Published in final edited form as:

J Cogn Eng Decis Mak. 2013 March 1; 7(1): 49–65. doi:10.1177/1555343412453461.

The Effect of Information Analysis Automation Display Content on Human Judgment Performance in Noisy Environments

Ellen J. Bass, Leigh A. Baumgart, and Kathryn Klein Shepley

University of Virginia, Charlottesville, Virginia, USA

Abstract

Displaying both the strategy that information analysis automation employs to make its judgments and variability in the task environment may improve human judgment performance, especially in cases where this variability impacts the judgment performance of the information analysis automation. This work investigated the contribution of providing either information analysis automation strategy information, task environment information, or both, on human judgment performance in a domain where noisy sensor data are used by both the human and the information analysis automation to make judgments. In a simplified air traffic conflict prediction experiment, 32 participants made probability of horizontal conflict judgments under different display content conditions. After being exposed to the information analysis automation, judgment achievement significantly improved for all participants as compared to judgments without any of the automation's information. Participants provided with additional display content pertaining to cue variability in the task environment had significantly higher aided judgment achievement compared to those provided with only the automation's judgment of a probability of conflict. When designing information analysis automation for environments where the automation's judgment achievement is impacted by noisy environmental data, it may be beneficial to show additional task environment information to the human judge in order to improve judgment performance.

Keywords

human-automation interaction; human-automated judge learning; automation display content; judgment analysis

Introduction

Information Analysis Automation

Information analysis automation augments human cognitive functions, such as working memory and inferential processing, in making a judgment regarding the state of the environment (Parasuraman, Sheridan, & Wickens, 2000). One class of information analysis automation processes input data and uses them in assessments of future states of the environment. For example, air traffic conflict prediction software can use current noisy

position and velocity inputs to derive a range of future aircraft positions and then predict the probability of a conflict with another aircraft. A human operator's ability to utilize the output of the automation depends on the represented information content such as the automation's judgment, information about the automation's judgment strategy, and information about the noisy environmental input data used by the automation.

Several studies have investigated the effects of automation type and level on human performance (see, for example, Endsley & Kaber, 1999; Endsley & Kiris, 1995; Kaber & Endsley, 1997, 2004; Lorenz, Di Nocera, Röttger, & Parasuraman, 2002; Wickens, Li, Santamaria, Sebok, & Sarter, 2010; Wickens & Xu, 2002). When compared to decision automation (that includes algorithms to generate and select between decision alternatives), information analysis automation (that includes algorithms to process and to integrate cues to assess the current state of the environment) has been shown to improve human performance particularly with imperfect automation (Crocoll & Coury, 1990; Rovira, McGarry, & Parasuraman, 2007; Sarter & Schroeder, 2001). However, few studies have investigated judgment performance under different conditions of display content derived from information analysis automation alone.

There may be benefits from information analysis automation that provides both an automated judgment and information related to the automation's judgment strategy. In an air traffic identification task, Seong and Bisantz (2008) found that human judgment performance improved when meta-information was provided regarding how the automation integrated input data to derive its judgment (i.e., its cue weighting strategy) compared to when only its automated judgment was provided. Pritchett and Vandor (2001) found that providing graphical information depicting intermediate steps of the automation's strategy led to decreased alert reaction time and therefore may have simplified the alert verification task. Sarter and Woods (1992, 1994) found that displays indicating the strategy of the automation improved operator responses and conformance to automated alerts compared to when strategy information was not available. Skjerve and Skraaning (2004) found that subjective measures of cooperation and time to detect critical events were improved for nuclear power plant monitoring tasks involving an interface designed to increase the observability of the automation's activity. Even automation that does not consistently apply its judgment strategy and is not consistently accurate can enhance human judgment performance over unreliable automation that does not display its strategy (Seong & Bisantz, 2002).

However, there may also be benefits from information analysis automation that provides additional information related to the noisy task environment, especially when those data are used by both the automation and the human to make a judgment. This information may provide insight into the environment as well as the unreliability of the automation's judgment as its performance is also degraded by the noisy input data. Such an ecological approach to design has been advocated (see, for example, Burns & Hajdukiewicz, 2004; Drews & Westenskow, 2006; Kirlik, 1995; Vicente, 2002; Vicente & Rasmussen, 1990; Woods, 1991). Instead of having the automation provide additional information about its strategy, it can instead help the human judge to understand the important relationships in the task environment. However, this ecological approach to automation design has mainly been

applied in the design of complex process control systems and not with information analysis automation.

Thus, while some studies suggest that providing insight into how the automation makes its judgments may improve human judgment performance, there may also be situations where information about the environment may improve human performance.

Human-Automated Judge Learning (HAJL)

In the present study, participants made judgments about the probability of an air traffic conflict in a noisy environment using the Human-Automated Judge Learning (HAJL) (Bass & Pritchett, 2008) process. HAJL builds on concepts from judgment analysis (Cooksey 1996), which is based on Brunswik's (1952, 1956) probabilistic functionalism, designating the organism-environment interaction as the primary unit of study. Cues in the environment may have associated uncertainty limiting the predictability of the criterion. Aspects of the judge's policy including the relationship between the cues and the judgment and the judge's consistency in applying the policy also impact the judgment process. Thus, investigating a judgment's achievement (i.e., correspondence between judgments and the environmental criterion) requires consideration of the relationship between the criterion and the cues and judgment and the cues.

