Selecting Methods for the Analysis of Reliance on Automation

Lu Wang
University of Virginia
Charlottesville, VA

Greg A. Jamieson
University of Toronto
Toronto, ON

Justin G. Hollands
Defence R & D Canada
Toronto, ON

Human reliance on imperfect automation has been the subject of many laboratory experiments. Across these studies, a diverse set of indices of reliance have been used, even in studies with similar experiment settings. This inconsistency of reliance measures makes the meta-analysis of experimental findings difficult. Moreover, few researchers have rigorously defined the optimal level of automation reliance in their settings, making it difficult to judge the appropriateness of that reliance. This paper attempts to guide researchers in selecting and interpreting measures of automation reliance behavior. We review the reliance analysis methods in existing research and propose four criteria for selecting among them. It is recommended that, where possible, future studies define optimal reliance to make unequivocal judgment about the appropriateness of reliance. In addition, more reliable and insightful conclusions can be obtained through the use of multiple measures.

INTRODUCTION

Much of the human-automation interaction literature rests on the assumption that automation is not error-free as a result of technology limitations. Consequently, users are often assigned the task of monitoring, verifying, and responding to automation feedback. For instance, fire detectors are often triggered by smoke from cooking instead of a real fire. In response to the alarm, rather than automatically activating sprinklers, the design assumes that users will open the windows and doors to let the smoke escape. This response is appropriate as long as there is no real fire. However, sometimes users ignore a correct alarm and fail to respond appropriately. In the tragic Seton Hall University residence fire of 2000, three students died after neglecting to heed alarms (Analla, 2001).

Both experimental and field studies have demonstrated that human operators often do not use imperfect automation effectively – either over-relying or under-utilizing it (Bainbridge, 1983; Parasuraman & Riley, 1997). In order to help operators better utilize imperfect automation, researchers have studied the factors that affect human reliance on automation, such as mental workload, self-confidence, and trust in automation (Lee & See, 2004). Two difficulties encountered in these studies are: 1) effectively measuring reliance behavior, and 2) defining optimal reliance for a particular setting (Moray & Inagaki, 2000). Previous studies have used a variety of distinctive methods to analyze automation reliance. Different reliance indices were used even in studies with similar settings (e.g., Parasuraman, Molloy, & Singh, 1993; Moray & Inagaki, 2000). This methodological divergence reflects an absence of guidance in the literature for choosing the most effective reliance analysis methods. This paper attempts to solve this critical methodological issue by reviewing the reliance analysis methods in existing research and proposing criteria for selecting the most informative analysis methods in future studies.

REVIEW

We begin with a review of the reliance indicators in the prevailing literature. It is useful to organize these measures into four distinct perspectives: users' general consistency with the automation, their performance of the automated task, their behavior pattern, and their decision criterion given certain automation feedback.

Consistency Indicators

When users are relying on automation, the automation feedback should influence their decision making to some extent. Hence, the consistency between automation feedback and the users' final decisions could reflect their general reliance on automation. A straightforward approach is to use the percentage of times that users accept automation feedback as a consistency indication (Biros, Daly, &

10.1518/107118108X350582

Gunsch, 2004). An alternate consistency indication is the correlation between automation feedback and the users' decision (Bisantz & Pritchett, 2003; Murrell, 1977). This approach is consistent with Brunswik's lens model (1956), which suggests that if a judgment task is based on multiple cues, the reliance on each cue could be indicated by the correlation between that cue and the final judgment.

Performance Indicators

Automation, when relied upon appropriately, is expected to help users better perform the automated task. Therefore, the performance of the automated task has been used to indicate users' reliance or inappropriate reliance on automation (e.g., Parasuraman, et al., 1993). For example, reliance has been measured by comparing users' performance when they receive correct feedback with their performance when they receive incorrect feedback (Maltz & Shinar, 2003). The more users rely on an automated system, the more likely their performance will improve when automation feedback is correct and degrade when automation feedback is incorrect. Some studies measure users' performance using their sensitivity (d') based on signal detection theory (Maltz & Shinar, 2003; Green & Swets, 1966), others use the percentage of correct responses (Dzindolet, Pierce, Pomranky, Peterson, & Beck, 2001; Galster, Bolia, Roe & Parasuraman, 2001).

