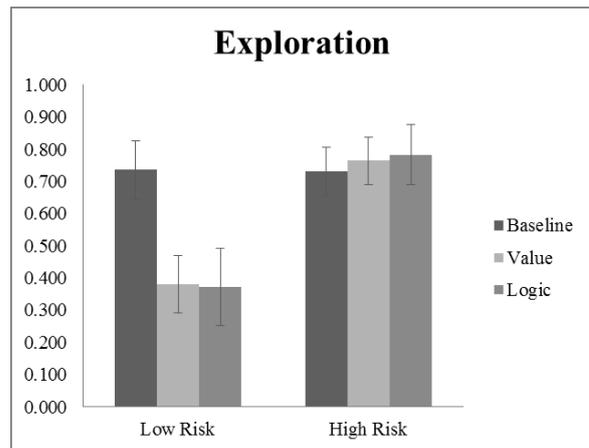
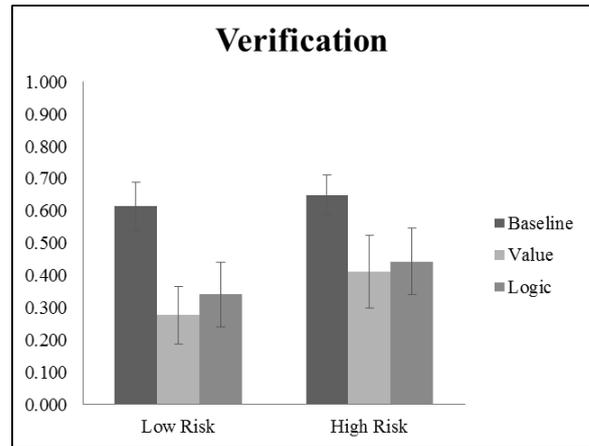


Fig. 4. Mean probabilities for ACFP recommendation verification (top), exploration (middle), and agreement (bottom) by transparency condition and risk level.

- and acceptability of smart systems in cars,” *Human Factors*, vol. 54, pp. 799-810, 2012.
- [16] T. Helldin, “Transparency for future semi-automated systems: effects of transparency on operator performance, workload and trust,” Ph.D. dissertation, Technology, Orebro Univ., Orebro, Sweden, 2014.
- [17] J. B. Lyons, K. S. Koltai, N. T. Ho, W. Johnson, D. E. Smith, and J. R. Shively, “Engineering trust in complex automated systems,” *Ergonom. Design*, vol. 24, pp. 13-17, 2016.
- [18] J. B. Lyons, G. G. Sadler, K. Koltai, H. Battiste, N. T. Ho, L. C. Hoffmann, D. Smith, W. Johnson, and R. Shively, “Shaping trust through transparent design: theoretical and experimental guidelines,” in *Proc. of the Appl. Human Factors and Ergonom. Conf.*, Orlando, FL, Jul. 2016, in press.
- [19] J. E. Mercado, M. A. Rupp, J. Y. Chen, M. J. Barnes, D. Barber, and K. Procci, “Intelligent agent transparency in human-agent teaming for multi-UxV,” *Human Factors*, in press.
- [20] H. Cramer, V. Evers, S. Ramlal, M. van Someren, L. Rutledge, N. Stash, L. Aroyo, and B. Wielinga, “The effects of transparency on trust in and acceptance of a content-based art recommender,” *User Model User-Adap. Inter.*, vol. 18, pp. 455-496, 2008.
- [21] R. Goritzlehner, C. Borst, J. Ellerbroek, C. Westin, M. M. van Paassen, and M. Mulder, “Effects of transparency on the acceptance of automated resolution advisories,” in *2014 IEEE Int. Conf. Syst., Man, and Cybern.*, San Diego, CA, 2014, pp. 2965-2970.
- [22] N. Meuleau, C. Plaunt, D. Smith, and C. Smith, “Emergency landing planner for damaged aircraft,” in *Proc. 21st Innovative Applicat. of Artificial Intell. Conf.*, pp. 114-121, 2009.
- [23] S. Brandt, J. Lachter, V. Battiste, W. Johnson, “Pilot situation awareness and its implications for single pilot operations: analysis of a human-in-the-loop study,” *Procedia Manufacturing*, vol. 3, pp. 3017-3024, 2015.