One decomposition of judgment achievement can be established by fitting the judgments and the criterion to symmetric linear models. The environmental linear model establishes a best fit between the environmental criterion and the cues. The judgment linear model establishes a best fit between the judgment and the cues. The Lens Model Equation (LME) (Hursch, Hammond, & Hursch, 1964; Tucker, 1964) partitions judgment achievement into lower level correspondences accounting for the contributions of the environment and judge:

$$r_a = GR_E R_S + C \sqrt{1 - R_E^2} \sqrt{1 - R_S^2}$$

Linear knowledge, G , computed as the correlation between the outputs of the environment and judgment policy models, measures how well the predictions of the linear model of the judge match predictions of the linear model of the environment and thus how well a modeled judgment policy captures the linear structure of the environment. Environmental predictability, R_E , measures the degree to which the criterion is predicted by the linear model of the environment. R_E is calculated as the coefficient of multiple correlation of the environmental linear regression model (regressing the environmental criterion on the cue values). Cognitive control, R_S , measures the degree to which a judgment is predicted by the linear model of the judge and is the coefficient of multiple correlation calculated by regressing human judgments on the cue values. Un-modeled knowledge, C , measures the extent to which the judgment and the environment share nonlinear components and is computed as the correlation between the residuals of the environment and the judgment policy models.

A triple system design is a model that considers two independent judges making judgments under the same task conditions. Except for the extension to three systems, the triple system model is based upon the same psychological constructs and mathematical techniques as the double system design lens model just described (Cooksey, 1996). This model includes the interaction among three entities involved in judgment: the environmental criterion, the judgment produced by one judge, and the judgment made by the other judge. It provides measures of both objective performance and agreement between the two judges' judgments.

H AJL combines a triple system design with a three-phase (training, interactive learning [IL], and prediction) process. The training phase measures are depicted in Figure 1. The environmental criterion (the actual probability of an air traffic conflict) is depicted on the left. The cues are independently used by the automation and the human to make a judgment. The degree of correlation between human (or automation) judgment and the environmental criterion reflects the degree of judgment achievement for the human (or the automation). The correlation between the human's judgment and the automation's judgment can be used to characterize conflict between them.

In the interactive learning (IL) phase of H AJL, the human and automation make independent initial judgments and the correspondence between their initial judgments measures their conflict. After viewing the automation's judgment, the human judge then provides a revised "joint" judgment (Figure 2). Achievement of the joint judgment is the degree of correlation between the joint judgment and the environmental criterion. Correspondence between the automation's judgment and the joint judgment measures the extent to which the human adapts to the automation. Correspondence between the human's initial and joint judgments provides a measure of compromise, where lower values indicate more compromise.

In the prediction phase of H AJL, the human judge makes a judgment independent of the automation and also predicts the automation's judgment (Figure 3). These judgments with the automation's judgment allow the following three measures to be calculated: predictive accuracy of the human (i.e., correspondence between the automation and his prediction of it), assumed similarity (i.e., correspondence between the human's judgment and his prediction of the automation), and actual similarity (i.e., correspondence between the human and the automation's judgments).

This three-phase protocol facilitates the measurement of human-automation interaction metrics related to judgment performance before, during, and after the availability of the automation. These include human judgment achievement, the degree to which the human judge compromises with and adapts to the automation, the ability of the human to predict the automation's judgments, and the actual and assumed similarity between the human's and the automation's judgments.

Using the H AJL process, this research aims to investigate the contribution of providing either information about the automation's judgment strategy, information about noisy task environment data used by the automation, or both on human judgment performance.

Method

Experimental Task

Participants were asked to make judgments about the probability of an air traffic conflict, a horizontal loss of separation of 5 nautical miles (nm). They monitored the progress of their own aircraft (the “ownship”) and another aircraft (the “traffic”) using a simulated egocentric traffic display. The ownship was flown by an autopilot, so the participants did not need to control the aircraft. In the simulation, the ownship's speed, altitude, and heading remained constant, while uncertainty (sensor noise) was introduced into the speed, lateral position, and heading of the traffic aircraft.

Participants

Sixteen male and sixteen female undergraduate engineering students ranging in age from 20 to 23 volunteered for this experiment. All participants were familiar with the use of computers and had no previous experience with the judgment task. All were paid and bonuses were awarded to top performers.

Apparatus

The air traffic simulation used for the task was adapted from Bass and Pritchett (2008). It included a Traffic Conflict Prediction System (TCPS), Data Entry Display (DED), Navigation Display (ND), and the Environmental Information Display (EID).

Traffic Conflict Prediction System (TCPS)—The TCPS information analysis automation calculated probability of conflict judgments using the displayed noisy airspeed, location, and heading at the time the calculation was made (Bass & Pritchett, 2008). It projected both ownship and traffic positions to the predicted point of closest approach (PCA), calculated the predicted horizontal miss distance, and determined the probability of conflict as the cumulative distribution function around the 5 nautical mile (nm) safe separation boundary accounting for the variance in the predicted horizontal miss distance, created by extrapolating the errors in current aircraft position, velocity, and heading. The environmental criterion (the actual probability of conflict) was calculated using the TCPS with no noise in the lateral position, speed, or heading. The cumulative distribution reflected the underlying actual distribution of conflicts in the environment.

Data Entry Display—The Data Entry Display (DED) (top of Figure 4) consisted of slide bars to enter probability judgments and to display the TCPS judgment. The first slide bar was used by the participant to enter his unaided probability of conflict judgment during all three phases. In the IL phase, a second slide bar displayed the judgment of the TCPS, while a third allowed the participant to make a revised joint judgment after seeing the judgment from the TCPS (the second slide bar) and any additional display content if available. During the prediction phase, a second slide bar was available for participants to enter their prediction of the TCPS's judgment.

