On the Independence of Compliance and Reliance: Are Automation False Alarms Worse Than Misses?

Stephen R. Dixon, Christopher D. Wickens, and **Jason S. McCarley,** University of Illinois, Savoy, Illinois

Objective: Participants performed a tracking task and system monitoring task while aided by diagnostic automation. The goal of the study was to examine operator compliance and reliance as affected by automation failures and to clarify claims regarding independence of these two constructs. **Background:** Background data revealed a trend toward nonindependence of the compliance-reliance constructs. **Method:** Thirty-two undergraduate students performed the simulation that presented the visual display while dependent measures were collected. **Results:** False alarm–prone automation hurt overall performance more than miss-prone automation. False alarm–prone automation also clearly affected both operator compliance and reliance, whereas miss-prone automation appeared to affect only operator reliance. **Conclusion:** Compliance and reliance do not appear to be entirely independent of each other. **Application:** False alarms appear to be more damaging to overall performance than misses, and designers must take the compliance-reliance constructs into consideration.

INTRODUCTION

For the past few decades, designers have attempted to reduce operator workload levels by introducing automated aids that assist or replace human functions. One example is the diagnostic aid that alerts operators to potential problems in the environment. The current study examines the effects of diagnostic aids on human performance in a supervisory control situation.

Imperfect Automation

It is sometimes tempting to assume that automation is a panacea for reducing workload levels and improving performance, but not all forms of automation are perfectly reliable. This is particularly true of diagnostic automation, which must make diagnoses and long-range predictions in a world of imperfect probabilistic information (Wickens & Dixon, 2007). Diagnostic automation designed to detect a state of the world can produce two forms of errors: false alarms and misses. A diagnostic aid's performance can be measured using signal detection theory (Green & Swets, 1988). The automation can have a liberal criterion (or beta) or a conservative criterion – that is, it can either commit more false alarms or more misses, respectively, or it can be neutral.

Because beta is typically set at the discretion of the system designer, it is critical for designers to understand the effects of different forms of automation errors before implementing a new automated system. This includes understanding the effects of the drivers of optimal beta – that is, how the probability of a signal, the costs of error, and the payoffs of correct responses interact with the operators' determination of where to set their criterion. The goal of the current study was to provide data that allow system designers to better make these judgments during the design process.

Compliance and Reliance

Recent theory has postulated that automation false alarms and misses have qualitatively different effects on operator dependence (Meyer, 2001, 2004). Compliance is what the operator typically does when the automation diagnoses a signal in the world, whereas reliance is what the operator does when the automation diagnoses noise in the world. An increase in false alarms is posited to reduce compliance, resulting in longer response times to automation alerts. In extreme cases, this

Address correspondence to Stephen R. Dixon, Department of Psychology, New Mexico State University, Las Cruces, NM 88003; sdixon@nmsu.edu. *HUMAN FACTORS,* Vol. 49, No. 4, August 2007, pp. 564–572. DOI 10.1518/001872007X215656.

results in a tendency to disregard those alerts entirely – the "cry wolf" effect (Dixon & Wickens, 2006; Wickens, Dixon, Goh, & Hammer, 2005). An increase in the automation's miss rate reduces reliance, causing the operator to allocate more attention to monitoring the raw data behind the automation in order to catch the possible automation misses. This diverts attentional resources from any concurrent task, causing a deterioration in performance in that task (Dixon & Wickens, 2006; Wickens & Colcombe, in press; Wickens, Dixon, Goh, et al., 2005).

Previously, it had been implied that compliance and reliance may be separate constructs (Meyer, 2004) and even possibly independent ones – that is, an increase in false alarms should affect only compliance and an increase in misses should affect only reliance, and indeed this could be seen to be an optimal cognitive response.

