Shaping Trust Through Transparent Design: Theoretical and Experimental Guidelines

Joseph B. Lyons, Garrett G. Sadler, Kolina Koltai, Henri Battiste, Nhut T. Ho, Lauren C. Hoffmann, David Smith, Walter Johnson and Robert Shively

Abstract The current research discusses transparency as a means to enable trust of automated systems. Commercial pilots (N = 13) interacted with an automated aid for emergency landings. The automated aid provided decision support during a complex task where pilots were instructed to land several aircraft simultaneously. Three transparency conditions were used to examine the impact of transparency on pilot's trust of the tool. The conditions were: baseline (i.e., the existing tool interface), value (where the tool provided a numeric value for the likely success of a particular airport for that aircraft), and logic (where the tool provided the rationale for the recommendation). Trust was highest in the logic condition, which is con-

J.B. Lyons (✉)
Air Force Research Laboratory, Wright-Patterson AFB, Dayton 45433, OH, USA
e-mail: joseph.lyons.6@us.af.mi

G.G. Sadler · K. Koltai · H. Battiste · N.T. Ho · L.C. Hoffmann
NVH Human Systems Integration, Canoga Park, Los Angeles, CA, USA
e-mail: garrett.g.sadler@gmail.com

K. Koltai
e-mail: kolina.koltai@gmail.com

H. Battiste
e-mail: hbattiste@gmail.com

N.T. Ho
e-mail: nhut.ho.51@gmail.com

L.C. Hoffmann
e-mail: lauren.c.hoffmann@gmail.com

D. Smith · W. Johnson · R. Shively
NASA Ames Research Center, Moffett Field, Los Angeles, CA, USA
e-mail: david.smith@nasa.gov

W. Johnson
e-mail: walter.johnson@nasa.gov

R. Shively
e-mail: robert.shively@nasa.gov

P. Savage-Knepshield and J. Chen (eds.), *Advances in Human Factors in Robots and Unmanned Systems*, Advances in Intelligent Systems and Computing 499,
DOI 10.1007/978-3-319-41959-6_11

sistent with prior studies in this area. Implications for design are discussed in terms of promoting understanding of the rationale for automated recommendations.

Keywords Trust · Transparency · Automation

1 Introduction

Advanced technology has great promise to support improved task performance across a variety of domains. Yet, advances in technologies such as automation, while beneficial to performance in stable (high-reliability) states, can have detrimental effects when they fail [1]. One paradoxical reason why automation can be devastating is that humans may form inappropriate reliance strategies when working with automation [2, 3]. Thus, the issue of trust in automation has emerged as an important topic for human factors researchers [4, 5]. Trust is a critical process to understand because trust has implications for reliance behavior—i.e., using or "relying" on a system when that reliance matters most. The trust process as it relates to automation is complex because the factors that influence trust range from human-centric factors such as dispositional influences (e.g. predisposition to trust) and experiential influences (learned trust aspects), to situational features [see 5 for a recent review]. Failure to establish appropriate trust can result in performance errors due to over-trust in technology where a human places unwarranted reliance on a technology, or alternatively, humans can under-trust technology by failing to use technology when that reliance is warranted. One key for researchers is to identify the set of variables that influences the trust process and to provide humans with the appropriate information to drive appropriate reliance decisions. The current paper discusses one such influence, the role of transparency and its influence on the trust process by presenting experimental data related to different transparency manipulations in a high-fidelity, immersive commercial aviation task environment involving automation support to a pilot.