Navigation Display—The Navigation Display (ND) (bottom of Figure 4) was an egocentric display. It contained a green aircraft icon representing the ownship. Concentric

circles around the ownship represented distances of 5, 10, 20, 30, 40, and 50 nm. A compass was displayed with the 40 nm circle. The heading of ownship was displayed on the compass and its speed was displayed under the ownship icon.

The traffic was always at the same altitude as ownship. Traffic appeared as a yellow triangle pointing in the direction of track heading. A one-line data block displayed indicated airspeed in knots. The heading of the traffic was displayed on the compass with a yellow hash mark. Traffic data were updated once a second.

The ND could display additional automation strategy information related to how the TCPS made its probability judgment: the projected positions of the ownship and traffic. Predicted ownship position was represented with a green ownship icon surrounded by a circular green 5 nm protected zone. Projected traffic was represented with a yellow dot surrounded by a yellow 2 standard deviation position error ellipse representing how the noisy input data (lateral position, speed, and heading) affected the projected location of the traffic at the point of closest approach (PCA).

Environmental Information Display—The Environmental Information Display (EID) (Figure 5) contained representations of sensor noise (environment information) in two sections: Traffic Speed Information and Traffic Heading Information. The Traffic Speed Information display (top of Figure 5) included a speed ruler that contained a grey hash mark for every traffic speed during the trial (displayed traffic speeds). The speed of the traffic at the end of each trial was a yellow hash mark on the speed ruler. The average and standard deviation of the displayed traffic speeds were represented with three red hash marks. The middle hash mark represented the average, and the outer red hash marks represented 1 standard deviation above and below the average speed. The final speed, average, and standard deviation were also displayed numerically below the speed ruler.

The Traffic Heading Information display on the EID (bottom of Figure 5) contained a compass that had a grey hash mark for every displayed traffic heading during the trial. A yellow hash mark on the compass indicated the final heading. The average and standard deviation of the displayed traffic headings were calculated and shown using red hash marks as with the average and standard deviation of the speed (average in the middle, standard deviation marks to the left and right of the average). The heading when the trial ended, along with the average and standard deviation of the displayed traffic headings, was also displayed numerically in the center of the compass.

Trials—A total of 180 different trials were used in this experiment. They were grouped into 6 sessions of 30 trials each. Each session had a high level of environmental predictability (0.971, 0.971, 0.969, 0.985, 0.989, and 0.974). Each trial consisted of the ownship in the center of the display and one traffic aircraft that flew at one of six possible headings ($\pm 45^\circ$, $\pm 90^\circ$, and $\pm 135^\circ$ degrees from ownship's heading) and five possible speeds (at the same IAS as ownship, ± 50 knots, ± 100 knots). The lateral position, airspeed, and heading errors for the traffic aircraft were normally distributed and had standard deviations of 500 meters, 15 knots, and 3 degrees, respectively. Every second, new position, speed, and

heading errors were added to the actual position, speed, and heading of the traffic aircraft and then displayed.

Procedure

Before the collection of experimental data, demographic and baseline trust (appendix) data were collected. A participant was considered to be a high or low trusting individual if his or her average was more than 2 standard deviations above or below the mean. Two participants were identified as being “low trusting of automation” individuals and one participant was identified as being a “high trusting of automation” individual. The three individuals were assigned to different experimental conditions.

For each participant, data were collected over 5 consecutive days. In the training phase (Days 1 and 2), every participant experienced 3 sessions of 30 trials each on the first day and 3 sessions of 30 trials each on the second day. In the IL phase (Days 3 and 4), the trials were reordered within each session and the sessions were presented in a different order as compared to the training phase. Three sessions were used during the prediction phase (Day 5). Again, the trials were reordered within each session and the sessions were shown in a different order than in the training and IL phases.

On Day 1, each participant read and signed a consent form. Each completed a demographic questionnaire including trust of automation questions. Each participant then received a briefing. Each was asked to make accurate judgments throughout the experiment but told that the judgment task was difficult because the displayed information would not be perfect. Each completed half of the training phase trials (3 sessions of 30 trials each). On Day 2, participants completed the second half of the training phase (3 sessions of 30 trials each).

For each trial in the training phase, participants only had access to the Data Entry Display and the Navigation Display (Figure 4 with only the first slider bar visible and without the additional automation strategy information related to how the TCPS made its probability judgment). After a random amount of time (uniformly distributed between 15 and 30 seconds), the trial froze for entry of the unaided probability of traffic conflict judgment. After the judgment entry, the trial continued in faster than real time until the participant started the next trial. In this way the participant could view the point of closest approach, a state related to the probability of conflict environmental criterion, but the actual environmental criterion was not displayed.

Days 3 and 4 composed the interactive learning phase (IL), which included the information analysis automation, TCPS. Day 3 started with a briefing tailored to the participant's assigned display content condition group. Each participant completed 2 days (Days 3 and 4) of 3 sessions each day with 30 trials in each session. For each trial in IL phase, participants made their unaided probability judgment in the same way as in the training phase. However, once they made their unaided judgment, they were provided with content from the TCPS, and depending on the experimental condition, additional automation strategy (as depicted in Figure 4) and/or environment information (as depicted in Figure 5). They then provided a revised joint probability judgment. The trial then continued in faster than real time so the participant could view the point of closest approach. After each of the 30 trials during the IL

phase, participants filled out the trust in automation questionnaire displayed on the computer screen. Participants used slide bars to enter responses on a 7-point scale.