Dixon and Wickens (2006) reported data partially consistent with this assertion. They had pilots fly a simulated unmanned aerial vehicle (UAV) mission, which consisted of tracking the UAV through a series of way points while searching for targets of opportunity along the way. Concurrently, the pilots were responsible for monitoring a set of four system gauges for possible system failures, aided by the implementation of a diagnostic aid that sounded an auditory alert when it determined (correctly or incorrectly) that a system failure had occurred. Importantly, the system gauges were highly distinguishable in nature (green and red zones) and likely required very few cognitive resources to analyze the raw data behind the alarms. The data revealed that increasing the automation miss rate affected only measures of reliance, whereas increasing the automation false alarm rate appeared to affect both compliance and reliance. However, because of the low power and sensitivity of the concurrent task measures, the authors were unable to find strong statistical evidence of the nonindependence of the constructs.

Wickens, Dixon, Goh, et al. (2005) replicated a portion of the previous experiment and added eye-tracking data to their analyses. They found behavioral data consistent with the previous findings, and their measures of visual scanning provided further evidence for the possible nonindependence of reliance and compliance. Specifically, the investigators noted that high false alarm rates induced a significant shift of attention toward the raw data in the alerted domain relative to perfect automation and, therefore, away from concurrent tasks. This shift should have been solely a symptom of a high automation miss rate according to the independence model. However, the same issues of low statistical power prevented the authors from making strong claims of nonindependence.

Wickens, Dixon, and Johnson (2005) repeated a similar version of the UAV paradigm, but instead they provided an unreliable diagnostic aid to the more difficult target search task as well as a perfectly reliable aid to the system gauge task. They found that the disruptive effects of automation false alarms on the concurrent task were at least as strong as those of the automation misses, yet another finding inconsistent with the reliance-compliance independence model. Similar effects have been found recently in a multiple-UAV paradigm (Levinthal & Wickens, 2005).

Collectively, these studies suggest that false alarm–prone automation may be at least as much if not more disruptive of multitask performance than is miss-prone automation (Bliss, 2003) because of the former's effect on concurrent task performance as well as automated task performance. However, this conclusion is based on only marginally significant performance trends with a low power measure (Dixon & Wickens, 2006). Furthermore, some of these findings may be a function of the perceived cost of false alarms and misses on total human-machine output.

The Current Study

The current study addresses the gaps in knowledge from the previously described experiments by providing a continuous and sensitive measure of concurrent task performance, hence allowing greater statistical power and experimental control. Furthermore, in the present automated task, participants encountered more than 35 examples of one class of automation error (misses or false alarms), whereas in the prior studies the disparity was far less (typically 2 or 3 of one class). Given these changes, the current study provides a stronger opportunity to examine the relationship between reliance and compliance in a way that was not available in the previous studies. To the extent that reliance and compliance are independent constructs influenced by automation misses and false alarms (FAs), respectively, then performance in the concurrent tracking task during nonalert periods should be equivalent between a perfectly reliable automation condition and an FA-prone condition because the operator should not be monitoring the systems gauge if there is no automated alert (i.e., reliance should be unaffected).

The second important addition to this study is that it continues to extend the conclusions beyond the relatively simple perceptual monitoring task used in the early UAV studies to one with highly demanding cognitive elements, a manipulation also subsequently done by Levinthal and Wickens (2005) and Wickens, Dixon, and Johnson (2005).

The current study involved two concurrent tasks: a continuous compensatory tracking task and a cognitively demanding systems monitoring task. In the latter, participants were required to calculate current values and report when the needle of the system gauge exceeded a certain acceptable range. Some participants performed the task unaided, and others performed the task with the aid of an automated diagnostic system that was perfectly reliable, FA prone, or miss prone.

We posited six hypotheses: (a) Perfect automation would benefit both the tracking task and systems monitoring task. (b) The system prone to automation misses would harm the tracking task because of a reduction in operator reliance, causing a shift of attention away from the tracking task in order to catch the automation misses in the systems monitoring task. (c) An increase in the automation miss rate should not affect measures of operator compliance (e.g., speed of response to an alert). (d) The system prone to automation FAs would harm the system monitoring task because of a reduction in operator compliance. (e) Automation FAs would also harm the tracking task, even when the alarm is silent, because of a reduction in operator reliance. (f) As a consequence of the previous hypotheses, automation FAs would be more harmful to overall performance than would automation misses.