Transparency represents a method for establishing shared awareness and shared intent between humans and machines [6]. Transparency is essentially a way for the human and the machine to be on the same page with regard to goals, processes, tasks, division of labor within tasks, and overall intent-based approach toward the interaction. Lyons [6] outlines several dimensions of transparency: intent, environment, task, analytic, team, human state, and social intent. The intent dimension involves understanding the overall purpose of the technology and how well this purpose matches the expectations of the human partner. Human expectations can be driven by naming schemes, physical appearance or other symbols, as well as by descriptions of the technology and prior experiences with similar technologies. The environment component involves educating the human (either through training or real-time display features) about how the technology senses information in the environment. The task dimension involves communicating the technology's limitations, capabilities, and task completion information to the human. The analytic

dimension involves sharing details about the rationale for behaviors taken or recommendations provided by the system as well as providing the human with an understanding of the programming of the technology (i.e., "how it works"). The team component involves understanding the division of labor between the human and the technology. The human state dimension involves communicating information about the human operator (e.g., stress, fatigue) to the technology. Finally, the social intent facet of transparency involves communicating information to the human regarding the planned pattern of interactions (e.g., style, timing, etiquette, etc.) between the human and the technology.

Previous research has found that transparency is beneficial to humans interacting with automated decision aids [7–9]. Transparency in these contexts has been shown to influence trust by conveying necessary information about the limitations, logic, or intent of a system. Transparency has also been explored in the context of automation for commercial aviation. Lyons and colleagues [9] systematically manipulated different levels of transparency in NASA's Emergency Landing Planner tool (ELP) [10]. The ELP was designed as an automated aid to support rapid decisions for commercial pilots to support effective diversion decisions. They sought to examine the potential benefits of added rationale for recommendations provided by the ELP by creating two additional display features to augment the existing ELP infrastructure. The first feature, termed value, added a numeric value reflecting the calculated probability for a successful landing for that particular diversion airport on the first attempt (i.e., without requiring a "go-around"). This subtle (but in no way simple) calculation was believed to increase the credibility of an option and provide the pilots with a quick estimate on the feasibility of a landing option. However, to make the ELP more transparent to the pilots, the study authors added a second feature which explained the rationale, termed logic, for the recommendation. This added information communicated the reasons why this diversion airport was a good or bad option. Using a static, low-fidelity task scenario, the authors found that trust was rated highest when the pilots were given the highest level of transparency for the ELP (i.e., the logic condition). It was unknown however, whether these same benefits would transfer to a more realistic task environment characterized by complexity and time constraints. The same design principles used in the aforementioned study (i.e., value and logic transparency) were used as a template for the current study, though the current study used a high-fidelity task environment to examine the effects of transparency on trust.

2 Method

2.1 Participants

The participants were 13 commercial transport pilots experienced with glass-cockpit instruments and with flight management systems (FMS). Participants were recruited locally from the San Francisco Bay Area through the San Jose State

University Research Foundation in conjunction with the Human Systems Integration Division at NASA Ames Research Center. They all had over 10000 h of flight experience as line pilots with the exception of a single participant who had between 3001 to 5000 h of experience. All participants had real-world experience making diversions from their filed flight plans for a variety of reasons including bad weather, traffic issues, mechanical failure, and/or medical emergencies. Participants were either employed by their airlines as Captains (66.7 %) or as First Officers (33.3 %). Two-thirds of participants had prior military flying experience. A majority of pilots (75 %) indicated that they were either "somewhat familiar," "familiar," or "very familiar" with flying in the study's simulated geographical area (Colorado-Utah-Wyoming).

2.2 Experimental Design

We used a within subject factorial design with three levels of transparency. The levels of the Transparency corresponded to providing the participant with no explanation (baseline) for the automation's diversion recommendation, just success probability (value), and success probability plus explanation (logic) for the automation's recommendation. This additive manipulation of transparency facets is consistent with similar methods in prior research [7].

Six experimental scenarios were constructed with six aircraft in each scenario, and presented to the participants in a singular fixed order. Each scenario was designed such that the best available landing options afforded a high success probability to three of the aircraft, but only a low success probability to the other three aircraft. The order in which the aircraft diversions occurred was experimentally prescribed for each of the six scenarios such that, when collapsed over participants, each scenario had an equal number of landings affording high and low Success Probability. Finally, the order of presentation of the transparency conditions was also counterbalanced. Each Transparency condition was presented in blocked fashion, with three blocks and two scenarios per block. This provided six potential block orderings, with each of these orderings given to at least two participants.