Behavior Indicators

Users' reliance on automation may also be revealed in certain behavior patterns. For example, in a cued target acquisition task, Maltz and Shinar (2003) checked whether participants scanned the cued targets prior to the non-cued targets or not. Other common behavior indicators include attention allocation, verification of automation feedback, and return to manual performance (e.g., Ezer, Fisk, & Rogers, 2007; Moray & Inagaki, 2000). The attention allocation could be measured directly using users' sampling rate (e.g., Bagheri & Jamieson, 2004), or indirectly from concurrent task performance based on the notion that the concurrent task competes for a limited cognitive resource with the automated task (Manzey, Bahner & Hueper, 2006). In general, the more users rely on automation, the less attention they would pay to the automated task, the less often they would try to verify the automation feedback, and the poorer they would perform without the automation. Depending on the experimental setting, the optimal reliance can sometimes be defined by determining the optimal sampling rate for each monitoring task (Moray & Inagaki, 2000).

Response Bias Indicators

Fire detectors and individual combat identification systems exemplify one class of automated systems that provides binary cues. These binary cues can be characterized as 'signal' and 'noise' in a general sense. For example, a fire detector only has two states, alerting (i.e., 'signal') or silent (i.e., noise).

For automation that provides binary cues, users' response bias lnß can be used as an indication of users' reliance (Maltz & Meyer, 2001; Meyer, 2001). Proponents of this method claim that users' expectation of signal probability is likely to change when they receive automation feedback. For example, when users think the fire alarm is reliable, their expectation of a real fire event will increase after hearing the alarm. According to the signal detection theory, this change of signal expectation should influence the setting of the users' response bias but not their sensitivity (Wickens & Hollands, 2000). Therefore, users' reliance on 'signal' and 'noise' feedback could be measured using users' response bias in situations that automation responds 'signal' and 'noise', respectively. Furthermore, the appropriateness of reliance can be described by the match between users' response bias and the optimal response bias. Equation 1 defines the optimal response bias lnß given 'signal' feedback. The first part of the equation is determined by the noise to signal ratio given the 'signal' feedback (i.e., the accuracy of the 'signal' feedback). The second part depends on the decision payoffs.

$$\ln \beta_{optimal} = \ln \left[\frac{P(Noise \mid Signal_{feedback})}{P(Signal \mid Signal_{feedback})} \times U \right]$$

…… Equation 1

An alternative to using the value of users' response bias is to use the shift of response bias between two situations to measure reliance (Maltz & Shinar, 2003; Murrell, 1977). For instance, the general reliance on an automated system could be indicated by the shift of response bias between the situations in which users receive 'signal' and 'noise' feedback. Similarly, the reliance on 'signal' feedback could be indicated by the shift of response bias from the manual condition to the condition in which users receive 'signal' feedback. Accordingly, the optimal reliance can be defined using the discrepancy between the values of the optimal response bias in two situations. Equation 2 expresses the optimal shift of response bias when users receive 'signal' feedback:

$$\ln\beta_{opti\_manual} - \ln\beta_{opti\_signal\_feedback}$$
$$= \ln[\frac{P(Noise)}{P(Signal)} \times U_1] - \ln[\frac{P(Noise \mid Signal_{feedback})}{P(Signal \mid Signal_{feedback})} \times U_2]$$

…… Equation 2

The response bias indicators could also be used to measure the reliance on automation with multiple discrete or continuous feedback states. For example, a common task for operators in a control room is to read meters and detect abnormal situations. The continuous meter feedback can be artificially divided into multiple categories by picking up several cutoff points (see Figure 1). For a certain range defined by the adjacent cutoff points, the operators' overall reliance on the feedback could be measured by their response bias when receiving feedback within that range. Correspondingly, the optimal reliance of feedback in a certain range could be defined using the percentage of times that an abnormal event happens (i.e., hit rate) or does not happen (i.e, false alarm rate) when the automation feedback is within that range.

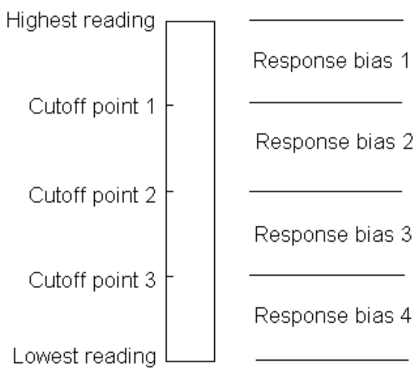

Figure 1. Response bias indicators for automation with continuous feedback

SELECTING RELIANCE MEASURES

Table 1 summarizes the reliance indicators reviewed thus far. With so many methods available to analyze the reliance on automation, how should a researcher choose appropriate methods for a specific study? Here we propose four criteria to consider when making this decision.