On the final day (Day 5), participants completed the prediction phase of 3 sessions of 30 trials (without the aid of TCPS). For each trial, participants made their unaided probability judgment in the same way as in the training phase using only the Data Entry Display and the Navigation Display (without the additional automation strategy information related to how the TCPS made its probability judgment). They also provided a judgment of what the automation would have predicted had it been available. In this phase, there was no feedback (i.e., the trial did not continue in faster than real time) so as not to confound the human's ability to predict the automation's judgment with further learning effects. During the prediction phase, participants were asked one question, which was displayed on the screen, (on a scale of 1 to 7) about their self-confidence in their ability to do the task.

Independent Variables

The display content provided to the participants during the IL phase was controlled in this study. The four display content conditions were automation judgment outcome only (O), automation judgment outcome plus environment information (OE), automation judgment outcome plus automation strategy information (OA), and automation judgment outcome plus environment information and automation strategy information (OEA). Each participant saw only one of the four display content conditions throughout the IL phase after the trial froze and the participant had made an unaided judgment.

Automation judgment only (O)—The additional display content during the IL phase was the TCPS's judgment of the probability of a conflict (slide bar labeled “Prediction System” on DED in Figure 4).

Automation judgment plus environment information (OE)—The additional display content during the IL phase included the TCPS's probability judgment (O) and task environment information related to sensor noise. This included the EID of Traffic Speed Information and the Traffic Heading Information (Figure 5), which were descriptive statistics of the noisy input data.

Automation judgment plus automation strategy information (OA)—The additional display content during the IL phase included the TCPS's probability judgment (O) and information about the TCPS's judgment strategy in the form of graphical representations of intermediate cues derived by TCPS leading to its probability of conflict judgment. This included the projected positions of ownship and traffic at the PCA, explicit representations of uncertainty with the traffic position error ellipse, and the ownship protected zone on the ND (Figure 4).

Automation judgment plus environment information and automation strategy information (OEA)—Participants were given the TCPS's outcome prediction (O) as well as the information given to the OE and the OA groups. *Session*. Session (group of 30 trials) was also treated as an independent variable in this study.

Dependent Variables

Measures based on the HAJL methodology (Table 1) related to judgment performance were calculated for each participant for each session (every 30 trials).

Data Analysis

The experimental design was a repeated measures, mixed model design. Display content condition, session (within each HAJL phase), and the session-display content interaction were fixed effects. Participants were nested within display content condition and were treated as a random effect in the model. Post hoc analysis was conducted using Tukey's Honestly Significant Difference (HSD). Wilcoxon Signed Rank tests were also used to determine if differences existed between unaided human judgment achievement and joint achievement values as well as between achievement values across phases.

Before calculating any derived measures, the inverse sine transformation was applied to all probability judgments to stabilize their variance (Box, Hunter, & Hunter, 1978). All derived measures are correlations between two sets of judgments. Therefore, before performing the data analysis described above, the correlations were transformed using Fisher's r to z transformation to obtain normally distributed variables as suggested by Cooksey (1996).

Results

The results of this experiment are presented using $\alpha = 0.05$ for significance and $\alpha = 0.1$ for a trend. The results from each of the three HAJL phases will be presented where the effects of display content and session on judgment achievement and use of automation are discussed. The display content-session interaction was never significant and is therefore not reported.

Training Phase

The average human judgment achievement, r_{a1} , during the training phase was low for all participants ($\mu = 0.38$, $\sigma = 0.24$). There was no significant difference between the participants assigned to the four display content conditions.

The effect of session was statistically significant ($F_{5, 168} = 19.482$, $p < 0.001$). A post hoc analysis using Tukey's contrasts indicates that Session 1 had significantly lower achievement than all other sessions and Session 6 had significantly higher achievement than all others. These data indicate that the participants were engaged in the experimental task and learning from practice. The lack of difference between the display content groups implies that subsequent judgment performance analysis should not be confounded by unaided judgment ability of the groups.

Interactive Learning Phase

The participants' unaided judgment achievement in the IL phase, r_{a1} ($\mu = 0.49$, $\sigma = 0.18$), was significantly higher than in the training phase as shown by a Wilcoxon Signed Rank test ($V = 12,652$, $p < 0.001$). This performance improvement may have been due to additional practice at the task as well as from learning from the automation.

Figure 6 and Figure 7 contain the unaided human judgment achievement (r_{a1}) by session and display content condition, respectively. There was no significant difference between the participants assigned to the four display content conditions in their unaided judgment achievement during the IL phase. However, session was significant ($F_{5, 168} = 4.8635, p < 0.001$). Post hoc analysis using Tukey's HSD indicates that the only significant difference between sessions for unaided human judgment was between Sessions 3 and 4 ($p < 0.001$) and between Sessions 3 and 6 ($p < 0.001$). The decrease from Sessions 3 to 4 may be explained by the fact that after Session 3, the participants had a break before starting Session 4 the next day. Participants may have needed a session to become refamiliarized with the task and the output from the automation.

The participants' joint judgment achievement after exposure to the automation was better than their unaided judgments in the IL phase. A Wilcoxon Signed Rank test indicates that average joint judgment achievement, r_{a2} ($\mu = 0.91, \sigma = 0.06$), was significantly higher than the average unaided judgment achievement ($V = 0, p < 0.001$) and much closer to the automation's judgment achievement ($\mu = 0.94, \sigma = 0.02$). Additionally, using Levene's test, the variance of joint judgments was significantly smaller compared to the variance of unaided judgments ($F_{191, 191} = 10.251, p < 0.001$) and the floor of judgment achievement was raised from -0.16 (unaided) to 0.66 (joint) across all participants.

