In Automation We Trust? Identifying Varying Levels of Trust in Human and Automated Information Sources

Carl J. Pearson, Allaire K. Welk, & Christopher B. Mayhorn
North Carolina State University, Raleigh, NC

Humans can easily find themselves in high cost situations where they must choose between suggestions made by an automated decision aid and a conflicting human decision aid. Previous research indicates that trust is an antecedent to reliance, and often influences how individuals prioritize and integrate information presented from a human and/or automated information source. Expanding on previous work conducted by Lyons and Stokes (2012), the current experiment measured how trust in automated or human decision aids differs along with perceived risk and workload. The simulated task required 126 participants to choose the safest route for a military convoy; they were presented with conflicting information regarding which route was safest from an automated tool and a human. Results demonstrated that as workload increased, trust in automation decreased. As the perceived risk increased, trust in the human decision aid increased. Individual differences in dispositional trust correlated with an increased trust in both decision aids. These findings can be used to inform training programs and systems for operators who may receive information from human and automated sources. Examples of this context include: air traffic control, aviation, and signals intelligence.

INTRODUCTION

Many complex tasks involve operators receiving decision-making input from automated and other human sources at the same time. One unfortunate example of this is where a Russian passenger jet and cargo plane in 2002 crashed in a mid-air collision. Like most large, commercial planes, these aircraft had automated information sources, which told two planes headed for a direct collision to change elevation in different directions. While one pilot flew his plane downward, as indicated by his automated system, the pilot in the opposing plane ignored the automated aid, which directed him to fly upward. Instead, this pilot listened to an air traffic controller who told him to descend as well, apparently unaware of the opposing plane's trajectory. The planes collided, in part due to a decision to trust and rely on the judgment and knowledge of a human information source over an automated information source. Clearly, the way humans decide between information presented by fellow humans and automation must be investigated, especially in risky, high-cost situations. Trust is one avenue that can be explored to determine how individuals determine which information source on which to rely (Lee & See, 2004).

Human-Human Trust

The process through which humans establish and develop trust in other humans has received a good bit of attention from the organizational management research community. Mayer, Davis, and Schoorman (1995) largely contributed to this body of work through the creation of their integrated model of organizational trust, which identified constructs of ability, benevolence, and integrity as factors that influence interpersonal perceptions of trustworthiness. They defined trust as "the willingness of a party to be vulnerable to another party" (Mayer et al., 1995). In this sense, trust is the precursor to reliance, when one actually commits to an action that makes them vulnerable, as shown by Serva, Fuller, and Mayer (2005). This demonstrates how interpersonal trust can be used as an indicator of reliance in human-to-human decision making.

Human-Automation Trust

Similarly, there has been a growing body of research in how humans trust automation. In particular, researchers have explored how humans calibrate their trust in automation and adjust after automation failure, as described in *Designing for Appropriate Reliance* by Lee and See (2004). Specific traits or situations have been investigated with regard to automation reliance, including situations involving heightened workload, where Biros, Daly, & Gunsch (2004) found that higher workload was related to an increased reliance on automation. These specific traits or situations have not been extensively explored in the context of simultaneous presence of human decision aids and automated decision aids.

Human Versus Automation Trust

While human-human trust and human-automation trust have both been investigated in previous research, very little work has examined how humans trust other humans or automation when both are present and in conflict. In one related study, Lyons and Stokes (2012) identified that humans relied on automation more heavily in higher risk scenarios when risk was manipulated; however, trust in human/automated sources was not measured. Furthermore, there were some methodological concerns with this study; in this within-subjects design, the consistency of the human information source's recommendations was questionable, which could have affected trust and subsequent reliance across trials. In addition to that, there was limited statistical power with only 40 participants.

Current Study Goals

This study was designed to extend previous work in the area by assessing the degree to which trust in an information source (human or automated) is influenced by the situational factors: workload and perceived risk, in addition to the individual factor: dispositional trust. On the basis of related research, several hypotheses were generated. The first hypothesis in this study was that the participants' trust in automation would be negatively related to a higher perceived workload (H1). The second hypothesis was that higher perceived risk of participants would be positively related to higher human trust (H2). The third and fourth hypotheses were that individuals with high dispositional trust would trust an automation information source (H3) and a human information source (H4) more than those who were generally less trusting.