2.3 Task/Apparatus

A dynamic commercial simulation environment was used for the current study in which an operator at an advanced ground station monitored and produced diversions for aircraft. This study utilized a subset of the functionalities of the whole prototype ground station, specifically six principal components: a Traffic Situation Display (TSD), an Aircraft Control List (ACL), Automatic Terminal Information

Service (ATIS) broadcasts, FAA-issued approach plates and airport charts, and pop-up windows containing evaluations of specific diversions provided by the Autonomous Constrained Flight Planner (ACFP) recommender system, and the ACFP itself (see Fig. 1). The ACFP is a tool being designed to support flight path monitoring and re-routing for NASA's Reduced Crew Operations (RCO) project [11], and which directly incorporated the ELP algorithm [10], served as the automated diversion recommendation aid during a complex landing scenario. Each of these diversions specified a runway at a specific airport, along with the route to that runway. The TSD provided participants with a visual display of the geographic area, convective weather cells, turbulence boxes, icons representing the locations of available airports, and information related to each aircraft's current state: location, heading, altitude, and indicated airspeed. Using the ACL, participants were able to toggle focus between the six simulated aircraft in the TSD and look up the selected aircraft's type (e.g., Boeing 747, Airbus A340, etc.). Local airport weather conditions were available to participants by requesting (from a menu accessed in the TSD) the ATIS broadcast for the corresponding landing site. Approach plate

Fig. 1 Example experimental ground station

information allowed participants to look up a schematic diagram for each available approach at a given airport in addition to legal requirements (e.g., weather ceiling minimums) necessary for the landing. Finally, the ACFP pop-up window interface provided participants with ACFP's recommendation for a landing site together with varying degrees and kinds of transparency information depending on the scenario's transparency condition (detailed below). In the scenario, participants were instructed to land all aircraft under their control, this resulted in the need to land 6 aircraft in each trial.

Following the examples set forth in [9] information presented to the participant in the ACFP window varied across scenarios using three hierarchical levels of transparency, identified here as baseline, value, and logic. In the baseline transparency condition (Fig. 2), participants were provided a recommendation from ACFP displaying the recommended landing site (airport and runway number), runway length (in feet), approach name/type, and distance to the landing site (in nautical miles). The value transparency condition (Fig. 2) included, in addition to the information presented in the baseline transparency condition, a "risk statement" that provided ACFP's evaluation 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"). It is important to note that a success probability of 55 % means that there is a 45 % chance of having to perform a "go-around" or follow-up attempts of the approach, not a 45 % chance of crashing. Finally, the logic transparency condition (Fig. 2) included all information presented in the low and medium conditions as well as statements to explain the ACFP's rationale behind its recommendation. These statements gave descriptions of relevant factors along the enroute, approach, and landing phases of flight that led to its determination for the recommendation See Figs. 3 and 4.

