ELSEVIER

International Journal of Industrial Ergonomics 28 (2001) 85–97

International Journal of
Industrial
Ergonomics

www.elsevier.nl/locate/ergon

# Assessment of operator trust in and utilization of automated decision-aids under different framing conditions

Ann M. Bisantz*, Younho Seong

*Department of Industrial Engineering, University at Buffalo, State University of New York, Amherst, New York 14260, USA*

Received 15 December 2000; received in revised form 24 January 2001; accepted 24 January 2001

**Abstract**

Computerized aids may be used to support decision-making and control in a variety of complex, dynamic arenas. For instance, such systems have been introduced into industrial settings as the means to implement automated control or support decision-making activities such as fault detection and recovery. Of interest in these systems is the extent to which operators utilize and trust such systems, in terms of their ability to successfully control systems, or the information or decision support they provide, particularly under conditions of potential failure. A theoretical framework to describe potential factors affecting these issues, and an experiment to investigate the role of failure cause on trust and system utilization, are described. Results provide some support for factors in the theoretical framework, and also demonstrated the use of an empirically developed trust scale.

**Relevance to industry**

As manufacturing environments increasingly rely on computerized and automated systems for control and human operator support, it is necessary to understand the situational factors which could impact operators' use of such systems. This paper describes a framework which could be used to investigate trust in industrial automation settings, as well as a rating scale which could be applied. © 2001 Elsevier Science B.V. All rights reserved.

*Keywords:* Trust; Automation; Decision-aids

## 1. Introduction

As automated systems to support control and decision-making in complex environments such as modern manufacturing systems proliferate, investigations into factors affecting the development of human trust in such systems, and the role trust plays in reliance and use of such systems, have gained importance. The dependability or reliability of systems have been linked to the development and maintenance of trust (e.g., Sheridan, 1988; Muir and Moray, 1996), and thus, questions related to automation failures are of relevance.

Issues of human trust are particularly interesting in situations when automation or decision-aids may fail not only through natural faults (e.g., hardware failures) or design deficiencies, but through the targeted, intentional interference of an adversary. We have addressed these issues primarily with respect to automated decision-support in a military context; however, concerns

*Corresponding author.
*E-mail address:* bisantz@eng.buffalo.edu (A.M. Bisantz).

PII: S0169-8141(01)00015-4

regarding computer hacking, or industrial sabotage make them relevant in non-defense environments as well. Additional fault-related factors, described in more detail below, also apply more generally. Regardless of the application environment, to design successful automated tools or decision-aids, it is necessary to understand the factors affecting operators' decisions to rely on, or use, those systems.

Generally, research from both social science and engineering perspectives agree that trust is a multi-dimensional, dynamic concept. For example, Deutsch (1958) claimed that trust included two concepts: expectation (predictability) and motivational relevance, while Rotter's (1967) defined trust in terms of an expectancy on the part of one individual or group that statements of another individual or group can be relied on. Rempel et al. (1985) concluded that trust would progress in three stages over time, from predictability, to dependability to faith. Muir and Moray (1996) extended these three factors, and developed an additive trust model that contained six components: predictability, dependability, faith, competence, responsibility, and reliability. Sheridan (1988) also suggested possible factors, including reliability, robustness, familiarity, understandability, explication of intention, usefulness, and dependence.

Empirical results have shown that people's strategies with respect to the utilization of an automated system may be affected by their trust in that system. For example, Muir and Moray (1996) and Lee and Moray (1994) studied issues of human trust in simulated, semi-automated pasteurization plants. In these experiments, participants were asked to control a simulated pasteurization process either by controlling pump and heating sub-systems, or by activating an automated controller, in order to produce pasteurized liquid. Different system aspects were altered to see how participants' trust in systems components, such as the automated controller, was affected. In particular, Muir and Moray (1996) altered the quality of the pump systems by introducing either random or constant errors in its ability to maintain a set-point, introduced errors into the pump's display of its pump rate (although actual pump rate was error-free), and the performance of the automated controller in setting and maintaining appropriate settings for the pump. Lee and Moray introduced faults into pump performance (Lee and Moray, 1992) or faults into either automatic or manual controllers (Lee and Moray, 1994). Trust was measured both subjectively, using rating scales and objectively, by logging participants' actions (e.g., hypothesizing that more or less use of an automated control system implied more or less trust in that automated system).

These studies showed, among other results, that operators' decisions to utilize either automated or manual control depended on their trust in the automation and their self-confidence in their own abilities to control the system. For instance, Lee and Moray (1994) found that reliance on an automated controller depended on the difference between participants' own self-confidence, and their trust in the automation. Additionally, results showed that trust depended on current and prior levels of system performance, the presence of faults, and prior levels of trust. For example, trust declined, but then began to recover, after faults were introduced (Lee and Moray, 1992). Lerch and Prietula (1989) found a similar pattern in participants' confidence in a system for giving financial advice: confidence declined after poor advice was given, then recovered, but not to the initial level of confidence. The magnitude of automation errors or faults (Lee and Moray, 1994) and the error consequences (Masalonis et al., 1998), also affect the degree to which trust is diminished.

Masalonis and Parasuraman (1999) assert that trust is one intervening variable between an automated system and its use: that is, people may or may not use a system because of their trust in it, and their trust in part depends upon their experience using or relying on the system. To understand and predict automation use, then, it is necessary to specify factors in system behavior which in turn affect operators' experiences, and thus their levels of trust.

A multi-dimensional taxonomy of factors affecting trust in systems has been developed, based in part by factors manipulated in the experiments cited above, which can be applied to address these issues (Llinas et al., 1998). The framework integrates and systematically varies a set of dimensions related to automation failures, which may affect trust in automated systems. The following dimensions are included in the framework:

### 2.1. Locus of failure

The location at which the potential for failure exists. Two potential dimensions contribute to this factor: component and surface-depth.