Display content condition ($F_{3, 168} = 5.979, p = 0.001$) and session ($F_{5, 168} = 9.5782, p < 0.001$) both significantly impacted participants' joint judgments. Figure 6 and Figure 7 also contain the joint judgment achievement (r_{a2}) by session and display content condition. Tukey's HSD post hoc analysis for the session effect on joint judgment indicates that Sessions 2, 3, 5, and 6 were significantly higher than Session 1 and also that Session 4 had a trend to be higher than Session 1. This could indicate that participants needed a session (Session 1) before they fully understood how to use the output from the TCPS to improve their joint judgment.

Tukey's HSD analysis also indicates that the automation judgment only display (O) had significantly lower joint achievement than both the OE (automation judgment plus environment information) ($p = 0.003$) and OEA (automation judgment plus environment information and automation strategy information) ($p = 0.02$) conditions. There was also a trend for the OA (automation judgment plus automation strategy) condition to be lower than the OEA condition ($p = 0.08$). There was no significant difference between the OE and OEA conditions, implying that for this particular study automation strategy information did not significantly help compared to when participants were also provided with environment information.

Neither compromise ($\mu = 0.49, \sigma = 0.18$) nor adaptation ($\mu = 0.93, \sigma = 0.05$) were found to be significantly different across display content condition in the IL phase. Providing additional environment or automation strategy information did not increase use of automation as measured by compromise or adaptation. Compromise was low and adaptation was high for all participants, indicating that regardless of the display content, participants were adapting their unaided probability judgments to match the automation's judgments.

However, session was significant for both compromise ($F_{5, 168} = 2.6983, p = 0.023$) and adaptation ($F_{5, 168} = 7.7129, p < 0.001$). Figure 8 depicts untransformed compromise and adaptation by session. Post hoc analysis indicated that Session 4 was significantly lower than Session 1 for compromise ($p = 0.045$). For adaptation, post hoc analysis indicated that Session 1 was significantly lower than Session 3 ($p < 0.001$) and Session 5 ($p < 0.001$), indicating that as participants gained experience with the automation (no matter the display content), their joint judgments more closely corresponded with the automation's judgments.

Prediction Phase

The participants' unaided judgment achievement in the prediction phase, r_{a1} ($\mu = 0.58, \sigma = 0.19$), was not significantly different based on session or display content condition. However, average unaided judgment was higher than during the IL phases, which may show that participants learned from the automation.

The participants' unaided judgment achievement in the prediction phase, r_{a1} , was significantly higher than their predictive accuracy ($\mu = 0.54, \sigma = 0.18$) ($V = 0, p < 0.001$). Participants were better at making probability of conflict judgments compared to predicting the automation's judgments overall. In this particular experiment, participants were not told that they would need to predict the automation in the instructions. Predictive accuracy was not significantly affected by either display content condition or session order.

Assumed similarity ($\mu = 0.90, \sigma = 0.16$) was significantly higher than actual similarity ($\mu = 0.56, \sigma = 0.18$) ($V = 4592, p < 0.001$) across all display content conditions (Figure 9). Participants assumed that their probability judgments would be much closer to the automation's judgments than they actually were, indicating that they were poorly calibrated with respect to understanding the difference between their performance and the automation's performance. Neither similarity measure was significantly affected by either display content condition or session order.

Discussion

Effect of Information Analysis Automation display content on Human Judgment

This research sought to investigate the impact of information analysis automation display content on human judgment performance. When participants were asked to make a joint judgment (a revision to their unaided judgment after viewing the automation's output in one of four display content conditions), judgment achievement significantly improved for all participants across all IL sessions and all display content conditions, showing the value of the automation on the human's revised judgment regardless of the display content condition. Additionally, the variance of joint judgment was reduced compared to the unaided judgment, indicating that the automation reduced variability and allowed for a range of operators with varying capabilities to maintain similar performance ranges. The automation also raised the floor of judgment achievement to support absolute levels of performance (see Figure 7).

Participants provided with additional display content pertaining to environment information (OE or OEA) had significantly higher joint judgment achievement compared to those

provided with only the automation's judgment (O). There was no significant difference between joint judgment achievement by participants in the OE and OEA display content groups. This implies that adding automation strategy information did not significantly help when participants were also provided with environment information for this judgment task.

In some cases, total system performance (as measured by human joint judgment achievement) was greater than either the human's unaided judgment achievement or the automation's judgment achievement alone. One participant in the OEA display content condition had greater joint judgment achievement than the automation's judgment achievement for every session in the IL phase. Two other participants in the OEA condition exhibited this performance in five of the six sessions in the IL phase. As a result of being provided with both automation strategy and environment information, these participants developed a strategy in which they were incorporating the automation's output into their joint judgments, rather than simply adapting to the automation's output. This is similar to behavior described in Bass and Pritchett (2008) where a participant was able to use the automation's output in his joint judgment based on the magnitude of sensor noise.

Effect of Information Analysis Automation display content on use and understanding of Automation

Neither compromise nor adaptation was found to be significant across display content conditions in the IL phase. This indicates that in this particular experiment, the display content did not affect the use of automation for the participants in terms of either adapting to the automation's judgments or maintaining consistency with their own judgments when asked to make joint judgments. However, compromise was low and adaptation was high for all participants, indicating that regardless of the display content, participants were adapting their unaided probability judgments to match the automation's judgments. Given the low unaided judgment achievement for all participants across all sessions, this was an appropriate strategy for the participants to employ.