Fig. 2 Screen capture of the baseline transparency condition

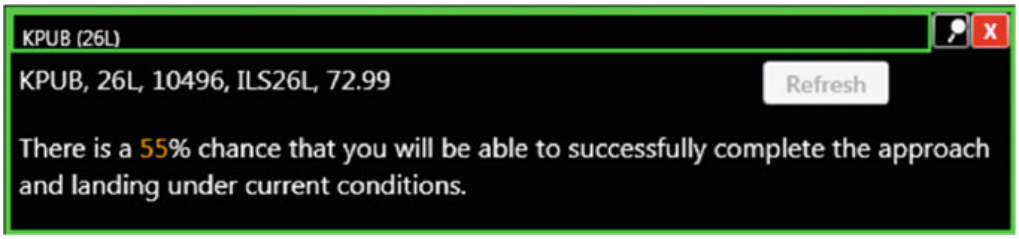

Fig. 3 Screen capture of the value transparency condition

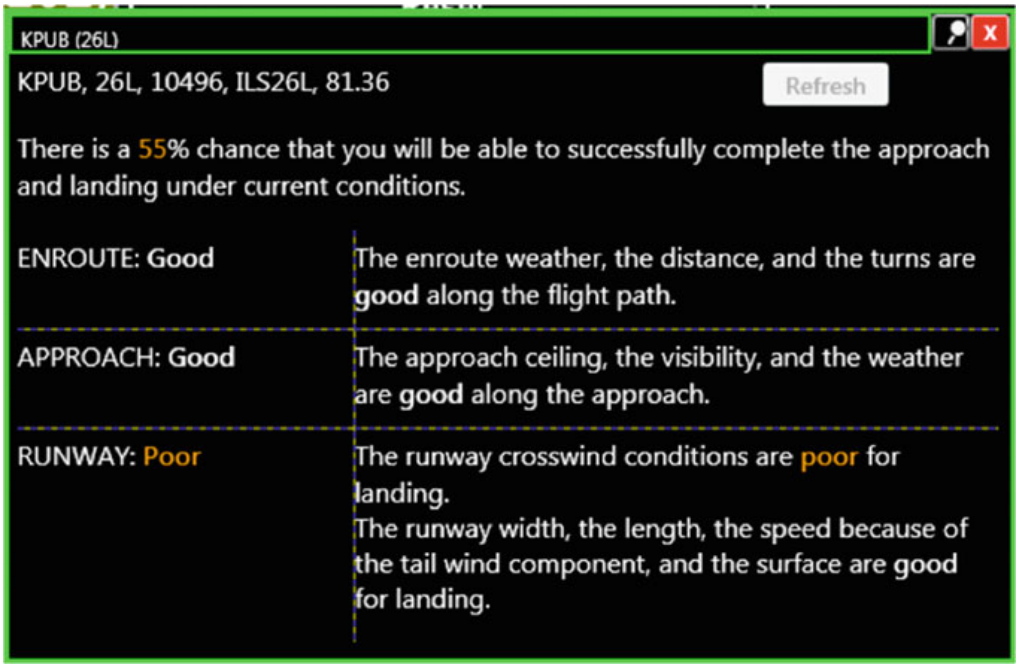

Fig. 4 Screen capture of the logic transparency condition

2.4 Measures

Trust was measured using a 7-item scale to gauge pilot's intentions to be vulnerable to the ACFP [9]. Participants 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 .88–.92. Example items included: "I think using the [ACFP] will lead to positive outcomes," "I would feel comfortable relying the recommendations of the [ACFP] in the future," and "when the task was hard I felt like I could depend on the [ACFP]."

3 Results

The order of transparency conditions was counterbalanced within a repeated measures design to maximize statistical power. To explore potential order effects of the transparency conditions over time, a repeated measures analysis was conducted. While there was no main effect of order, nor a main effect of time on trust (all p' s >0.05), there was a significant time by order interaction, $F(5, 7) = 12.44$, $p < 0.05$. As depicted in Fig. 5, the interaction effect follows a quadratic trend such that participants tend to report higher trust when they interact with the logic form of transparency either early (e.g., 5) or later in the task (e.g., 3 and 1). It is also clear that participant interactions with the baseline transparency resulted in lower trust