METHOD

Participants

Thirty-two undergraduate students from the University of Illinois participated in the experiment and were paid $9/hr, plus performance bonuses. Participants were made aware of the incentives and were told that the two tasks should receive equal priority.

Apparatus and Stimuli

The experimental simulation ran on a Dell GX270 computer, with a 21-inch (53-cm) Dell monitor using 1280 × 1024 resolution. Figure 1 presents a sample display for the experiment.

The experimental display was subdivided into two areas of interest separated by approximately 12° of visual angle (center to center). In the top portion of the display was the two-dimensional tracking task. Participants used a joystick with first-order dynamics to track the target disturbed

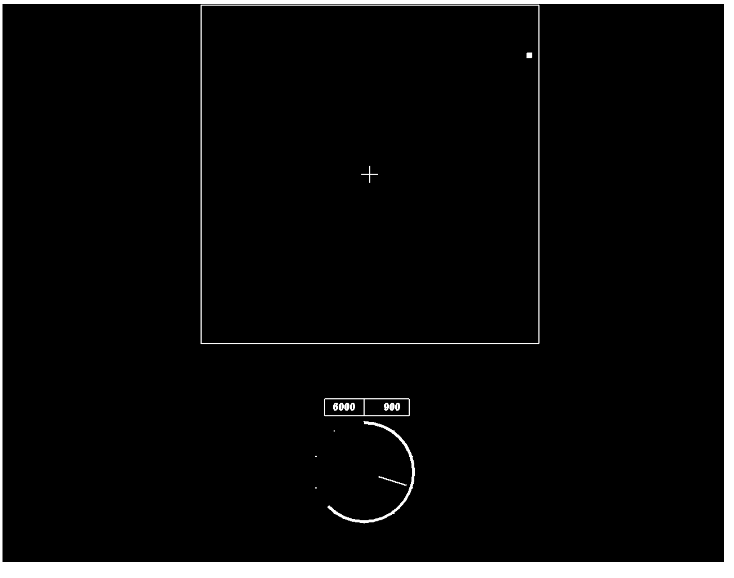

Figure 1. Sample screen shot of experimental display.

by a quasi-random input with a bandwidth of 0.43 Hz. A negative drift in error was added so that if the participant did not exert appropriate force on the joystick, the ball would quickly float toward the outer edges of the box.

The system monitoring gauge was located in the bottom display and represented the value of a generic real-world variable (e.g., altitude). The gauge had 10 small white ticks spaced equidistantly around the outside of an imaginary circle. A yellow bar that "filled" the outside of the gauge denoted units of 1,000, 2,000, and so forth. A yellow needle that rotated around the inside of the gauge denoted units of 100, 200, et cetera. Thus, the value of the sample gauge depicted in Figure 1 is approximately 5975. The yellow needle was driven by the sum of four sine waves ranging in bandwidth from 0.04 to 0.43 Hz. The yellow bar "filled" the gauge in a linear fashion, according to whatever the current value was, as dictated by the random movements of the needle.

Above the gauge were two white boxes with white numerical values. The number in the left box denoted the ideal value for a "safe" system. The number in the right box denoted the range of "safety" for the system. Thus, for the example shown in Figure 1, the participant had to keep the gauge within 5500 ± 800. If the gauge went out of this range, it denoted a system failure (SF). If an SF occurred, the participant was expected to press a button on the keyboard as quickly as possible. When an SF occurred, the needle stayed out of the acceptable value range until it was detected or the trial ended. The SF task was purposely designed to be a challenging task that required both visual and cognitive (working memory) resources. In a real-world application, the system gauge might be considered poorly designed in that it is difficult to read; however, it was designed for this experiment as much for its theoretical value as for its practical value and in recognition of the fact that many real-world gauges indeed are poorly designed.

