

## OCCASIONAL AUTOMATION FAILURES ON EASY TASKS UNDERMINES TRUST IN AUTOMATION

Poornima Madhavan, Douglas A. Wiegmann, and Frank C. Lacson

Aviation Human Factors Division, Institute of Aviation,

University of Illinois at Urbana-Champaign,

Savoy, IL 61874.

Automation users often mistrust diagnostic aids that are imperfectly reliable (Parasuraman & Riley, 1997), particularly on occasions when automation errors appear *conspicuous* (c.f. Dzindolet, Peterson, Pomranky, Pierce, & Beck, in press), or *simple* to the user. Madhavan, Wiegmann, and Lacson, (2003) found that when operators catch easy mistakes made by an automated aid, it bolsters their self-confidence in their own ability to outperform the aid even on difficult tasks, resulting in automation under-trust and under-reliance. Specifically, Madhavan et al. proposed the “easy-errors hypothesis” that aids that miss targets *exclusively* on “easy” tasks are more likely to be mistrusted than aids that generate misses only on “difficult” tasks.

However, such is rarely the case in the real world where systems are designed such that the probability of the system generating an error on an easy task is lower than that of generating an error on a difficult task. Easy automation errors are typically interspersed among more difficult errors. Therefore, the main purpose of this study was to examine the effects of such *occasional* easy errors on automation trust and reliance.

In addition to misses, research has shown that automated aids that produce a large number of false-alarms create under-trust in automation (Parasuraman & Riley, 1997;

Gupta, Bisantz, & Singh, 2001). While false-alarms have been shown to affect operator trust more than misses on difficult tasks (Cotte', Meyer, & Coughlin, 2001; Gupta, et al., 2001; Maltz & Shinar, 2003), the effect of *easy* false-alarms on trust relative to *easy* misses remains unexplored. Therefore, the second objective of this study was to examine whether *easy* false-alarms might affect user trust and reliance in a manner akin to *easy* misses.

### Method

Sixty students from the University of Illinois completed 200 trials of a computerized task where they detected the presence of a target "X" among an array of alphanumeric characters. Out of the 200 trials, 50% were "target trials" where the "X" was present and 50% were "noise trials" wherein the "X" was absent. Out of these, 50% were "easy" and 50% were "difficult" trials.

Three groups performed the task with a diagnostic aid that had a hit/correct rejection rate of .70, and a miss/false alarm rate of .30. The control group performed the task unaided. The reliability of the aid on each type of trial varied as shown in Figure 1. Participants rated their confidence in each diagnosis on a scale ranging from 1 (no confidence) to 5 (very confident). Feedback was given as to whether each diagnosis was correct. Aided participants estimated their trust and perceived reliability of the aid on a post-experimental questionnaire.

### Results

Trust in the aid was lower for both easy-errors groups compared to the difficult-errors group ( $M = 5.2$ ,  $SD = 1.9$ ), (easy-miss: ( $M = 3.27$ ,  $SD = 2.12$ ),  $d = .96$ ; easy-false-alarm: ( $M = 3.6$ ,  $SD = 2.29$ ),  $d = .79$ ). Trust did not differ between the easy-errors

groups. Perceived aid reliability for the easy-miss group ( $M = 51.8\%$ ,  $SD = 28.63$ ) was significantly lower than for the easy-false-alarm, ( $M = 69.67\%$ ,  $SD = 15.86$ ),  $d = .77$ , and difficult-errors groups, ( $M = 75.67\%$ ,  $SD = 11.32$ ),  $d = 1.1$ .

Agreement rates revealed that the difficult-errors group never disagreed with the aid on *easy trials* given their aid never made errors on easy trials. Likewise, the easy-miss group never disagreed with the aid on easy-noise trials, and the easy-false-alarm group never disagreed with the aid on easy-target trials. The easy-miss and easy-false-alarm groups agreed with the aid on 80% of the easy-target and easy-noise trials respectively, as their aids were 80% accurate on these trials. The easy-false-alarm group ( $M = 66\%$ ,  $SD = .21$ ) agreed with the aid less than the easy-miss group, ( $M = 79\%$ ,  $SD = .17$ ),  $d = .72$ , on *difficult-target trials*. The easy-false-alarm group also complied with the aid less, ( $M = 56\%$ ,  $SD = .11$ ), than the easy-miss group ( $M = 61\%$ ,  $SD = .008$ ,  $d = .64$ , on *difficult-noise trials*. Overall, the easy-false-alarm group had fewer hits on difficult-target trials and fewer false-alarms on difficult-noise trials relative to the easy-miss group.