#### 2.1.1. Component dimension

Automated decision aids or controllers could produce erroneous or anomalous outputs due to failures at multiple system levels. For instance, failures in the environment or controlled system of interest, the automation or decision support algorithms, or in the human–computer interface, could contribute to an anomalous or unexpected behavior of the system–automation combination. In the pasteurization experiments (Lee and Moray, 1994; Muir and Moray, 1996) the quality of system performance was manipulated at different system levels. Faults or random errors were introduced at the level of the environment—the process control system itself (i.e., the pumps), and at the level of a control intervention (i.e., the automated controller). Three levels can be described, as follows:

- *Environment*. The physical environment corresponds to the actual situation that is taking place, and system being controlled. For example, in a process control setting, the states of pumps and heaters can be observed and controlled, and those components can be subject to failure.
- *Automated aiding system*. The algorithms and processes which automatically control processes, or provide automated decision-support based on measurement of state variables and subsequent processing, could also be subject to fault.
- *Interface*. Finally, one can consider a third, interface level in which the results of the algorithms are displayed to the operator, in order to aid decision-making or system monitoring. Inaccurate results, or manipulated displays, could be presented to the operator.

#### 2.1.2. Surface-depth dimension

A second related dimension along which investigations of performance in automated systems can vary is a surface-depth dimension. The surface level corresponds to the information available about the environment (as formalized in Brunswik's Lens Model; Cooksey, 1996; Hammond et al., 1975), whereas the depth level corresponds to the actual state of the environment. In an industrial process control environment, surface level features would be the observable outputs from sensors, or algorithms. Depth level features would be the actual operations of the sensors or algorithms themselves. The manipulations performed by Muir and Moray (1996) can be described in terms of this dimension. Muir and Moray (1996) manipulated both the characteristics of the pump itself (the depth level) and the display of the pump rate (the surface level). This surface-depth dimension can be applied at the environment and aiding system dimensions described above, resulting in five combinations. Since the interface level does not perform any processing, but is a representation of the results of the aiding system level, it is best considered at the surface level.

### 2.2. Causes of failure or corruption

Systems can fail due to different intents. For instance, system components or automation may fail due to unintentional hardware or software difficulties. In some environments, failures may be introduced intentionally, through acts of sabotage (e.g., through military style "information attacks", acts of computer hacking, or industrial sabotage).

### 2.3. Time patterns of failure

This dimension reflects the dynamic or time-dependent characteristics of the degradation. Failures, sabotage, and subterfuge can occur not only as failures or degradations at a particular point in time, but also in a continuing fashion.

Additionally, failures can occur with patterns that are either predictable or unpredictable.

## 3. Research questions

The research questions of interest in this experiment were to see if and how the cause of an automation failure differentially influenced subjective assessments of trust in an automated decision-aid, and selection of information to use (i.e., implying trust in the information). Inclusion of failure cause as a factor in the taxonomy described above in essence hypothesizes that differences in levels of that factor will impact trust and automation use. By testing this hypothesis, this study can provide evidence regarding the adequacy of the taxonomy.

Measurement of trust is also an issue. Many researchers have used questionnaires to measure subjective feelings of trust (Larzelere and Huston, 1980; Rempel et al., 1985; Singh et al., 1993; Lee and Moray, 1994; Muir and Moray, 1996), however, these questionnaires have been based on theoretical rather than empirical notions of trust dimensions. Jian et al. (2000) used a three-phase experiment, in which words related to trust were collected, rated, and clustered, to empirically develop a twelve item trust questionnaire (shown in Table 1). In the current study, data based on this trust questionnaire was collected to assess the sensitivity of the questionnaire to different points in theoretical framework.

Table 1
Items in the trust questionnaire

| The system is deceptive |
|---|
| The system behaves in an underhanded manner |
| I am suspicious of the system's intent, action, or output |
| I am wary of the system |
| The system's action will have a harmful or injurious outcome |
| I am confident in the system[a] |
| The system provides security[a] |
| The system has integrity[a] |
| The system is dependable[a] |
| The system is reliable[a] |
| I can trust the system[a] |
| I am familiar with the system[a] |

[a] Positively framed questions.

## 4. Method

### 4.1. Participants

Participants were recruited from the university community and were paid $6.50/h for their participation. There were 10 participants in each condition, for total 30 participants. Among them, 20 participants were male. There were 23 students who had taken one or more probability-related courses. The average age was 26 years.

### 4.2. Apparatus

Experiments were run using a low fidelity simulation of an anti-air warfare task. Air contacts, shown using military symbols, were overlayed on a map displayed on a simulated radar screen. Participants could select air contacts, obtain information about the contacts, and make identifications of contacts as hostile or friendly using a mouse and pull-down menus. Two sources of information could be selected for each contact. An information window gave contact range, bearing, radar signature, altitude, and speed. Contact parameters were assigned probabilistically and were not completely diagnostic of an aircraft being hostile or friendly (e.g., both a friendly and hostile aircraft could share the same radar signature or speed). Additionally, participants could access a decision-aid which provided upper and lower limits of a probability range that the aircraft was friendly. This interval was generated based on the actual probability that a contact with a particular pattern of altitude, speed, and radar was friendly. The displayed interval was computed based on this probability, plus and minus randomly generated errors that ranged from 0% to 5% (two errors were generated so that the true probability did not always fall at the center of the interval). The simulation was displayed on 17″ high resolution, color monitors, and participants interacted with the simulation using a keyboard and mouse.