Providing automation strategy information (OA and OEA conditions) in the IL phase did not appear to help participants understand the automation, as seen by low predictive accuracy compared to unaided judgment achievement in the prediction phase of the experiment. This indicates that providing participants with automation strategy information did not help them understand the automation any better than those provided with environment information (OE condition). It is possible that the representation of automation strategy information did not help participants or that its incremental contribution was too small. Also, participants may have lacked the domain expertise to understand and make use of the automation strategy information.

Participants assumed that their predictions of the automation would be closer than they actually were when comparing assumed similarity ($\mu = 0.90$, $\sigma = 0.16$) to actual similarity measures ($\mu = 0.56$, $\sigma = 0.18$). There was also no correlation between assumed similarity and unaided judgment achievement in the prediction phase ($r = 0.002$). All participants assumed their judgments would match the automation's judgments regardless of how successful their unaided judgments were. Similar results were found by Bass and Pritchett (2008). However, one participant in the OA display content group had much lower assumed

similarity ($\mu = 0.16$, $\sigma = 0.21$) compared to actual similarity ($\mu = 0.71$, $\sigma = 0.04$) averaged across the three sessions in the prediction phase (the only participant in this experiment with that pattern). It is possible that this participant recognized that he did not understand the automation's strategy and did not expect his judgments to align with the automation's judgment (despite seeing automation strategy information in the IL phase).

Implications for Information Analysis Automation design

These results have implications for information analysis automation design. In this experiment, the automation's imperfect judgments were tied directly to the noisy input data (imperfect sensors providing information regarding the traffic aircraft speed and heading). Although providing participants with information regarding the automation's judgment strategy may improve performance compared to those receiving only the automation's judgment, this information does not appear as beneficial to participants as information pertaining to the uncertainty in the environment when making their joint judgments.

It is possible that additional environment information allowed the participants to exploit the automation more effectively as they understood how automation judgment achievement varied based on factors in the environment. This result supports the human judgment literature related to the impact of cognitive feedback on judgment performance. In a review of over 20 human judgment experiments involving cognitive feedback, Balzer et al. (1989) found that environment information is the component of feedback with the greatest effect on human judgment performance. Environment information (graphical and statistical information related to the predictability of the environmental criterion, cue weights, and the function form relating the cues to the criterion) also increased human judgment performance in a baseball prediction task over other forms of cognitive feedback (Balzer et al., 1992). Furthermore, both trained and untrained judges also performed better with environment information during a medical diagnosis task (Gattie & Bisantz, 2006). Thus, for some judgment tasks, it may be that decision makers may not necessarily need to understand the underlying algorithm(s) used by the automation if they understand how the automation performs under different conditions, similar to results found in Masalonis and Parasuraman (2003). It is likely that when designing information analysis automation to be used in noisy environments where the automation's judgment achievement is correlated with noisy input data, it is better to show additional environment information than the automation's judgment strategy. However, research should investigate where this pattern no longer holds (i.e., where environmental noise increases, reducing the automation's judgment achievement, as well as placing a ceiling on human judgment performance given the provided cues, regardless of the environmental information provided).

In this experiment, the automation's achievement was high compared to the participant's unaided judgment achievement. However, if this were not the case, display content could have a different effect on both judgment performance and also trust of the automation. For example, when participants could observe explanations of why automation was making errors, they tended to increase adaptation to the automation, even when unwarranted (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003) and particularly when trust exceeded self-confidence (Lee & Moray, 1994). High automation error rates have also

resulted in lower subjective measures of trust in automation in numerous studies (de Vries, Midden, & Bouwhuis, 2003), particularly when participants are aware of conditions affecting the automation's reliability (Masalonis & Parasuraman, 2003).

There are different ways to represent environmental and automation strategy information. Thus, in order to fully generalize the results found here, further research using different content and representations of environment and automation strategy information in different domains should be conducted. Additionally, it would be interesting to investigate the benefit of providing environment or automation strategy alone, without including an automated judgment for the participants to consider.

A limitation of this study is that the participants were undergraduate students performing a simplified air traffic conflict prediction task with no secondary tasks to perform. Information analysis automation display content conditions should also be tested with trained operators in more naturalistic environments. In particular, it is unclear if the benefit of additional information from the automation (whether environment or automation strategy) would hold in settings where conflict detection is only one of many tasks to perform. Also, it may be that experienced participants may have been better able to understand the strategy information or alternatively have produced better independent judgments and therefore relied less on the automation.

Utility of HAJL Protocol and Measures

The HAJL protocol supports measuring judgment achievement, human-automation interaction, and automation understanding. HAJL measures in combination can further help one understand the human-automation interaction (Bass & Pritchett, 2008). For example, patterns in the unaided judgments and compromise/adaptation can identify over- or underreliance on the automation. In this experiment, unaided judgment achievement was much lower than the automation's judgment achievement, indicating appropriate use of the automation for all participants. Additionally, the assumed and actual similarity measures can point to the inability of the human judges to understand that there are substantive differences between their judgments and the automation.

HAJL also enables examining individual human judgment behavior. For example, the only participant who had higher actual similarity measures compared to assumed similarity also had higher unaided judgment achievement compared to predictive accuracy for all three sessions in the prediction phase. Furthermore, this participant's joint judgment achievement was higher than the automation's judgment achievement in four of the six sessions in the IL phase. These results indicate that although the participant did not fully understand the automation, he was aware of this lack of understanding and was still able to use the automation strategy information on the display to enhance his joint judgment achievement and total system performance.