Confidence on *easy trials* was high and did not differ across groups. Confidence of the easy-miss group, ( $M = 3.57$ ,  $SD = .43$ ) on *difficult-target trials* was significantly higher than that of the difficult-errors, ( $M = 3.22$ ,  $SD = .61$ ),  $d = .68$ , and control groups, ( $M = 3.27$ ,  $SD = .45$ ),  $d = .71$ . Likewise, the easy-false-alarm group, ( $M = 3.12$ ,  $SD = .66$ ), was significantly more confident than the difficult-errors, ( $M = 2.54$ ,  $SD = .74$ ),  $d = .85$ , and control groups, ( $M = 2.64$ ,  $SD = .59$ ),  $d = .80$  on *difficult-noise-trials*.

### Conclusions

Results revealed that compared to participants utilizing automation that made only difficult errors, participants using aids that generated easy errors trusted the aid less,

underestimated its reliability, and were highly self-confident regardless of the accuracy of their diagnoses. These findings corroborate the “easy-errors hypothesis” (Madhavan, et al., 2003) that easy errors generated by a diagnostic aid undermine automation trust and reliance. Furthermore, the present study extends the findings of Madhavan, et al., by demonstrating that the undermining effect of easy errors on trust is salient even when the easy errors (a) are occasional, and (b) comprise easy false alarms. However, the undermining effect of false alarms was not significantly different from that of misses suggesting that the “simplicity” of automation errors appears to have a stronger influence on trust and reliance than the type of error generated.

The easy-errors groups were significantly more confident than the other groups, though this inflated confidence was not justified by better performance. Such overconfidence could result from participants overestimating their ability to detect errors on difficult trials when the aid committed obvious errors on easy trials. In the real world, such “easy” misses or “easy” false-alarms may occur when the algorithms used by automation to discriminate between noise and signal is insufficient to capture all instances of a particular target. Additional research is needed to examine the generalizability of the present findings to real world complex systems.

## References

- Cotte', N., Meyer, J., & Coughlin, J. F. (2001). Older and younger drivers' reliance on collision warning systems. In *Proceedings of the Human Factors and Ergonomics Society 45<sup>th</sup> Annual Meeting* (pp. 277-280). Santa Monica, CA: Human Factors Society.
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (in press). The role of trust in automation reliance. *Human Factors*.
- Gupta, N., Bisantz, A. M., & Singh, T. (2001). Investigation of factors affecting driver performance using adverse condition warning systems. In *Proceedings of the Human Factors and Ergonomics Society 45<sup>th</sup> Annual Meeting* (pp. 1699-1703). Santa Monica, CA: Human Factors Society.
- Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2003). Automation failures on tasks easily performed by operators undermines trust in automated aids. In *Proceedings of the 47<sup>th</sup> Annusl Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA: Human Factors Soceity.
- Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human Factors*, 45 (2), 281-295.
- Parasuraman, R., & Riley, V. (1997). Humans and automation use: Use, misuse, disuse, abuse. *Human Factors*, 39 (2), 230-253.

	Noise Trials (n = 100)		Target trials (n = 100)	
	Easy Noise (n = 50)	Diff Noise (n = 50)	Diff Targets (n = 50)	Easy Targets (n = 50)
<b>Difficult Only</b> (n = 15)	p (CR) = 1.00 (0 False Alarms)	p (CR) = .40 (30 False Alarms)	p (H) = .40 (30 Misses)	p (H) = 1.00 (0 Misses)
<b>Easy Miss</b> (n = 15)	p (CR) = 1.00 (0 False Alarms)	p (CR) = .40 (30 False Alarms)	p (H) = .60 (20 Misses)	p (H) = .80 (10 Misses)
<b>Easy FA</b> (n = 15)	p (CR) = .80 (10 False Alarms)	p (CR) = .60 (20 False Alarms)	p (H) = .40 (30 Misses)	p (H) = 1.00 (0 Misses)
<b>Control</b> (n = 15)				

} p (H) = .70  
 p(CR) = .70  
 60 Errors  
 70%  
 Reliable

Figure 1. Graphical Representation of Experimental Groups