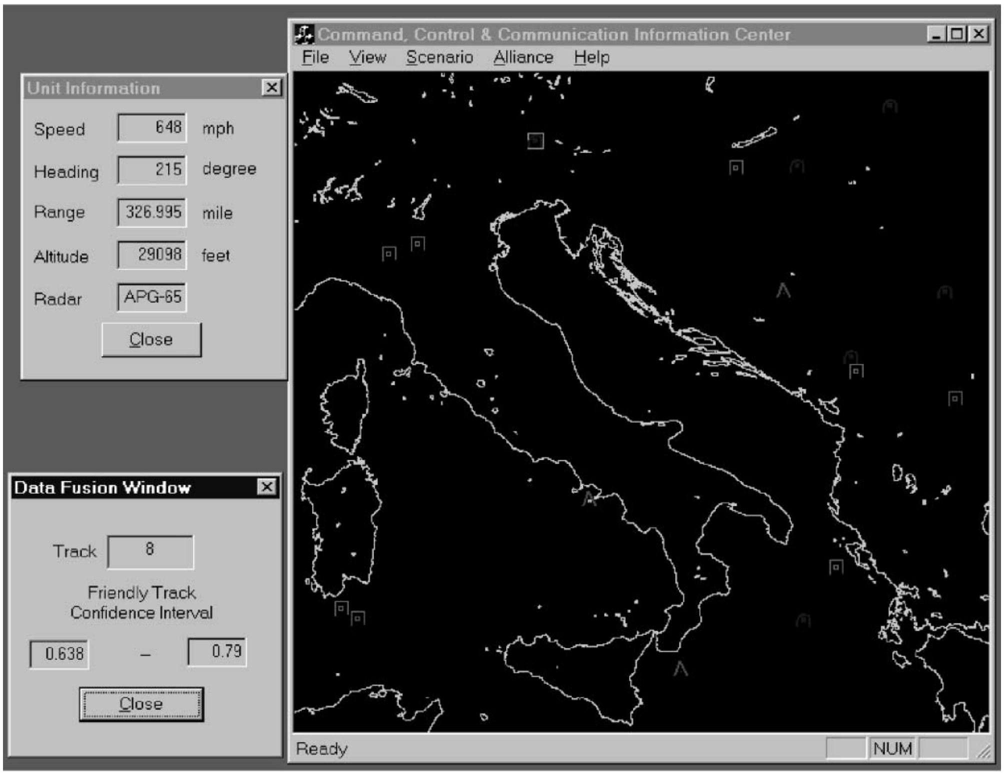
### 4.3. Independent variables

The primary independent variable of interest was failure cause. Training materials provided to the participants differed in the description provided about the potential failure of the decision-aid: participants in the sabotage condition were told that the decision-aid provided to help them make decisions "may be subject to intentional interference with the computer system by enemy forces which may cause the aid to produce unreliable estimates". Participants in the non-intentional failure condition were told that the aid "may be subject to occasional hardware or software problems which may cause the aid to produce unreliable estimates". Participants in the control condition were not told about any potential for failures.

Participants also completed six sessions (described below) which were treated as independent variables in the analysis.

### 4.4. Experimental task

During the experiment, participants clicked on unknown contacts, requested either the information window or decision-aid window to obtain information, and made identifications of the contact as hostile or friendly (see Fig. 1). Participants performed the task for six, 20 min sessions. In each session, there were between 37 and 50 contacts to identify. During the first two sessions, the decision-aid provided accurate probability intervals as described above. During the first 10 minutes of the third session, an error was introduced into the interval. This error was



Speed 648 mph
Heading 215 degree
Range 326.995 mile
Altitude 29098 feet
Radar APG-65

Track 8
Friendly Track
Confidence Interval
0.638 – 0.79

Fig. 1. Picture of the simulation showing the map-based radar display, data fusion (decision-aid) window, and track information window. Participants could use the mouse to select tracks or contacts on the radar display, and then use the pull down menus to select to see either or both of the data fusion and track information windows.

computed by either shifting the intervals to one direction 5% or widening each direction 2.5%. Ten minutes after the onset of the simulation for session 3, the participant was informed that an error in the decision-aid had been detected and corrected (except for the control condition, in which no message was generated). Notification was provided because the intent of the experiment was not to assess fault detection, but rather operator responses to a known fault. Finally, three sessions without errors were completed after the error session.

Thus, in terms of the above taxonomy, this study investigated differences between sabotage and non-intentional causal factors, at the automation component level. The type of failure was a degradation, at the depth level of the surface-depth dimension.

In order to investigate the impact of trust, particularly after the decision-aid error, on use of the information sources, participants were limited in the number of times they could access the information window and the decision-aid window. Participants could access these windows a total of 1.5 × the number of contacts in the session (that is, they could not access both windows for all contacts). Both types of windows disappeared from the screen 5 s after they were requested. Participants' scores reflected this tradeoff, as well as their correctness in identifying contacts. Participants were informed of the scoring method, and were shown their score after each session.

Participants were given task instructions to read, and given a 10 min practice session during which time the experimenter showed the participant the mouse and menu functions. Failure cause (sabotage, hardware failure, or unspecified) was a between-subjects variable, while session was a within-subjects factor. Ten participants were included each cell, and performed the experiment for all six sessions. Dependent variables included the correctness of participants' actions identifying unknown contacts, the number of times participants accessed either the decision-aid (data fusion) window or the information window, the score shown to participants, and the subjective trust ratings.

A 12 item trust questionnaire was given after sessions 1, 3, and 6. Questionnaire items are shown in Table 1, and were developed experimentally (see Jian et al., 2000, for a detailed description of the questionnaire). Participants were asked to "rate the decision-aiding system on the following scales". The questionnaire was implemented on a computer; participants could move a slider bar between end points labeled "not at all" and "very much". The scale was also marked with seven divisions.

## 5. Results

Dependent measures included task performance measures, information use measures, and responses to the questionnaire. An analysis of variance was used to investigate difference in these measures between conditions. Failure cause was treated as a between-subject, fixed factor, and session was treated as a within-subject, fixed factor. For the questionnaire responses, question was a within-subjects factor.

First, the percentage of contacts correctly identified was analyzed. The percent correct measure is a ratio of the number if aircraft identified correctly in a session, and the total number of aircraft identified in a session (which varied from session to session). Results are shown in Fig. 2. Averaged across failure conditions, mean percent correct ranged from 61% (session 2) to 85% (session 3), with an overall mean of 74%. There was a significant fault by session interaction ($F_{10,135} = 2.539$, $p = 0.008$), significant session effect ($F_{5,135} = 67.98$, $p < 0.000$), and marginally significant fault condition effect ($F_{2,27} = 2.634$, $p = 0.09$). Inspection of Fig. 2 indicates, however, that though there was a significant interaction, the pattern of results was highly similar across session.