For some participants, performance on the systems monitoring task was aided by automation. The automated aid sounded an auditory alert (i.e., a synthesized human voice pronouncing the word "one") when an SF occurred. The automation, expressed in the framework of signal detection theory, could provide hits (alarm with true SF), misses (no alarm with true SF), FAs (alarm with no SF), or correct rejections (no alarm with no SF).

Trials

There were 100 trials that each lasted exactly 30 s. At the beginning of each trial, the target value (in the left numeric box above the SF gauge) changed to a new random value between 1,000 and 9,000, rounded to the nearest 100. The target range (in right numeric box above the SF gauge) changed to a new random value between 100 and 900, rounded to the nearest 100. The SF gauge itself reset to the target value and then immediately began oscillating.

An SF occurred on 50 trials, with SF and non-SF trials randomly ordered. SFs (and automation FAs) always occurred within a temporal window beginning 5 s and ending 12 s from the start of the 30-s trial interval, thus giving the participant at least 18 s to detect the failure. There was never more than one SF or alert from the automation per trial. Trials lasted the entire 30 s, regardless of whether or not an SF occurred or was detected. During each trial, the participant was allowed to make only one SF response (e.g., if they responded "yes" before an SF actually occurred, then it would be classified as an operator FA) and was not allowed to retract a response once executed. Once the 30-s trial ended, participants were no longer able to respond to that particular trial. At the end of each trial, the screen flashed either green or red to inform participants whether their response was correct or incorrect, respectively.

Procedure and Design

After filling out a consent form and reading the instructions, each participant completed 20 practice trials followed by 80 experimental trials. There were four experimental conditions: (a) baseline condition (no automated aid); (b) A100 condition (40 hits, 0 FAs, 0 misses, 40 correct rejections [CRs]); (c) FA60 condition (40 hits, 32 FAs, 0 misses, 8 CRs); and (d) M60 condition (8 hits, 0 FAs, 32 misses, 40 CRs). Participants were told that the automation would be either perfectly reliable or "not perfectly reliable" and, in the latter case, which way the automation criterion would be set. Thus, it can be assumed that the participants were immediately aware of the potential for automation failures as well as which type of failure they would encounter.

RESULTS

One participant in the M60 condition was dropped because of unusually poor performance levels (beyond the third standard deviation below the group mean) in the tracking task. For the most part, analysis entailed a one-way omnibus ANOVA followed by three planned comparisons: (a) baseline versus A100, (b) baseline versus the combination of FA60 and M60 in a planned comparison (i.e., weights of –1, 0.5, 0.5, respectively), and (c) FA60 versus M60. Because only three a priori comparisons were made, familywise error rates were not adjusted. Any post hoc tests used a Bonferroni correction.

Tracking Error

Tracking error was calculated only during the period between the beginning of a trial and the onset of either an SF or an automation FA, as this was the period of time in which variations in attentional reliance caused by the different conditions were expected. These data represented between 5 and 12 s of time at the beginning of each trial. Figure 2 presents these data as a function of condition, using the solid black bars.

A one-way ANOVA revealed a reliable main effect of condition, $F(3, 27) = 6.64$, $p < .01$. Planned comparisons showed that perfect automation ($M = 98$) may have improved performance over baseline ($M = 131$) at a level approaching significance, $t(14) = 1.53$, $p = .07$. The baseline condition showed better performance relative to the average of the two unreliable conditions, $t(14) = 2.55$, $p = .01$, whereas the difference between the FA60 ($M = 243$) and M60 conditions ($M = 182$) was not statistically significant, $t(13) = 1.32$, $p > .10$.

Effect of FAs on Reliance

Because tracking error was measured during the time before an SF or alert occurred, we could assume that any performance deficits in the tracking task for the FA60 condition indicated a reduction in operator reliance. That is, such deficits would demonstrate that the operator was putting attentional resources into the SF task even when there was no alert, causing an increase in tracking error. A separate analysis done only on trials in which there was no SF and the automation was silent revealed the same results. These data are shown with the striped bars of Figure 2. A one-way ANOVA on these data revealed a main effect of condition, $F(3, 27) = 5.09$, $p < .01$, and a post hoc comparison between the FA60 and the A100 conditions, $t(14) = 3.78$, $p < .01$, provides clear evidence that operator reliance was reduced.