Table 1. Summary of different reliance indicators

Indicator Category	Indicators/Measures
Consistency	1. Percentage of opportunities in which users follow automation feedback 2. Correlation between users' decision and automation feedback
Performance	Difference between performance (i.e., sensitivity, error rate) of the automated task when receiving correct feedback and when receiving misleading feedback
Behavior	1. Sequence of visual scan 2. Attention allocation (e.g., visual sampling rate, concurrent task performance) 3. Actions taken to verify automation feedback 4. Return to manual performance
Response Bias	1. Response bias when receiving certain automation feedback 2. Response bias shift between two situations

The Definition of Optimal Reliance

The automation reliance literature places inadequate emphasis on defining optimal reliance on automation despite its importance in judging the appropriateness of reliance behavior. The consequences of this neglect have not entirely escaped human factors scientists. "The reason that current research does not unequivocally support the presence of complacency is that none of the research known has rigorously defined optimal behaviour in supervisory monitoring" (Moray & Inagaki, 2000, p. 365).

Researchers should exercise caution when drawing conclusions from reliance analysis measures that do not afford a benchmark. For example, in one study a performance indicator showed that participants made more errors when the aid was incorrect than when it was correct (Dzindolet et al., 2001). The intuition after reading this analysis result might be that the participants over-relied on the aid. However, without considering the relative accuracy of the automation and participants' manual performance, this result may not be sufficient to support this conclusion. Take an extreme case as an example. If an aid is highly reliable, say the error rate is 1%; meanwhile participants' manual performance is inferior, say the participants' error rate is 50%. If participants rely completely on the aid, the error rate when the aid is incorrect (100%) will be much larger than the error rate when the aid is correct (0%). However, complete reliance is obviously appropriate in this case. This example illustrates that it is preferable to use reliance indicators that allow for optimal reliance definitions. However, few empirical studies have established the optimal reliance using the optimal sampling rate or response bias (e.g., Moray & Inagaki, 2000; Meyer, 2001).

Specificity of Reliance

Similar to the definition of specificity of trust (Lee & See, 2004), specificity of reliance concerns whether an evaluation emphasizes users' general reliance on the whole automated system or specific reliance on one type of automation feedback. It is expected that more specific reliance analysis would yield greater insight into users' reliance behavior. Take the automation with binary cues for example, there are two reasons justifying the preference for separately analyzing reliance on each type of feedback. First, experimental evidence shows that users' reliance on 'signal' and 'noise' feedback are qualitatively different, as they are influenced by automation's false alarm rate and miss rate, respectively (Dixon & Wickens, 2006). Investigators assessing the general reliance on automation run the risk of neglecting the independent changes of reliance on each type of feedback. Second, it is difficult to define the optimal reliance on the automation as a whole because the optimal reliance on the 'signal' feedback and 'noise' feedback can be different. For example, in the case of some combat identification systems, the 'friend' feedback always correctly identifies friendly soldiers, but the 'unknown' feedback does not guarantee that a target is hostile. To study reliance on combat identification systems, it is necessary to discriminate reliance on the 'friend' feedback from reliance on the 'unknown' feedback.

All four categories of reliance indicators are capable of indicating the general reliance on the whole system. However, consistency indicators are generally not suitable for indicating reliance on specific feedback. Instead, reliance on specific feedback can be measured with performance indicators, behavior indicators, and response bias indicators (e.g., Mosier, Skitka, Heers, & Burdick, 1998).

Compatibility with Experiment Design

A specific experimental design may make some reliance analysis methods more or less feasible. Before selecting a method, it is important to consider whether a reliance indicator makes sense in the experimental setting. For example, an attention allocation method may not be effective when there is no concurrent task in addition to the automated task.

In particular, the response bias indicators seem to impose stricter requirements on experiment design. In addition to perceived signal probability, response bias is also influenced by perceived decision payoffs (Wickens & Hollands, 2000). Therefore, when using response bias to indicate reliance on automation feedback, it is important to keep the decision payoffs consistent across experimental conditions. Moreover, since the shift in response bias relative to changes in the decision payoffs is likely to be less than the ideal magnitude (Wickens & Hollands, 2000), the prescribed decision payoffs are better set to be neutral (i.e., U=1). Otherwise, the deviation from the optimal response bias might be attributed to the suboptimal response to the decision payoffs instead of inappropriate reliance on the automation. When researchers are not confident about the appropriateness of participants' responses to the decision payoff structures, the optimal shift of response bias should be used as an indication of appropriate reliance. An examination of Equation 2 shows that, even if users' responses to decision payoffs are not optimal, this influence will be eliminated from the reliance measure as long as it is consistent across experiment conditions. That is, when U_1 equals U_2, both terms can be eliminated from the definition of optimal reliance shift.

