Differences in Trust between Human and Automated Decision Aids

Carl J. Pearson
North Carolina State University
2310 Stinson Dr
Poe Hall 700
Raleigh, NC 27695
cjpearso@ncsu.edu

Allaire K. Welk
North Carolina State University
2310 Stinson Dr
Poe Hall 700
Raleigh, NC 27695
akwelk@ncsu.edu

Dr. William A. Boettcher
North Carolina State University
Campus Box 8102
Raleigh, NC 27695
william_boettcher@ncsu.edu

Dr. Roger C. Mayer
North Carolina State University
2801 Founders Dr
Nelson Hall 1328
Raleigh, NC 27607
rcmayer@ncsu.edu

Sean Streck
North Carolina State University
1021 Main Campus Drive, Suite 310
Raleigh, NC 27606
smstrec2@ncsu.edu

Dr. Joseph M. Simons-Rudolph
North Carolina State University
2310 Stinson Dr
Poe Hall 714C
Raleigh, NC 27695
jmrudolp@ncsu.edu

Dr. Christopher B. Mayhorn
North Carolina State University
2310 Stinson Dr
Poe Hall 728
Raleigh, NC 27695
cbmayhor@ncsu.edu

ABSTRACT

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 humans often rely on automation or other humans, but not both simultaneously. Expanding on previous work conducted by Lyons and Stokes (2012), the current experiment measures 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 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 for operators who may receive information from human and automated sources. Examples of this context include: air traffic control, aviation, and signals intelligence.

CCS Concepts

• Human-Centered Computing → Human-Computer Interaction → Laboratory Experiments

HotSoS '16, April 19-21, 2016, Pittsburgh, PA, USA
© 2016 ACM. ISBN 978-1-4503-4277-3/16/04…$15.00
DOI: http://dx.doi.org/10.1145/2898375.2898385

Keywords

Trust; reliance; automation; decision-making; risk; workload; strain

1. 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 their plane down as told by automation, the pilot in the opposing plane ignored the automation aid, which directed them to fly up. Instead, they listened to an air traffic controller who told them to descend as well, apparently unaware of the opposing plane's trajectory. The planes then collided, in part due to a decision to trust the judgment and knowledge of a human information source over an automated information source. Clearly, the way humans decide between fellow humans and automation must be investigated, especially in risky, high-cost situations.

1.1 Human-Human Trust

The organizational management literature contains a good deal of research that investigated how humans trust other humans. 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 of trustworthiness. They defined trust as: the willingness of a party to be vulnerable to another party. 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 shows how trust can be used as an indicator of reliance in human-to-human decision making.

1.2 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.

1.3 Human Versus Automation Trust

While human-human reliance and human-automation reliance both have been investigated in previous research, very little work has been done on how humans rely on humans or automation when both are present and in conflict. Lyons and Stokes (2012) began this exploratory research and found that humans relied on automation more heavily in higher risk scenarios when risk was manipulated. 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 reliance across trials. In addition to that, there was limited statistical power with only 40 participants.

1.4 Current Study Goals

The current study was conducted with stimuli adopted from the initial Lyons and Stokes (2012). Given the work by Biros et al. (2004) that described how workload affects reliance, the current study explored this topic within the context where both human and automation was present. 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. The fifth hypothesis was that the human information source would be relied on more overall than the automation (H5).

2. METHOD

2.1 Participants

The sample pool of 126 participants was obtained from an undergraduate group at North Carolina State University, in exchange for class extra credit. They were mostly directed to the study from enrollment in Political Science coursework. 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.

2.2 Measures & Materials

First, participants were presented with a self-report scale of dispositional trust adapted from a portion of the International Personality Item Pool (Goldberg, 2015). 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 was used as a measure of perceived workload following the experimental task. A single item on a 5-point Likert scale was used to assess perceived risk.

The trust in the human 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. Two questions were removed as they referred to trust in relationship to an organization, which did not apply to our task in any way. 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 prompted that "the system" referred to the map in the Convoy Leader software task. The measure of reliance was taken as the behavioral selection of the route as recommended by the automated tool or human information source.

Convoy Leader allows a participant to decide the best route for a ground-based convoy to traverse a war-torn, hostile city based on data provided for specific routes by automated or human sources. It involves an automated tool (a map in this situation) with past improvised explosive device (IED) locations and known enemy territory, as can be seen in figure 1. There are three routes possible on the map. The map recommends 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 recommends 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 exists but is not recommended by either information source.

Figure 2: Screenshot of human decision aid video

2.3 Procedure

First, participants completed an informed consent before they filled out the dispositional trust questionnaire. After that, the experimental task was explained to the participant. The task was to decide the safest route to take with the convoy, given conflicting information from the automated tool and the human. Participants were told they would have to choose the safest route. Then they were randomly divided into three conditions in which the stimuli presentation varied: concurrent where both information sources were presented at the same time, 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 as they saw fit. After that, they filled out the trust questionnaires, 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. 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.

3. RESULTS

A series of descriptive statistics were calculated to assess the normality and skew of the variables involved, as can be seen in Table 1. All values of skew and kurtosis were small, with the skew ranging from -.41 to .2, and kurtosis ranging from -.38 to .79.