SF Detection Rate

There were no significant differences found in operator beta across conditions, $F(3, 27) = 2.3$, $p > .10$. For all other detection rate analyses, the signal detection measure d' was used. Perfect scores (e.g., zero misses or FAs) were adjusted by assuming ½ of a miss or FA.

A one-way ANOVA revealed a main effect of condition, $F(3, 27) = 8.84$, $p < .001$. Planned comparisons revealed no significant difference between the baseline condition ($M = 3.03$) and A100 condition ($M = 3.20$), $t(14) < 1.0$. The baseline condition produced performance better than the average of the two unreliable conditions, $t(14) = 2.43$, $p = .01$. The FA60 condition ($M = 2.04$) produced

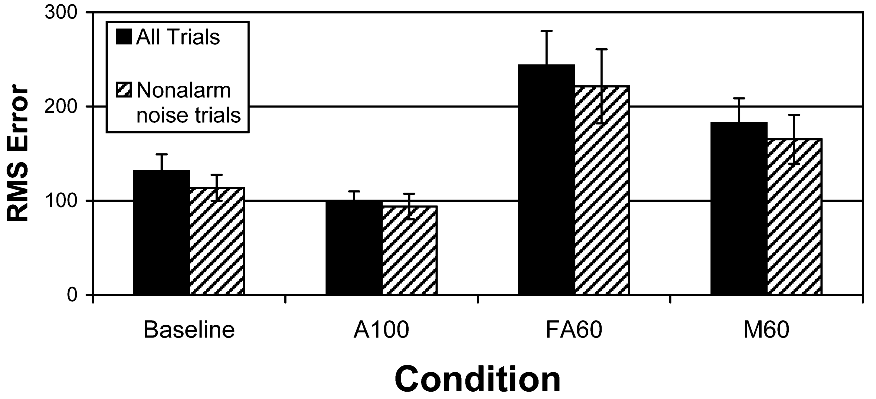

Figure 2. Tracking error as a function of condition. The solid black bars represent all trials in the experiment, and the striped bars represent only nonalarm noise trials. *SE* bars are included. RMS = root mean square.

performance worse than that of the M60 condition ($M = 2.61$), $t(13) = 3.08$, $p < .01$. Post hoc tests revealed that the baseline condition was performed better relative to the FA60 condition, $t(14) = 3.15$, $p < .01$, but did not differ significantly from the M60 condition, $t(13) = 1.38$, $p > .10$.

Further analysis was done in the automation conditions to determine the likelihood of a yes/no operator response based on the type of automation response, helping to shed light on operator agreement or disagreement with the automation. Table 1 presents these data. When there was a signal, all groups tended to agree with automation but did so less with FA-prone automation ($M = .93$) than with miss-prone automation ($M = 1.00$), $t(14) = 3.75$, $p < .01$. In contrast, when the automation was silent, the operator was less likely to agree in the miss-prone condition ($M = .82$) than in the FA-prone condition ($M = .92$), $t(13) = 2.14$, $p < .05$. These findings are consistent with the postulation that FA-prone automation reduces compliance but that miss-prone automation reduces reliance.

SF Response Times

SF response times are presented in Figure 3. The solid black bars represent all trials in the experiment, and the striped bars represent only true alarm signal trials. Note that the baseline condition did not have any true alarms, but the corresponding trials were included as a measure of what "baseline" results were.

A one-way ANOVA on the data for all trials revealed a main effect of condition, $F(3, 27) = 9.85$, $p < .001$. Planned comparisons revealed that participants in the baseline condition ($M = 0.82$ s) performed more poorly than those in the A100 condition ($M = 0.42$ s), $t(14) = 2.26$, $p < .05$, but better than the average of the two unreliable automation conditions, $t(14) = 2.48$, $p = .01$. The FA60 condition ($M = 1.91$) was performed much more poorly than the M60 condition ($M = 0.88$), $t(13) = 2.71$, $p < .01$. Post hoc tests revealed that performance in the baseline condition was better than in the FA60 condition, $t(14) = 3.35$, $p < .001$, but was not significantly different from that in the M60 condition, $t(13) < 1.0$.