Use of the two sources of information—the information window, and the decision-aid window, was analyzed in several ways. First, the overall percent use of the information window was computing by dividing the number of times that window was selected by the total number of times either the information or decision-aid window was selected. Note that the overall percent use of the information window and the decision-aid window

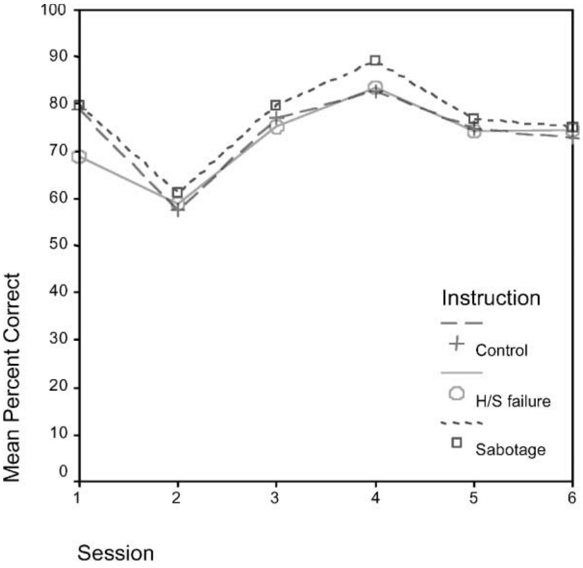


Fig. 2. Mean overall percentage of information window use for different failure conditions.

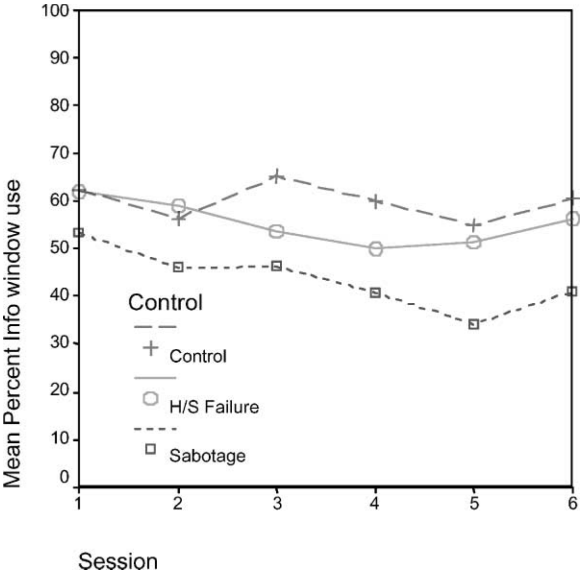


Fig. 3. Mean percent use of the information window per number of contacts per session.

necessarily sum to 100; thus, only the overall percent use of the information window was analyzed, and is shown in Fig. 3. There was a significant session effect ($F_{5,135} = 2.487$, $p = 0.034$). From Fig. 3, use of the information window tended to decrease from sessions 3 to 5, then recover in the final session. It follows that use of the decision-aid window thus increased from sessions 3 to 5, then declined in the final session. Inspection of Fig. 3 also suggests that participants in the sabotage condition tended to use the information window less, and thus the decision-aid window more, although the difference between fault conditions was not statistically significant ($F_{2,27} = 1.213$, $p = 0.313$). There was no significant session by fault condition interaction ($F_{5,135} = 0.54$, $p = 0.849$).

Use of each information sources, per total number of contacts in a session, was also investigated. These measures scale the measures of information use by the number of contacts to be investigated and identified. The use of the information window and the decision-aid window were analyzed. While there was no significant effect of fault condition on the use of the decision-aid window ($F_{2,17} = 0.291$, $p = 0.750$), there was a significant impact of fault condition on information window use ($F_{2,27} = 2.943$, $p = 0.07$). Inspection of Fig. 4 indicates that participants in the sabotage condition seemed least likely to select the information window on a track-by-track basis, and participants in the control condition the most likely, while across the three conditions there was a trend toward decreased information window use after the third session. There were no significant session effects, or interactions between session and fault condition, for either use of the decision-aid window or information window.

Inspection of cases where only one of the two windows was selected shows a similar pattern, though there were not significant differences between fault condition (see Fig. 5). Use of the decision-aid window changed over significantly over sessions ($F_{5,135} = 3.852$, $p = 0.003$), tending to increase after the third session and decreasing from the fifth to sixth session. There were no significant session by fault condition interactions for either measure.

In general, across the information use measures, there appears to be a trend for participants in the sabotage condition to make less use of the information window than participants in the other