Cost

Finally, the choice of reliance analysis method should also include the cost. Although sampling rates can shed a great deal of light on users' reliance on automation, they often require expensive equipment, demand time–consuming data processing, and introduce complexities in defining optimal reliance (Bagheri, 2004). Similarly, the response bias shift and the return to manual performance indices might require comparisons to experimental conditions that might otherwise be excluded from a study.

DISCUSSION

This paper presents a review of four analysis methods that have been used in previous studies to measure automation reliance. In addition, it also identifies four criteria that could guide researchers in selecting the most informative reliance analysis methods for their studies.

The definition of optimal reliance has been neglected in most existing research. Future studies should place more emphasis on defining optimal reliance, since it will lead to more informative and less biased interpretation of the results. Moreover, it is often preferable to examine reliance in higher resolution. More thorough results could be obtained by measuring reliance on specific feedback rather than the whole system. Finally, researchers should also consider whether a reliance index is cost effective and meaningful in a particular experiment design. It is wise to design experiments that afford multiple reliance analysis methods. This is because different indicators consider the reliance behavior from different perspectives. Analysis results that take advantages of multiple methods are expected to generate more insight and offer greater reliability over single methods. Several recent studies serve as excellent examples of the value of this approach (e.g., Maltz & Shinar, 2003; Manzey et al., 2006).

REFERENCES

Analla, T. (2001). A fire alarm lesson from Seton Hall: SJSU officials pushing safety after dorm blaze here. *Spartan Daily*. Retrieved May 22, 2008, from http://media.www.thespartan daily.com/media/storage/paper852/news/2001/05/15/Campus News/A.Fire.Alarm.Lesson.From.Seton.Hall-1501630.shtml.

Bagheri, N., & Jamieson, G. A. (2004). The impact of context-related reliability on automation failure detection and scanning behaviour. In *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics* (pp. 212-217).

Bagheri. N. (2004). *The effect of automation reliability on operator monitoring*. Unpublished master's thesis, University of Toronto, Toronto, Ontario, Canada.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779.

Biros, D. P., Daly, M., & Gunsch, G. (2004). The influence of task load and automation trust on deception detection. *Group Decision and Negotiation, 13*(2), 173-189.

Bisantz, A. M., & Pritchett, A. R. (2003). Measuring judgment in complex, dynamic environments: A lens model analysis of collision detection behavior. *Human Factors, 45,* 266–280.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2^{nd} ed.). Berkeley: University of California Press.

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors, 48*(3), 474-486.

Dzindolet, M. T., Pierce, L., Pomranky, R., Peterson, S., & Beck, H. (2001). Automation reliance on a combat identification system. In *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 532-536). Minneapolis/St. Paul, MN: Human Factors and Ergonomics Society.

Ezer, N., Fisk, A. D., & Rogers, W. A. (2007). Reliance on automation as a function of expectation of reliability, cost of verification, and age. In *Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society* (pp. 6-10). Baltimore, MD: Human Factors and Ergonomics Society.

Galster, S. M., Bolia, R. S., Roe, M. M., & Parasuraman, R. (2001). Effects of automated cueing on decision implementation in a visual search task. In *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 321–325). Santa Monica, CA: Human Factors and Ergonomics Society.

Green D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics.* New York Wiley.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80.

Maltz, M., & Meyer, J. (2001). Use of warnings in an attentionally demanding detection task. *Human Factors, 43*(2), 217-226.

Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human factors, 45*(2), 281(15)-296.

Manzey, D., Bahner, E., & Hueper, A. (2006). Misuse of automated aids in process control: Complacency, automation bias and possible training interventions. In *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 220-224).San Francisco, CA: Human Factors and Ergonomics Society.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human factors, 43*(4), 563-572.

Moray, N., T. Inagaki (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science, 1*(4), 354-365.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International journal of aviation psychology, 8*(1), 47-63.

Murrell, G. A. (1977). Combination of evidence in a probabilistic visual search and detection task. *Organizational Behavior and Human Performance, 18,* 3-18.

Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology, 3*, 1-23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.