Effect of Misses on Compliance

The following analysis was done to determine if automation misses had any effect on operator compliance. Applying the logic used to examine accuracy, if compliance in the M60 condition was perfect, then on trials in which the automation alert sounded, the response times to the SFs should have been equivalent to those in the A100 condition – that is, the operators would have known that the automation did not commit FAs and that when it sounded, it was always correct. To test the effects of compliance, we analyzed data for trials on which an alarm occurred separately.

A one-way ANOVA on these SF response times revealed a main effect of condition, $F(3, 27) = 7.54$, $p < .01$. Planned comparisons between the M60 condition ($M = 0.37$) and A100 condition ($M = 0.41$), $t(13) < 1.0$, revealed that the misses in the M60 condition did not appear to affect operator compliance at all and that performance in the M60 condition following a true alarm was better than performance in the baseline condition ($M = 0.91$), $t(13) = 2.10$, $p < .05$. As expected, the FA60 condition ($M = 1.69$) did degrade operator compliance, $t(14) = 3.53$, $p < .01$.

TABLE 1: Operator Response as a Function of Automation Accuracy

	State of the World			
	Signal		Noise	
Condition	Hit	Miss	FA	CR
FA60	.93	—	.35	.92
M60	1.00	.05	—	.82
A100	.96	—	—	.93

Note. Operator agreement rates are shown. The first and third columns tend to be measures of compliance, whereas the second and fourth columns tend to be measures of reliance. FA = false alarm, CR = correct rejections.

DISCUSSION

Previous studies have indicated that automation FAs and automation misses have qualitatively different effects on operator performance (Meyer, 2004; Dixon & Wickens, 2006) – that is, automation FAs tend to adversely affect operator compliance, whereas automation misses tend to adversely affect operator reliance. The current study was able to provide a stronger opportunity to examine the degree of independence of the reliance-compliance constructs implied by Meyer (2001, 2004) and expanded upon by Dixon and Wickens (2006).

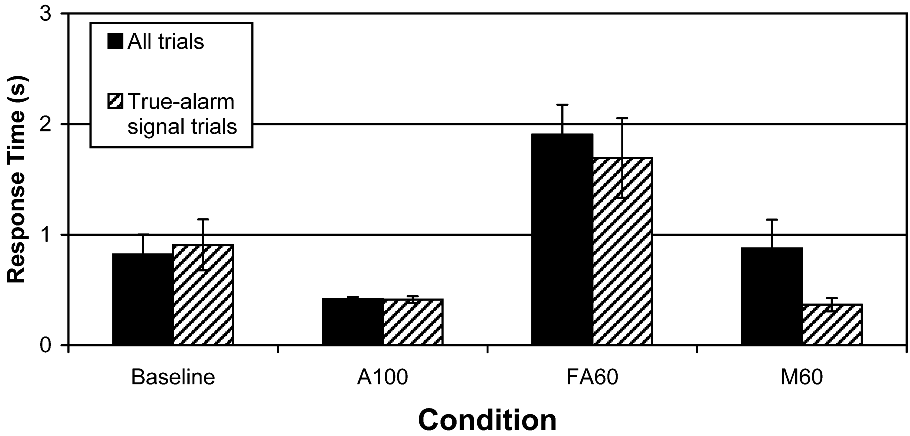

Figure 3. System failure (SF) response times as a function of condition. The solid black bars represent all trials in the experiment, and the striped bars represent only true alarm signal trials. *SE* bars are included.

The current study replicated the finding that perfect automation is beneficial to overall human-automation performance, as predicted in Hypothesis a. However, this benefit was not as profound in the tracking task, in which unaided performance was nearly as good as perfect automation. This may have been attributable to increased effort on the part of the participants and/or to the participants' ability to divide attention between the two tasks more effectively than we expected. Perhaps if we had made the systems monitoring task even more difficult than it was, that task would have become more of an attention sink, resulting in poorer tracking task performance.