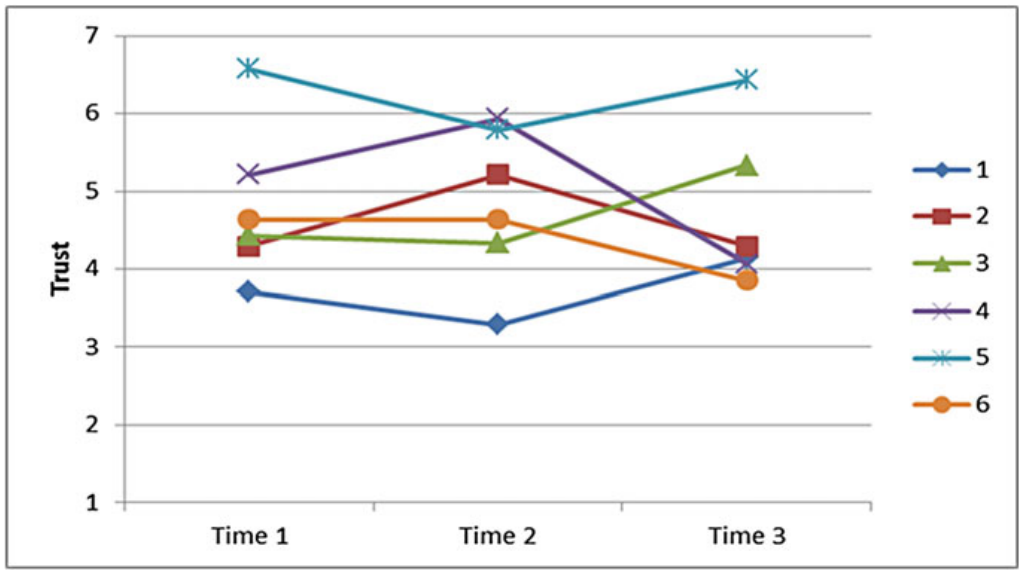

Fig. 5 Time by order interaction predicting trust

when it followed either the logic or value transparency (e.g., 4 and 6). Given that the order of the transparency conditions did have an influence on trust overtime, we used a repeated measures ANCOVA to examine the impact of transparency condition on trust while including order as a covariate. As shown in Fig. 6, trust was highest in the logic condition and lowest in the baseline condition. These

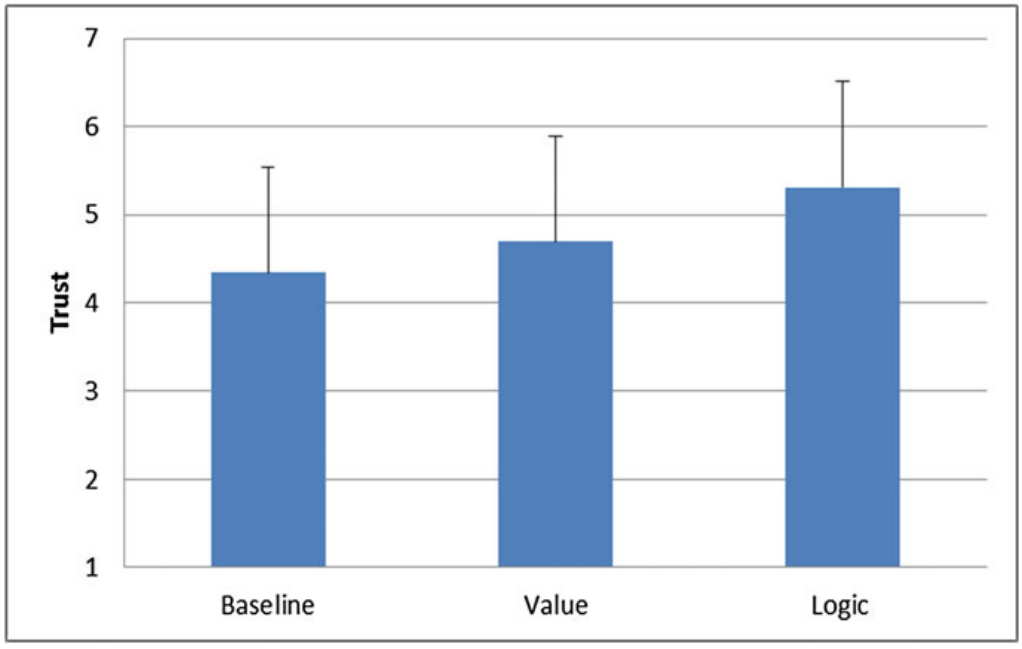

Fig. 6 Means for trust by condition

differences were reliable, $F(2, 22) = 4.39$, $p < 0.05$, demonstrating that trust was influenced by transparency condition and that the highest level of trust was associated with the logic-based form of transparency.