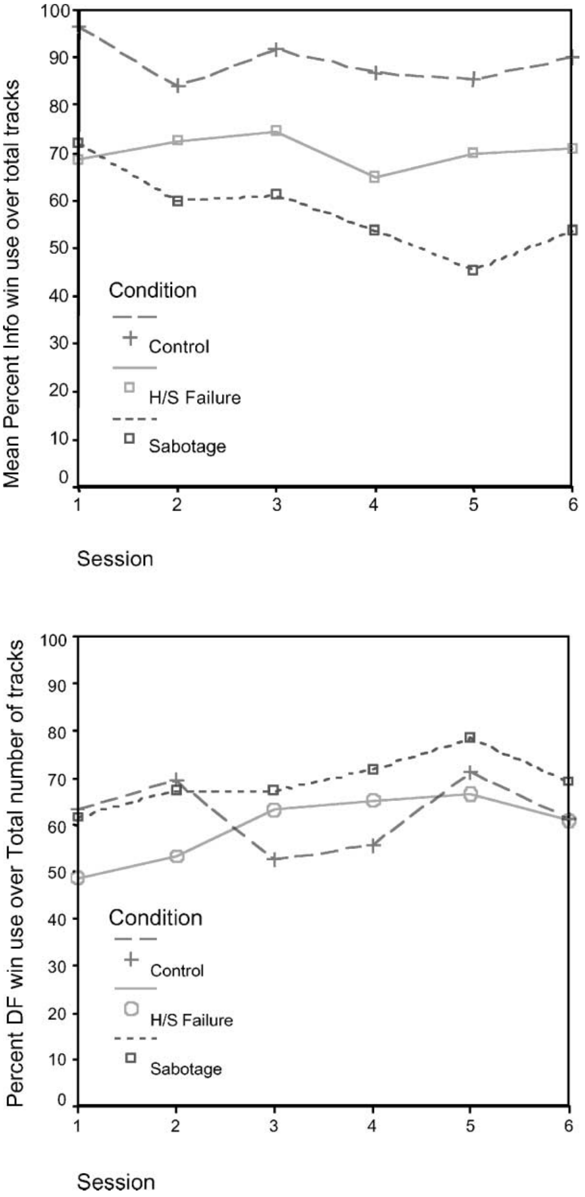


Fig. 4. Percent of use of either information window (top) or decision-aid window (bottom) per contact, by fault condition.

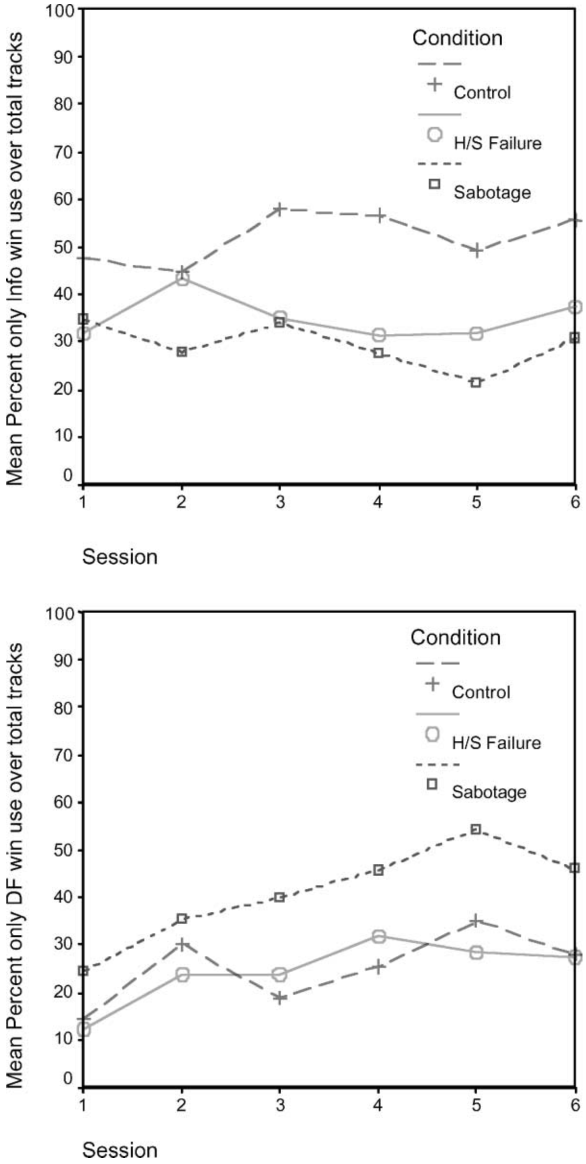


Fig. 5. Percent of contacts where either the information window (top) or decision-aid window (bottom) was used exclusively, by fault condition.

two conditions. Also, there was a trend for participants across conditions to make less use of the information window and more use of the decision-aid window after the third session, perhaps indicating that participants attributed the failure to the information window rather than the decision-aid window.

Questionnaire data was also analyzed. For scoring, responses on the scale were coded from 0 (not at all) to 7 (very much). First, analysis of the responses to all 12 questions indicated that there was a main effect of question ($F_{11,297} = 4.768$, $p < = 0.000$): questions did tap into different concepts of trust.