Table 1.

Descriptive statistics of all variables used in multiple regressions

Measure	N	Mean	SD	Skew	Kurtosis
Perceived Risk*	126	47.8	1.79	-0.41	-0.38
Personality Trust	126	4.5	9.30	-0.59	0.79
Automation Trust	126	31.1	5.48	-0.39	0.04
Human Trust	126	67.4	9.38	-0.40	0.18
Workload	126	3.1	1.07	0.20	-0.25

Note: higher numbers in mean indicates higher value of measured trait or perception

3.1 Human Trust and Perceived Risk

Perceived risk and dispositional trust were assessed as predictors of trust in the human source. Bivariate correlations were first analyzed to assess collinearity in the predictors. There was no significant correlation between perceived risk and general trust where $r(125) = .12$, $p = .167$. This suggests there is negligible collinearity among predictor variables.

A multiple regression analysis was conducted to discover if general trust and perceived risk were predictors of participants' trust in the human (see table 2). The model was found to be significant, $F(2,123) = 28.55$, $p < .001$. The two predictors explained 32% of the variance ($R^2=.32$). It was found that higher risk predicted higher trust in the human ($\beta=.20$, $p=.009$) and higher general trust predicted higher trust in the human ($\beta=.50$, $p<.001$).

The positive relationship between perceived 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 general trust and trust in automation rejects the fourth null hypothesis in support of the fourth alternative hypothesis (H4): higher participant general trust was related to higher trust in the human.

Table 2.

Multiple regression model with perceived risk and general trust predicting human trust

Variable	B	SE B	β
General Trust	.51	.08	.50**
Perceived Risk	1.04	.39	.20*

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

3.2 Automation Trust and Workload

The effect of workload and general trust were assessed as predictors of automation trust. Bivariate correlations were first analyzed to assess collinearity in the predictors. There was no significant correlation between workload and general trust where $r(125) = .02$, $p = .809$. This suggests there is negligible collinearity among predictor variables.

A multiple regression analysis was conducted to discover if general trust and workload were predictors of participants' trust in automation (see table 3). The model was found to be significant, $F(2,123) = 11.93$, $p < .001$. The two predictors explained 16% of the variance ($R^2=.16$). It was found that higher workload predicted less trust in automation ($\beta=-.23$, $p=.007$) and higher dispositional trust predicted higher trust in automation ($\beta=.34$, $p<.001$).

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 general trust and trust in automation rejects the third null hypothesis in support of the third alternative hypothesis (H3): higher participant general was related to higher trust in automation.

Table 3.

Multiple regression model with workload and general trust predicting automation trust

Variable	B	SE B	β
General Trust	.20	.05	.34**
Workload	-1.16	.44	-.23*

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

3.3 Reliance on Information sources

There was a marginal difference between reliance on information sources, with the human information source being most frequently relied on across participants. However, these results were not statistically significant; there was no significant difference in actual reliance between the human information source and automated tool information source (H5).

4. DISCUSSION

This study further explored the nature of humans' decisions to trust conflicting sources between automated decision aids and human decision aids. Building on the work of Lyons and Stokes (2012), the results from the current experiment differed from the results of the previous paper. Where Lyons and Stokes (2012) found that humans displayed increased trust in automation in riskier situations, the current results indicate that trust in automation had no difference based on risk. Separate from risk as investigated by Lyons and Stokes (2012), it was found in this current study that trust in automation was lower with a higher workload perceived by participants. Directly in contrast to previous results, the current study found that humans had more trust in the human information source in riskier situations, as opposed to more trust in automation. A key difference between the methodologies used in these papers is that Lyons and Stokes measured reliance directly, while this study measured trust and reliance. These are feasibly different constructs, which could promote different patterns in the ostensibly conflicting results. This was the case in the current study where trust differed significantly across the predictors of workload or perceived risk, but reliance did not have significant differences across predictors.

One possible explanation for these conflicting findings might involve cognitive overhead, where the potential benefits of automation are possibly 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.

While the marginal reliance results were aligned with the trust indicators, they were non-significant. Reliance and trust were known to be distinct constructs, but may have more variation within a single situation than anticipated. The lack of significance in reliance could also be attributed to that fact that the automated tool used in this study had a clear answer, but did not explicitly take over the task for the user. For this reason it was not true automation, which could have also affected the ultimate reliance in the automation measured in this experiment.

4.1 Future Research

Based on the findings of the current study and that published by Lyons and Stokes (2012), it might be beneficial to measure both reliance and trust more distinctly in the same study. A distinction could be made theoretically between the two if there is an interaction among results. Much of the research existing on trust in automation involves designs where automation failure occurs and how that trust is (or is not) regained. This 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 as information might vary for particular contexts of use.

4.2 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.

5. REFERENCES

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

[2] 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.

[3] Goldberg, L. (2015). A Broad-Bandwidth, Public-Domain, Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Models. Retrieved November 29, 2015, from http://ipip.ori.org/newBroadbandText.htm

[4] 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.

[5] 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.

[6] 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.

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

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