Hypothesis b predicted that the miss-prone automation would harm the tracking task by causing operators to shift attention away from the tracking task in order to catch the potential automation misses. This proved to be correct, as the performance in the miss-prone condition suffered relative to that in the perfectly reliable condition, matching previous findings by Dixon and Wickens (2006). Consistent with Hypothesis c, the data showed that automation misses had no significant effect on operator compliance.

Hypothesis d predicted that the FA-prone automation would damage the systems monitoring task by reducing operator compliance. The data agreed strongly with this hypothesis, as both the SF detection rates and response times suffered relative to those in the perfectly reliable automation condition and even dropped far below baseline performance. Although the data in Table 1 show that operators were inclined to agree with the automation when it correctly detected an SF, the increased response times suggest that this agreement occurred was only after the participant double-checked the raw data. When the automation presented an FA, operators incorrectly agreed only one third of the time. These two factors indicate low operator compliance.

Importantly, FA-prone automation also adversely affected operator reliance, as predicted by Hypothesis e, confirming what Wickens, Dixon, Goh, et al. (2005) and Dixon and Wickens (2006) suggested based on trends seen in their data. When the automation was silent, operators in the FA condition should have completely ignored the systems monitoring gauge and focused their entire attention on the tracking task. Instead, the data revealed that the tracking task performance in the FA condition was not only worse than that in the reliable automation condition but was also worse than that in the baseline condition. This implies that the reliance-compliance constructs may not be entirely independent of each other.

Thus, our Hypothesis f – that the FA-prone condition would be more harmful to overall performance relative to the miss-prone condition – proved to be correct both qualitatively and quantitatively. First, the FA-prone automation adversely affected both operator compliance and reliance, whereas the miss-prone automation appeared to reduce only operator reliance. Second, FA-prone automation hurt performance more on the automated task than did miss-prone automation, (e.g., the "cry wolf" effect) and hurt performance at least as much as miss-prone automation on the concurrent task. The current data provide convincing evidence that automation FAs not only produce effects on automation dependence that are qualitatively different from those produced by automation misses but also are quantitatively more harmful to performance than are misses.

The strong effect of FA-prone automation on reliance – as measured by tracking error prior to the occurrence of an alert – can be explained only by assuming that attention was diverted from the tracking task to monitor the raw data in the system failure task. The data thus imply that the high FA rate reduced operators' trust in and dependence on the alerting system in general (i.e., not just by reducing compliance) and did so as much as did the miss-prone system. Thus the logical statement that the two constructs, reliance and compliance, should be independently affected by miss rate and false alarm rate is not reflected in the participants' monitoring strategies.

Although the current data do not speak to the issue directly, one possible reason for the relatively broad effects of false alarms on operator dependence is that false alarms, which were each accompanied by a salient perceptual event (an auditory cue), are simply more noticeable or memorable errors than are automation misses (Madhavan, Wiegmann, & Lacson, 2006). Further research will be necessary to test this speculation.

In addition to the previous explanation, confirming Hypothesis e on the basis of the current data, one can also speculate how two additional nonreliance-related FA factors can disrupt concurrent task performance in the interval after the alert sounds. First, the greater number of true alerts for the FA60 (N = 40) than for the M60 (N = 8) condition will cause a much greater frequency of attention diversion away from tracking in order to respond to the system failures. Second, even the false alerts (N = 32 here), because they require participants to check the raw data of the monitoring task in order to determine whether or not a system error has truly occurred, will cause a diversion of attention away from tracking.

Note that if such diversion in the second case does not occur, and the false alerts are totally ignored, then a corresponding cost will be evident in the compliance cost of the FA-prone system. Although postalert tracking error was not measured here, we remain relatively certain that such interference would have been shown, given the known limits of time-sharing between separated visual tasks.