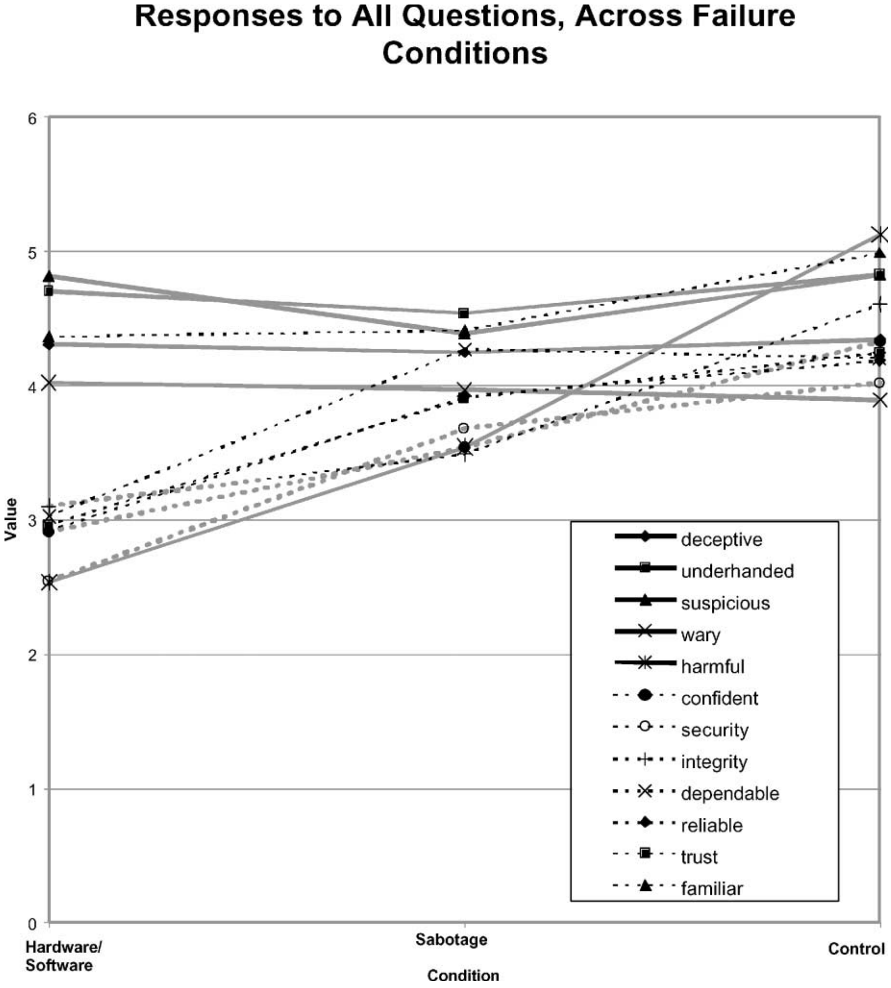


Fig. 6. Responses to trust questionnaire, by failure condition.

There was also a significant question by fault condition interaction ($F_{22,297} = 1.64$, $p = 0.037$), question by session interaction ($F_{22,594} = 1.607$, $p = 0.04$), and question by session by fault condition interaction ($F_{44,594} = 1.424$, $p = 0.041$). There was not a significant session or session by condition effect ($F_{2,54} = 0.136$, $p = 0.873$; $F_{4,54} = 0.474$, $p = 0.754$, respectively). Further interpretation can be made by inspecting Fig. 6. For the control and sabotage conditions, responses for both positively framed questions (displayed with dashed lines) tended to be similar to, and intermingled with, responses for negatively framed questions (displayed with solid lines). For the hardware/software failure condition, responses to negatively framed questions tended to be clustered together, and higher, than positive responses. Responses to most positively framed questions appeared highest in the control condition, and lowest in the hardware/software failure condition. This indicates that participants tended to have higher negative and lower positive feelings regarding trust when they were told that failures should be attributed to hardware or software failures.

Further analysis was conducted of negatively framed, and positively framed questions, separately. For negatively framed questions (Figs. 7 and 9), there was a significant difference between

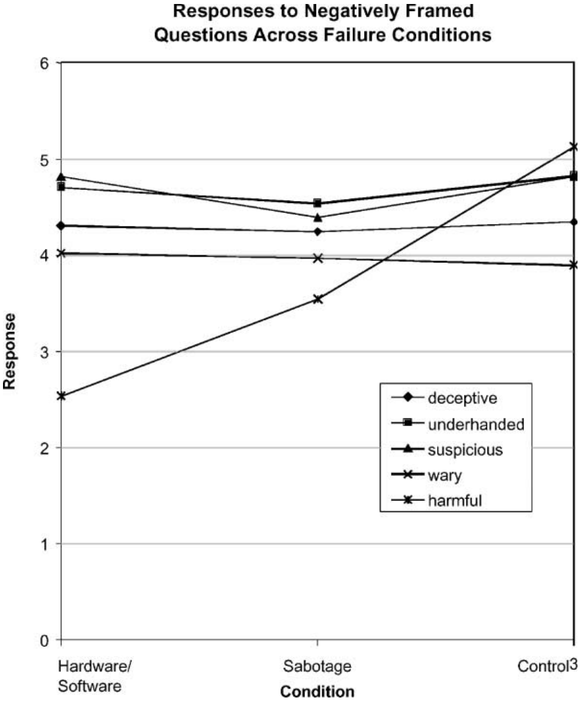


Fig. 7. Responses to negatively framed questions by failure condition.

questions ($F_{4,108}$ =5.061, $p=0.011$), a significant question by condition interaction ($F_{8,108}=10.348$, $p=0.033$), and a significant question by condition by session interaction ($F_{16,216}=1.337$, $p=0.047$) again indicating that the questions in fact tapped into different concepts of trust as conditions varied. In particular, the notion of harm appeared different across conditions, with least agreement with the statement "The system's action will have a harmful or injurious outcome" occurring in the hardware–software condition, and most in the control condition. Additionally, responses tended to decline across sessions in the sabotage condition. Similar tests were performed on the positively framed questions (Figs. 8 and 9). There was a significant effect of question ($F_{6,162}=8.502$, $p<0.000$), significant question by session interaction ($F_{12,324}$, $p=0.005$) and significant question by session by condition interaction ($F_{24,324}=1.575$, $p<0.044$). Agreement with the question regarding familiarity tended to rise across sessions, particularly for the sabotage and control conditions, while overall responses declined slightly for the sabotage condition, and though lower overall, rose for the hardware/software condition.

Finally, responses to all negatively framed questions, and all positively framed questions, were pooled, and analyzed to see if there was an overall trend to assessing trust in the information systems positively or negatively. As shown in Fig. 10, responses to positively framed questions were significantly lower than negatively framed questions ($F_{1,27}=2.96$; $p=0.096$). There were no significant interactions with session or condition.

## 6. Discussion

In general, results from the experiment indicated that participants showed some tendency to reduce their use of one of the two information sources (the information window) after a fault occurred, and that participants in the sabotage condition tended to be less likely to use the information window. These results indicate that participants in the sabotage condition were more suspicious of the information window rather than the decision-aid window, suggesting possibly that participants may attributed the potential for faults to the information window and not the decision-aid. Future experiments could clarify this point by expanding the training and explanations provided to participants, as well as the magnitude and description of the errors in the aid. Results from the questionnaire data indicated that while participants in the sabotage condition showed a decline in their assessments of positive trust factors over sessions, their assessments of negative trust factors also declined. Additionally, participants in the hardware/software fault condition tended to give the lowest scores to positive characteristics of trust, indicating that subjectively, they had the least trust in the decision-support information.