4 Discussion

Trust of automated systems remains a highly pertinent topic for researchers given the burgeoning nature of advanced technology. Technology offers the promise of improved performance and reduced workload for human users/managers of technology, yet these benefits will only be realized when the technology is designed in such a way as to foster appropriate reliance. One such method involves adding transparency features to automated systems. The present research explored the impact of transparency on trust using a high-fidelity simulation involving an automated aid in commercial aviation.

Consistent with prior research, the current study demonstrated that higher levels of transparency engender higher trust of automation. Specifically, the use of logic-based explanations for the recommendations was found to promote trust. This is consistent with a prior study that used similar transparency manipulations [9], however that previous study was conducted using low-fidelity methods. The current results naturally extend prior research by demonstrating the benefits of logic-based transparency in a high-fidelity task simulation using commercial pilots as the human operators. Clearly, when automated aids offer recommendations to humans they should include information related to the rationale or the key drivers of the recommendation, as this will help to foster trust in the automation. The rationale provided by the automation will help to reduce uncertainty on behalf of the human.

Future research should continue to explore the impact of transparency on the trust process. Future studies might consider a variety of different forms of transparency. The SA-based model of transparency highlights the importance of perception, comprehension, and projection and their additive effects [7]. Perhaps most importantly, Mercado and colleagues [7] found that higher levels of transparency modulated trust with no detriment to cognitive workload. This is critical as added information has the potential for overloading operators, which is counterproductive. Further, future research should explore an expanded view of transparency as outlined in [6].

References

1. Onnasch, L., Wickens, C.D., Li, H., Manzey, D.: Human performance consequences of stages and levels of automation: an integrated meta-analysis. Hum Factors **56**, 476–488 (2014)
2. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum Factors **46**, 50–80 (2004)

3. Lyons, J.B., Stokes, C.K.: Human-human reliance in the context of automation. Hum Factors **54**, 111–120 (2012)
4. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multirobot control: a review of the human factors issues. IEEE Transactions on Human-Machine Systems, 13–29 (2014)
5. Hoff, K.A., Bashir, M.: Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors **57**, 407–434 (2015)
6. Lyons, J.B.: Being transparent about transparency: a model for human-robot interaction. In: Sofge, D., Kruijff, G.J., Lawless, W.F. (eds.) Trust and Autonomous Systems: papers from the AAAI spring symposium (Technical Report SS-13-07). AAAI Press, Menlo Park, CA (2013)
7. Mercado, J.E., Rupp, M.A., Chen. J.Y.C., Barnes, M.J., Barber, D., Procci, K.: Intelligent agent transparency in human-agent teaming for multi-UxV management. Human Factors (in press)
8. Wang, L., Jamieson, G.A., Hollands, J.G.: Trust and reliance on an automated combat identification system. Hum Factors **51**, 281–291 (2009)
9. Lyons, J.B., Koltai, K.S., Ho, N.T., Johnson, W.B., Smith, D.E., Shively, J.R.: Engineering trust in complex automated systems. Ergon in Design **24**, 13–17 (2016)
10. Meuleau, N., Plaunt, C., Smith, D., Smith, C.: Emergency landing planner for damaged aircraft. In: Proceedings of the Scheduling and Planning Applications Workshop (2008)
11. Brandt, S.L., Lachter, J., Battiste, V., Johnson, W.: Pilot situation awareness and its implications for single pilot operations: analysis of a human-in-the-loop study. Procedia Manufacturing **3**, 3017–3024 (2015)