METHOD

Participants

A sample of 126 undergraduate participants was obtained from an introductory Political Science course at North Carolina State University. Students received course credit in exchange for participation. The average age was 19 years old (*SD* = 2.9, *M* = 19.2). The balance of males and females was fairly even with 66 males and 60 females. The experiment was deployed using *Qualtrics* online survey software.

Measures & Materials

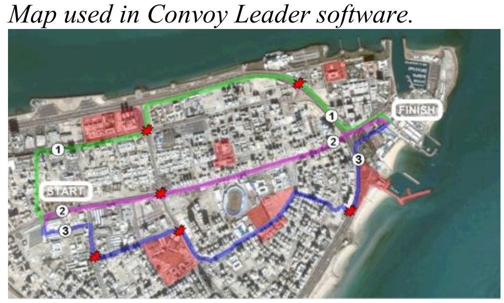
There were multiple self-report scales and measurements used in this experiment. Participants were presented with a self-report scale of dispositional trust (Merritt & Ilgen, 2008). There were 10 questions (e.g.: "I believe in human goodness"), and composite scores were created by averaging response values from a 7-point scale. The NASA TLX (Hart & Staveland, 1988) was used as a measure of perceived workload, and was assessed following the experimental task. The NASA TLX is a multidimensional measure which was created to assess workload retroactively, based on self-report. A single item on a 5-point Likert scale was used to assess perceived risk.

The interpersonal trust scale was adopted from Mayer et al. (1995). There were 21 questions on a 5-point scale of agreement to disagreement (e.g.: "I feel very confident about the human's skills"), where 1 = strongly disagree and 5 = strongly agree. Composite scores were created by averaging responses. One question was removed as it referred to trust calibration over time, which did not apply to our task because participants only interacted with each source once.

The trust in automation scale was borrowed from Bisantz and Seong (2001). There were 11 questions on a 5-point scale of agreement to disagreement (e.g.: "I am wary of the system"), also where 1 = strongly disagree and 5 = strongly agree. Composite scores were created by averaging responses. Participants were informed that "the system" referred to the map in the Convoy Leader software task.

The Convoy Leader platform tasked participants with deciding the best route for a ground-based military convoy to traverse a war-torn, hostile city. Data were provided for specific routes by automated and human sources. An automated tool (a map in this situation) provided information regarding past improvised explosive device (IED) locations and areas of insurgent activity, as illustrated in Figure 1. There were three possible routes on the map. The map recommended a path implicitly based on the choice with the fewest IED sites and least adjacent enemy territory to the route.

Figure 1.
Map used in Convoy Leader software.



A human intelligence officer recommended a different path based on information they obtained separately from the automated tool's information. The human information source appeared as a video clip where he verbally described the recommendation over a 30 second period, as can be seen in Figure 2. A third route option was present but was not recommended by either information source.

Figure 2.
Screenshot of human decision aid video.

Procedure

First, participants provided informed consent before they filled out the dispositional trust questionnaire (Merritt & Ilgen, 2008). After that, the experimental task was explained to the participant. The task was to decide the safest route for the convoy. Participants were randomly assigned to one of three between subject conditions in which the stimuli presentation varied. Information was presented concurrently, with both information sources being presented simultaneously, the automated tool was presented first followed by the human source, and the human source presented first followed by the automated tool. After viewing both information sources, participants chose the route they believed to be safest. Lastly, they filled out the interpersonal (Mayer et al., 1995) and automation trust (Bisantz & Seong 2001) questionnaires, the NASA TLX workload measure, and the perceived risk question.

The data were collected and analyzed in *SPSS* to statistically test the hypotheses. The variables analyzed were dispositional trust, trust in the automated tool after the task, trust in the human after the task, perceived risk of the situation, and perceived workload. Cronbach's alpha was calculated for all of the variables, with the exception of perceived risk, as it had a single item. Most of the variables had a high level of reliability: in the trust predisposition $\alpha = .87$, in automation trust $\alpha = .87$, in human trust $\alpha = .90$. Only in the NASA TLX was the reliability low ($\alpha = .49$). This is most likely due to the nature of the questions and the experimental task. For example, one question asks about the physical difficulty of the task. This may be interpreted by participants differently, and possibly on the low end, given that the test was deployed over a computer with a simple mouse and keyboard as input methods. These answers most likely differed from questions in the TLX asking about mental workload, which would be higher.