Although the HAJL protocol employs a three-phased experiment, which might take additional time to complete, the phases support measuring the impact automation has on judgment performance but also allows one to elicit additional quantitative measures assessing human-automation interaction. The IL phase supports quantitatively measuring

use of automation through the compromise and adaptation measures. The prediction phase supports quantitatively measuring the human's understanding of the automation through the predictive accuracy and similarity measures. These quantitative measures are different from others who have used subjective techniques to measure use (Masalonis & Parasuraman, 2003) and trust of automation (Lee & Moray, 1994).

Acknowledgments

This work was supported in part by grant number UVA-03-01, Sub-Award 3029-VA ("Human-Automation Interaction Methodologies for Personal Air Vehicle Systems With Pilots With Varying Skills from the National Institute of Aerospace") from the National Institute of Aerospace and grant number T15LM009462 from the National Library of Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Aerospace, the National Library of Medicine, or the National Institutes of Health. The authors thank the students who participated in the experiment and the anonymous reviewers who provided helpful feedback on previous versions of this article.

Appendix

Please rate your agreement with the following statements on a scale from 1 to 7. (1 = not at all; 7 = very much)	
<p>I trust what other people tell me.</p> <p>1 2 3 4 5 6 7</p> <p><input type="range"/></p> <p>Not at all Very much</p>	<p>I believe that people are basically moral.</p> <p>1 2 3 4 5 6 7</p> <p><input type="range"/></p> <p>Not at all Very much</p>
<p>I trust a computer program to work properly.</p> <p>1 2 3 4 5 6 7</p> <p><input type="range"/></p> <p>Not at all Very much</p>	<p>I trust automation to do a task better than a human.</p> <p>1 2 3 4 5 6 7</p> <p><input type="range"/></p> <p>Not at all Very much</p>
<p>I distrust people.</p> <p>1 2 3 4 5 6 7</p> <p><input type="range"/></p> <p>Not at all Very much</p>	<p>I suspect hidden motives in people.</p> <p>1 2 3 4 5 6 7</p> <p><input type="range"/></p> <p>Not at all Very much</p>
<p>I expect all software to have bugs.</p> <p>1 2 3 4 5 6 7</p> <p><input type="range"/></p> <p>Not at all Very much</p>	<p>I would not let a computer do a job that I could do myself.</p> <p>1 2 3 4 5 6 7</p> <p><input type="range"/></p> <p>Not at all Very much</p>

Trust Questionnaire

References

- Balzer WK, Doherty ME, O'Connor R. Effects of cognitive feedback on performance. *Psychological Bulletin*. 1989; 106(3):410–433.

- Bass EJ, Pritchett AR. Human-Automated Judge Learning: A methodology for examining human interaction with information analysis automation. *IEEE Transactions on Systems, Man, and Cybernetics*. 2008; 38(4):759–776.
- Box, G.; Hunter, WG.; Hunter, JS. *Statistics for experimenters: An introduction to design, data analysis, and model building*. New York: Wiley; 1978.
- Brunswik, E. The conceptual framework of psychology. In: Neurath, O.; Carnap, R.; Morris, C., editors. *International encyclopedia of unified science*. Vol. 1. Chicago: University of Chicago Press; 1952.
- Brunswik, E. *Perception and the representative design of psychological experiments*. Berkeley: University of California Press; 1956.
- Burns, CM.; Hajdukiewicz, JR. *Ecological interface design*. Boca Raton, FL: CRC Press; 2004.
- Cooksey, R. *Judgment analysis: Theory, methods, and applications*. San Diego, CA: Academic Press; 1996.
- Crocoll, WM.; Coury, BG. Proceedings of the 34th Annual Meeting of the Human Factors and Ergonomic Society. Santa Monica, CA: Human Factors and Ergonomic Society; 1990. Status or recommendation: Selecting the type of information for decision aiding; p. 1524–528.
- de Vries P, Midden C, Bouwhuis D. The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies*. 2003; 58(6):719–735.
- Drews FA, Westenskow DR. The right picture is worth a thousand numbers: Data displays in anesthesia. *Human Factors*. 2006; 48(1):59–71. [PubMed: 16696257]
- Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP. The role of trust in automation reliance. *International Journal of Human-Computer Studies*. 2003; 58(6):697–718.
- Endsley MR, Kiris EO. The out-of-the-loop performance problem and level of control in automation. *Human Factors*. 1995; 37(2):381–394.
- Endsley MR, Kaber DB. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*. 1999; 42(3):462–492. [PubMed: 10048306]
- Gattie G, Bisantz A. The effects of integrated cognitive feedback components and task conditions on training in a dental diagnosis task. *International Journal of Industrial Ergonomics*. 2006; 36(5): 485–497.
- Hursch CJ, Hammond KR, Hursch JL. Some methodological considerations in multiple-cue probability. *Psychological Review*. 1964; 71:42–60. [PubMed: 14105718]
- Kaber DB, Endsley MR. Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*. 1997; 16(3):126–131.
- Kaber DB, Endsley MR. The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*. 2004; 5(2):113–153.
- Kirlik, A. Requirements for psychological models to support design: Towards ecological task analysis. In: Flach, JM.; Hancock, PA.; Caird, JK.; Vicente, KJ., editors. *Global perspectives on the ecology of human-machine systems*. Vol. 1. Hillsdale, NJ: Lawrence Erlbaum; 1995. p. 68–120.
- Lee J, Moray N. Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*. 1994; 40(1):153–184.
- Lorenz B, Di Nocera F, Röttger S, Parasuraman R. Automated fault-management in a simulated spaceflight micro-world. *Aviation, Space, and Environmental Medicine*. 2002; 73(9):886–897.
- Masalonis, AJ.; Parasuraman, R. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomic Society. Santa Monica, CA: Human Factors and Ergonomic Society; 2003. Effects of situation-specific reliability on trust and usage of automated air traffic control decision aids; p. 533–537.
- Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*. 2000; 30(3):286–297.