Overall, the results are interesting, for several reasons. First, given the tendency for some differences in aid utilization between fault conditions, as well as differences in subjective assessments of certain characteristics of trust, the inclusion of this dimension—causal factors—in the theoretical framework described above is supported. Future work in this area, both in terms

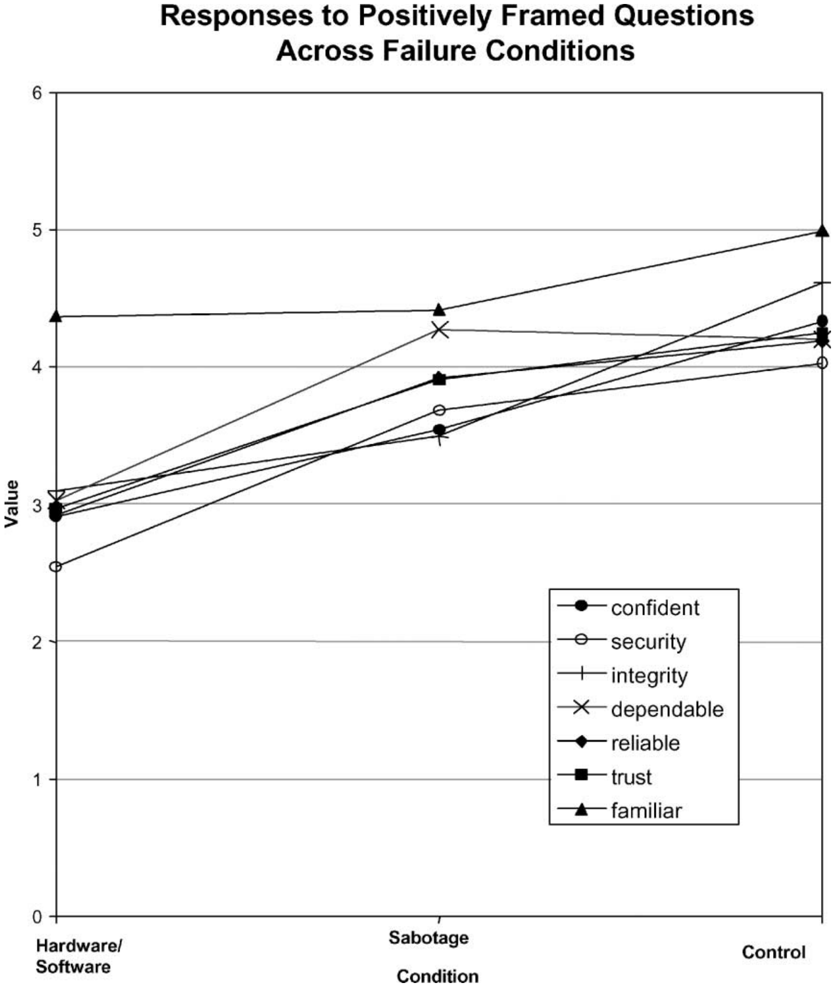


Fig. 8. Responses to positively framed questions, by failure condition.

of laboratory experimentation, and the development of displays and aids to support interaction with automated systems, should continue to consider such a dimension. For example, any automated algorithms to detect potential faults should attempt to identify and display the type of fault with respect to intentional or natural failure conditions. Second, the tendency for a pattern of less use of the information window, then some recovery by the last session is consistent with prior work on dynamic patterns of trust, suggesting both that sabotage of a decision systems may induce disuse for a period of time but not permanently induce distrust. Although results presented here were consistent with prior studies, they were of limited magnitude. Subjective assessments of trust also did not follow a clear temporal pattern. Further work is needed to more definitively determine the types of failures, and the magnitudes of such failures, that could cause more extreme effects on trust in, and use of, decision-aids.

Finally, results from the trust questionnaire indicate that participants responded differently to different items in the trust questionnaire, and that though they had overall higher negative beliefs about the system than positive beliefs, there was not a clear differentiation in responses between