RESULTS

A series of multiple regressions were conducted to determine which of the collected variables predicted trust in automation and trust in the human decision aid. Included in the models as predictors were the NASA TLX, perceived risk, and dispositional trust. A full list of descriptive statistics can be found in table 1.

Table 1.

Descriptive measures of variables in Regression models

Measure	Mean	SD
1. Dispositional Trust	4.52	1.79
2. NASA TLX	67.44	9.38
3. Perceived Risk	3.08	1.07
4. Bisantz Trust (Automation)	47.81	9.30
5. Mayer Trust (Interpersonal)	31.13	5.48

Note: variables 1-3 are predictor variables, variables 4 and 5 are separate outcome variables

Human Trust and Perceived Risk

A multiple regression analysis was conducted to determine if dispositional trust and perceived risk, among other variables, were predictors of participants' trust in the human source. A summary of significant predictors can be seen in table 2. The overall model was found to be significant, $F(3,122) = 20.08$, $p < .001$. The predictors explained 33% of the variance (R^2=.33). It was found that higher risk predicted higher trust in the human (β=.23, p=.004) and higher dispositional trust predicted higher trust in the human (β=.50, p<.001).

The positive relationship between perceived risk and trust in the human rejects the second null hypothesis in support of the second alternative hypothesis (H2): those who experienced higher perceived risk tended to have more trust in the human. The positive relationship between dispositional trust and interpersonal trust rejects the fourth null hypothesis in support of the fourth alternative hypothesis (H4): higher dispositional trust predicted higher trust in the human source.

Table 2.

Significant variables in multiple regression model with perceived risk and dispositional trust predicting interpersonal trust

Variable	B	SE B	β
Dispositional Trust	.51	.08	.50**
Perceived Risk	1.20	.41	.23*

Note: * p < .01, ** p <.001

Interpersonal Trust Antecedents

To further explore the relationship among trust in human information sources, dispositional trust, and perceived risk, four antecedents of trust behavior as identified by Mayer, Davis, and Schoorman (1995) were examined. These antecedents exist as sub-measures within the full instrument; they are trust, ability, benevolence, and integrity. Multiple regression analyses were conducted with the previously discussed predictor variables on each of the four antecedent subscales, with each trust antecedent as the dependent variable. All models were found to be significant: trust [$F(3,122)$=10.91, p<.001, R^2=.21], ability [$F(3,122)$=9.49, p<.001, R^2=.19], benevolence [$F(3,122)$=16.35, p<.001, R^2=.29], and integrity [$F(3,122)$=11.90, p<.001, R^2=.23]. Perceived risk and dispositional trust were found to be significant predictors in all multiple regression models of the trust antecedents, as can be seen in tables 4-7 in the appendix.

Automation Trust and Workload

A multiple regression analysis was conducted to discover how the previously discussed predictor variables influenced participants' trust in automation (see table 3). The model was found to be significant, $F(3,122) = 8.19$, p <.001. The predictors explained 23% of the variance (R^2=.23). It was found that higher workload significantly predicted less trust in automation (β=-.25, p=.005) and higher dispositional trust significantly predicted higher trust in automation (β=.33, p<.001).

The negative relationship between workload and trust in automation rejects the first null hypothesis in support of the first alternative hypothesis (H1): those who experienced higher workload tended to have less trust in automation. The positive relation between dispositional trust and trust in automation rejects the third null hypothesis in support of the third alternative hypothesis (H3): higher participant dispositional trust was positively related to higher trust in automation.