Thus it is clear that FA-prone automation affects total performance by degrading reliance (affecting prealert concurrent task performance), by time-sharing disruption (affecting postalert concurrent task performance), and by reduced compliance (affecting alerting task performance), whereas miss-prone automation degrades only the first of these (reliance), thereby supporting Hypothesis f.

More research needs to be done to determine whether these three explanations, which are not necessarily in opposition to one another, are viable explanations for why automation FAs appear to affect operator reliance. Further research also needs to be done to determine whether the two types of automation errors affect different cognitive processes, or whether a single-process model is sufficient in explaining the data. The current data suggest that the two constructs are not entirely independent of each other. Although this may imply that a single-process model is more likely to explain the data, this single dissociation in the data is not sufficient to reject a multiple-process model. A multiple-process model does not require the assumption of selective influence (Dunn & Kirsner, 1988). Indeed, it is possible to have nonselective effects on operator dependence (e.g., automation FAs affecting operator reliance) and still confirm a multiple-process model. One effective method of determining the viability of a multiple-process model is to perform a state-trace analysis on the data; however, the current data are not sensitive enough to warrant this analysis.

The current study expands on the findings from previous studies and provides a more sensitive analysis of the qualitative differences between automation FAs and misses, thereby allowing generalization of its implications beyond the specific UAV paradigm used here. Subsequently, the current data allow designers of automated systems to more accurately weigh the impact of automation FAs and misses on operator performance when deciding where to set the bias threshold in future systems.

ACKNOWLEDGMENTS

This research was sponsored by ARMY MicroAnalysis and Design 6021.000-01. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the U.S. Army. The authors wish to acknowledge the support of Ron Carbonari (in developing the simulation) as well as the helpful comments of three anonymous reviewers.

REFERENCES

Bliss, J. (2003). An investigation of alarm related accidents and incidents in aviation. *International Journal of Aviation Psychology, 13,* 249–268.

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,* 474–486.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95*, 91–101.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics.* New York: Wiley.

Levinthal, B., & Wickens, C. D. (2005). *Supervising two versus four UAVs with imperfect automation: A simulation experiment* (AHFD-05-24/MAAD-05-7). Savoy, IL: University of Illinois, Aviation Human Factors Division.

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors, 48,* 241–256.

Maltz, M., & Shinar, D. (2003). New alternative methods in analyzing human behavior in cued target acquisition. *Human Factors, 45,* 281–295.

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

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46,* 196–204.

Spence, C., & Driver, J. (1997). Audiovisual links in attention: Implications for interface design. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics, Vol. 2, Job design and product design* (pp. 185–192). Hampshire, UK: Ashgate Publishing.

Wickens, C. D., & Colcombe, A. (in press). Performance consequences of imperfect alerting associated with cockpit display of traffic information. *Human Factors.*

Wickens, C. D., & Dixon, S. R. (2007). Is there a magic number 7 (to the minus 1)? The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8*(3), 201–212.

Wickens, C. D., Dixon, S. R., Goh, J., & Hammer, B. (2005). Pilot dependence on imperfect diagnostic automation in simulated UAV flights: An attentional visual scanning analysis. Paper presented at the *13th Annual International Symposium of Aviation Psychology* held April 18–21 in Oklahoma City, Oklahoma.

Wickens, C. D., Dixon, S. R., & Johnson, N. R. (2005). *UAV automation: Influence of task priorities and automation imperfection in a difficult surveillance task* (AHFD-05-20/MAAD-05-6). Savoy, IL: University of Illinois, Aviation Human Factors Division.

Stephen R. Dixon is an assistant professor of psychology at New Mexico State University. He received his Ph.D. in psychology from the University of Illinois in 2006.

Christopher D. Wickens is a senior scientist at Alion Science Corporation, Micro Analysis & Design Operations, Boulder, Colorado, and professor emeritus at the University of Illinois at Urbana-Champaign. He received his Ph.D. in psychology from the University of Michigan in 1974.

Jason S. McCarley is an assistant professor of psychology at the University of Illinois. He received his Ph.D. in psychology from the University of Louisville in 1997.

Date received: November 21, 2005
Date accepted: June 26, 2006