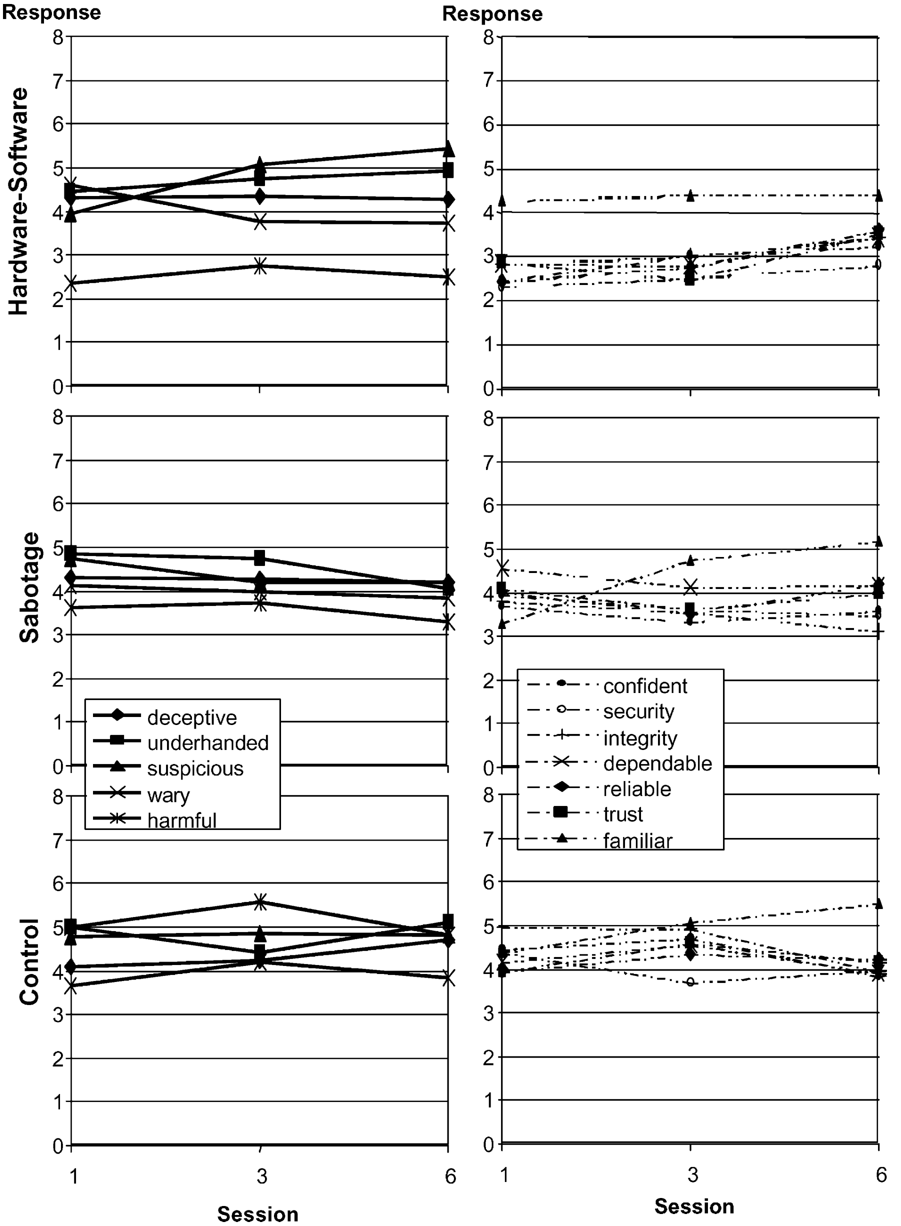


Fig. 9. Responses to negatively and positively framed questions, by failure condition and session.

responses to positive and negative scale items. These results support those found in developing the trust questionnaire—particularly, the fact that there could be different concepts within trust, and that both positively and negatively framed concepts could be associated with trust—and thus provide some validation for the questionnaire items. Additionally, there were difference in responses to the questionnaire based on fault condition, suggesting both that the questionnaire was sensitive to differences in task conditions, and also that the different levels of fault conditions

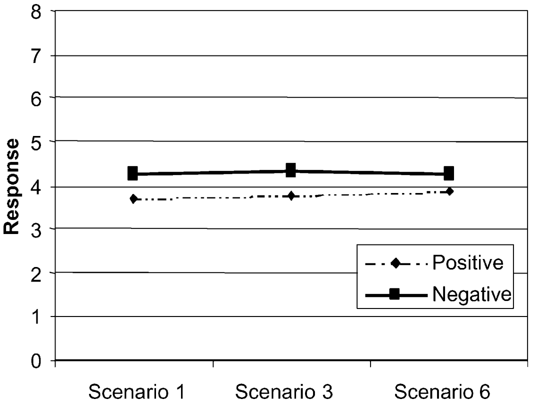


Fig. 10. Responses for negatively and positively framed questions, by session.

were in fact affecting participants' subjective responses of trust differently.

Overall, results indicated that as suggested by the proposed theoretical framework, differences in fault causation may impact operators' trust in, and use, of information systems, and that the trust questionnaire could identify differences in concepts of trust as conditions varied.

## Acknowledgements

The authors acknowledge the support of the Air Force Research Laboratory, Human Effectiveness Directorate (G. Kuperman, sponsor) and UB Center for Multi Source Information Fusion (J. Llinas, director) for support of this work.

## References

Cooksey, R.W., 1996. Judgment Analysis: Theory, Methods, and Applications. Academic Press, San Diego, CA.

Deutsch, M., 1958. Trust and suspicion. Journal of Conflict Resolution 2 (4), 265–279.

Jian, J.Y., Bisantz, A.M., Drury, C.G., 2000. Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics 1 (4), 53–71.

Hammond, K.R., Stewart, T., Brehmer, B., Steinmann, D.O., 1975. Social judgment theory In: Kaplan, M.F., Schwartz, S. (Eds.), Human Judgment and Decision Processes. Academic Press, New York.

Larzelere, R.E., Huston, T.L., 1980. The Dyadic Trust Scale: Toward understanding interpersonal trust in close relationships. Journal of Marriage and the Family 42 (3), 595–604.

Lee, J.D., Moray, N., 1992. Trust, control strategies and allocation of function in human–machine systems. Ergonomics 35 (10), 1243–1270.

Lee, J.D., Moray, N., 1994. Trust, self-confidence, and operators' adaptation to automation. International Journal of Human–Computer Studies 1 (40), 153–184.

Lerch, F.J., Prietula, M.J., 1989. How do we trust machine advice? In: Salvendy, G., Smith, M.J. (Eds.), Designing and Using Human–Computer Interface and Knowledge Based Systems. Elsevier Science Publishers, North-Holland, Amsterdam, pp. 410–419.

Llinas, J., Bisantz, A., Drury, C.G., Seong, Y., Jian, J., 1998. Studies and analyses of aided adversarial decision-making. Phase 2: Research on Human Trust in Automation. April, 1998. Center for Multisource Information Fusion, State University of New York, Buffalo.

Masalonis, A.J., Duley, J.A., Galster, S.M., Castano, D.J., Metzger, U., Parasuraman, R., 1998. Air traffic controller trust in a conflict probe during Free Flight. Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, pp. 1601–1606.

Masalonis, Parasuraman, 1999. Trust as a construct for evaluating automated aids: past and future theory and research. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, pp. 184–188.

Muir, B.M., Moray, N., 1996. Trust in automation: Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics 39 (3), 429–460.

Rempel, J.K., Holmes, J.G., Zanna, M.P., 1985. Trust in close relationships. Journal of Personality and Social Psychology 49 (1), 95–112.

Rotter, J.B., 1967. A new scale for the measurement of interpersonal trust. Journal of Personality 35, 651–665.

Sheridan, T.B., 1988. Trustworthiness of command and control systems. Proceedings of IFAC Man–Machine Systems, Oulu, Finland, pp. 427–431.

Singh, I.L., Molloy, R., Parasuraman, R., 1993. Automation-induced complacency: development of the complacency-potential rating scale. The International Journal of Aviation Psychology 3 (2), 111–122.