Table 3.
Significant variables in multiple regression model with workload and dispositional trust predicting automation trust

Variable	B	SE B	β
Dispositional Trust	.20	.05	.33**
Workload	-1.26	.49	-.25*

Note: * $p < .05$, ** $p < .01$

DISCUSSION

This study further explored the extent to which individuals trust automated and human decision aids when both are presented and conflict. Compared to the seminal work of Lyons and Stokes (2012), the results from the current experiment differed. Where Lyons and Stokes (2012) found that humans displayed increased reliance in automation in riskier situations, the current results indicate that trust in automation was unaffected by perceived risk. Furthermore, results from the current study demonstrated that trust in automation was lower when participants reported higher perceived workload. Lastly, and in direct contrast to previous research, the current study found that humans reported increased trust in the human information source in riskier situations, as opposed to increased trust in automation.

One possible explanation for these conflicting findings might involve cognitive overhead, where the potential benefits of automation are outweighed by the additional load of engaging with automation (Kirlik, 1993). In the context of this experiment, the added work of assessing the trustworthiness of the automated tool could have been perceived as too great when compared to the potential benefits of using the automation. This plausibly explains why an operator may not trust automation in a higher workload situation, when there are fewer resources to assess trustworthiness of the automation.

Future Research

Based on the findings of the current study and the work published by Lyons and Stokes (2012), it seems that it may be beneficial to measure both behavioral reliance and trust in the same study. This experimental implementation could contribute to a theoretical distinction between the two constructs. Much of the research existing on trust in automation involves designs where automation failure occurs and how trust in that automated source is (or is not) regained. This topic would be worth investigating among situations with dual information sources (human and automation), as they are both liable to make errors in distinct ways. There may also be differences across populations (such as civilian versus military) sampled such that information might vary for particular contexts of use.

Conclusions

In this study, human trust in conflicting human and automated information sources was investigated. It was found that humans tend to trust other humans more in situations of higher perceived risk and that humans tend to trust automation less where there is a higher perceived workload. In many high cost situations, such as aviation or signals intelligence, this situation occurs frequently and must be better understood so that optimal choices by the operator can be guided. This research supports that humans, in a situation of conflicting information between automation and another human, will tend to have a greater amount of trust in the human information source when there is a high perceived risk and a lesser amount of trust in the automation information source when there is a higher perceived workload.

REFERENCES

Biros, D., Daly, M., & Gunsch, G. (2004). The Influence of Task Load and Automation Trust on Deception Detection. *Group Decision and Negotiation*, 173-189.

Bisantz, A., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, 85-97.

Kirlik, A. (1993). Modeling Strategic Behavior in Human-Automation Interaction: Why an “Aid” Can (and Should) Go Unused. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 221-242.

Lee, J., & See, K. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50-80.

Lyons, J., & Stokes, C. (2011). Human-Human Reliance in the Context of Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 112-121.

Madhavan, P. & Wiegmann, D. A. (2007). Similarities and Differences between human-human and human-automation trust: and integrative review. *Theoretical Issues in Ergonomics Science,* 277-301.

Mayer, R., Davis, J., & Schoorman, F. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, 709-734.

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

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press.

Serva, M., Fuller, M., & Mayer, R. (2005). The reciprocal nature of trust: a longitudinal study of interacting teams. *Journal of Organizational Behavior*, 625-648.

Appendix

Table 4.
Significant predictors in multiple regression model predicting Trust subscale in Mayer (1995) trust model

Variable	B	SE B	β
Dispositional Trust	.09	.02	.36**
Perceived Risk	.34	.11	.25*

Note: * p <.01, ** p < .001

Table 5.
Significant predictors in multiple regression model predicting Ability subscale in Mayer (1995) trust model

Variable	B	SE B	β
Dispositional Trust	.12	.03	.37**
Perceived Risk	.31	.15	.18*

Note: * p < .05, ** p <.01

Table 6.
Significant predictors in multiple regression model predicting Benevolence subscale in Mayer (1995) trust model

Variable	B	SE B	β
Dispositional Trust	.17	.03	.49**
Perceived Risk	.30	.15	.16*

Note: * p < .05, ** p <.001

Table 7.
Significant predictors in multiple regression model predicting Integrity subscale in Mayer (1995) trust model

Variable	B	SE B	β
Dispositional Trust	.12	.02	.43**
Perceived Risk	.24	.12	.17*

Note: * p < .05, ** p <.001