- Pritchett, AR.; Vandor, B. Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomic Society. Santa Monica, CA: Human Factors and Ergonomic Society; 2001. Designing situation displays to promote conformance to automatic alerts; p. 311-315.
- Rovira E, McGarry K, Parasuraman R. Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*. 2007; 49(1):76. [PubMed: 17315845]
- Sarter NB, Schroeder B. Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors*. 2001; 43(4):573–583. [PubMed: 12002006]
- Sarter NB, Woods DD. Pilot interaction with cockpit automation: Operational experiences with the flight management system. *The International Journal of Aviation Psychology*. 1992; 2(4):303–321.
- Sarter NB, Woods DD. Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *The International Journal of Aviation Psychology*. 1994; 4(1):1–28.
- Seong, Y.; Bisantz, AM. Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomic Society. Santa Monica, CA: Human Factors and Ergonomic Society; 2002. Judgment and trust in conjunction with automated decision aids: A theoretical model and empirical investigation; p. 423-427.
- Seong Y, Bisantz A. The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*. 2008; 38(7-8):608–625.
- Skjerve A, Skraaning G. The quality of human-automation cooperation in human-system interface for nuclear power plants. *International Journal of Human-Computer Studies*. 2004; 61(5):649–677.
- Tucker LR. A suggested alternative formulation in the developments of Hirsch, Hammond & Hirsch and by Hammond, Hirsch & Todd. *Psychological Review*. 1964; 71:528–530. [PubMed: 14216901]
- Vicente KJ. Ecological Interface Design: Progress and challenges. *Human Factors*. 2002; 44(1):62–78. [PubMed: 12118874]
- Vicente KJ, Rasmussen J. The ecology of human-machine systems II: Mediating “direct perception” in complex work domains. *Ecological Psychology*. 1990; 2(3):207–249.
- Wickens, CD.; Li, H.; Santamaria, A.; Sebok, A.; Sarter, NB. Human Factors and Ergonomics Society Annual Meeting Proceedings. Santa Monica, CA: Human Factors and Ergonomic Society; 2010. Stages and levels of automation: An integrated meta-analysis; p. 389-393.
- Wickens, CD.; Xu, X. Automation trust, reliability and attention, HMI 02 03(AHDF Technical Report No AHFD-02- 14/MAAD-02-2). Savoy: University of Illinois Aviation Human Factors Division; 2002.
- Woods, DD. The cognitive engineering of problem representation. In: Weir, GRS.; Alty, JL., editors. *Human-computer interaction in complex systems*. New York: Academic Press; 1991. p. 169-188.

Biographies

Ellen J. Bass is an associate professor in the Department of Systems and Information Engineering at the University of Virginia. She has over 28 years of research and design experience in the systems engineering of complex, dynamic systems. Her research involves modeling human judgment and decision making in complex domains in order to inform the design of decision support systems and training. She applies this research in the domains of air transportation, bioinformatics, emergency management, healthcare, health promotion, and meteorology.

Leigh A. Baumgart is a PhD candidate at the University of Virginia in the Department of Systems and Information Engineering. She received her BA in physics at the State University of New York at Geneseo and her MS in systems engineering at the University of Virginia.

Kathryn Klein Shepley is an associate program manager for the Department of Terminal and Tower Systems Engineering and Evolution at the MITRE Corporation's Center for Advanced Aviation System Development. She received her BA in economics and MS in systems engineering from the University of Virginia.

Figure 1.
Measures From the Human-Automated Judge Learning Training Phase.

Figure 2.
Measures From the Human-Automated Judge Learning Interactive Learning Phase.

Figure 3.
Measures From the Human-Automated Judge Learning Prediction Phase.

Figure 4. Data Entry Display and Navigation Display From the Interactive Learning Phase Prior to Making a Joint Judgment in the Display Content Condition That Included Automation Strategy Information (OA and OEA).

Figure 5. Environment Information Display Available During the Interactive Learning Phase Prior to Making a Joint Judgment in the Display Content Condition That Included Environment Information Related to Sensor Noise (OE and OEA).

Figure 6.
Human Unaided and Joint Judgment Achievement During the IL Phase by Session Order.

Figure 7.
Human Unaided and Joint Judgment Achievement During the IL Phase by Display Content.

Figure 8.
Untransformed Compromise and Adaptation During the IL Phase by Session Order.

Figure 9.
Untransformed Assumed and Actual Similarity During the Prediction Phase by Display Content.

Table 1
Dependent Measures

HJL Phase	Derived Measure	Description	Correlation Between
Training	r_{a1}	Human unaided judgment achievement	Environmental criterion and human unaided judgments
IL	r_{a1}	Human unaided judgment achievement	Environmental criterion and human unaided judgment
	r_{a2}	Human joint judgment achievement	Environmental criterion and human joint judgments
	Compromise	Degree to which humans compromise their unaided judgment	Human's unaided and joint judgments
	Adaptation	Degree to which humans adapt their unaided judgments to match the automation	Automation's judgments and the human's joint judgments
Prediction	r_{a1}	Human unaided judgment achievement	Environmental criterion and human unaided judgments
	Predictive accuracy	Degree to which human is able to predict the automation's judgment	Human's predictions of the automation's judgments and the automation's judgments
	Actual similarity	Degree to which human's unaided judgment matches the automation's judgment	Human's unaided judgments and automation's judgments
	Assumed similarity	Degree to which humans' prediction of the automation matches their unaided judgment	Human's unaided judgments and human's predictions of the automation's judgments