MANAGEMENT OF MULTIPLE UAVS WITH IMPERFECT AUTOMATION

Brian R. Levinthal & Christopher D. Wickens
University of Illinois at Urbana Champaign
Aviation Human Factors Division
Savoy, Illinois

Forty-two participants manipulated two or four unmanned aerial vehicles (UAVs) while monitoring an adjacent display for camouflaged tanks. They were sometimes supported in the tank-detection task by an automated target recognition system that operated at either a reliability of .9 with a threshold designed to provide an equal number of false-alarms and misses, .6 with a low threshold producing more automation false alarms, or .6 with a high threshold producing more automation misses. As taskload doubled, performance on the UAV task was significantly reduced. The effect was not mediated by the presence of automated aids, though the aids did influence performance in the tank-detection task. Tank detection was improved by both the highly-reliable aid and (to a lesser extent) by the miss-prone aid, but was not improved (and sometimes hurt) by the false-alarm prone condition. The results support the independence model of reliance and compliance proposed by Meyer (2001).

INTRODUCTION

Unmanned aerial vehicles (UAVs) have recently drawn attention because of their ability to improve pilot safety while maintaining many of the mission-critical capabilities of a manned flight. Manipulation of UAVs can be costly, however: the Shadow and Hunter, two UAVs operated by the army, require several operators to perform a single mission. Therefore, there has been substantial interest in reducing the demands associated with a single UAV, with the goal of allowing a single operator to manage multiple vehicles. While there is some evidence to support the feasibility of multi-UAV control (Cummings and Guerlain, 2004; Ruff et al., 2002), there remains some concern that simultaneous management of multiple vehicles might, in general, constitute an unacceptable degree of mental workload. This workload can be mitigated by the presence of an automated decision aid (Dixon, Wickens, and Chang, 2005), but because of the uncertainty inherent to the UAVs' airspace, such aids are likely to be imperfect.

Imperfect automated decision aids are susceptible to committing two forms of error, false alarms and misses, whose relative prevalence is determined by the aid's alert threshold (i.e. signal detection beta). Meyer (2001, 2004) has argued that this tradeoff between false alarms and misses produces two potentially independent cognitive states: compliance and reliance. Compliance reflects the operators' tendency to agree with an automated aid when it provides an alert, and is reduced when the aid commits false alarms (which lead to the 'cry wolf' effect; Bliss, 2003). Reliance refers to the operators' assumption that a system is functioning normally while the alert is silent, and is reduced when the aid commits misses. High reliance leads to complacency (Parasuraman et al., 2003), while low reliance reduces the operators' 'spare capacity' for performing concurrent tasks (Dixon and Wickens, in press). In Meyer's (2001) original conception of compliance and reliance, the two states were portrayed as independent. Recent evidence, however, has not supported this claim (Wickens, Dixon, Goh, and Hammer, 2005; Wickens, Dixon, and Johnson, 2005), and instead indicates that under some circumstances, a high false-alarm rate can be shown to drive both reliance and compliance. Contrary to the independence hypothesis proposed by Meyer, these results support an alternate 'False-Alarms Hurt' model.

The following experiment examines the effects of an automated decision aid's level of reliability and alert threshold in the context of multiple-UAV control using a highly realistic UAV simulator. Pilots were responsible for the simultaneous operation of multiple vehicles while performing a vigilance task that was sometimes aided by automation. As far as is known, this is the first study to have jointly examined the effects of UAV number and unreliable automation.

We hypothesized 1) that overall performance would decrease when the number of UAVs under participants' control increased, and 2) that the "FA-Hurts" model would most accurately characterize the relationship between reliance and compliance. That is, miss-prone automation would impose costs only on the concurrent task, while FA-prone automation would impact both the automated task (through the "cry wolf" phenomenon), and the concurrent task, because of the disruptive interruptions of false alarms.

METHODS

Forty-two students, ages 19 – 29 from the University of Illinois at Urbana-Champaign participated in a two-hour study, and were compensated approximately $10 per hour for their time.

A SIL (Systems Integration Lab) UAV simulator (from Micro-Analysis and Design, Colorado) was used for this experiment. Figure 1, below, shows the arrangement of two displays for this experiment, which provided raw data and controls relevant to the pilots' two tasks. For the UAV-management task, pilots were presented with an interface designed to replicate those currently in use by the Army. Pilots interacted with the controls via touch-screen monitor.

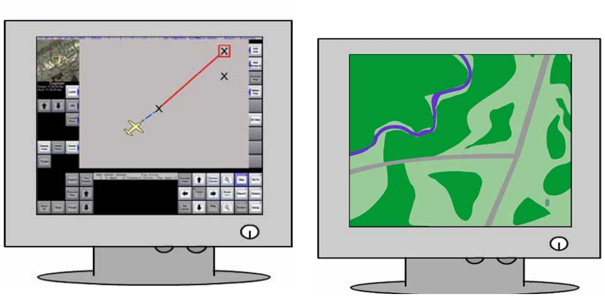

Figure 1. An example of the setup for this experiment.

Pilots navigated each UAV through a series of branching waypoints. UAV-flight between waypoints was automated, but pilots were required to perform a decision rule upon reaching each waypoint in order to select the next correct branch. To choose the correct waypoint, pilots subtracted the last two digits of the X and Y coordinates of the UAV, and selected the northern-most waypoint when the difference was greater than 50 (otherwise, they chose the southern-most waypoint). Thus, the demands associated with UAV flight were predominantly cognitive, and not physical. When UAVs reached a waypoint, they circled indefinitely in an 'idle' mode until the pilot selected the next waypoint. The accumulation of 'idle' time served as a dependent measure of UAV-task performance. In two experimental sessions, pilots were responsible for either two UAVs (low-workload) or four UAVs (high-workload). UAVs reached a waypoint (i.e. pilots were required to perform the decision rule) once every 15 seconds during low workload sessions, and every 7.5 seconds during high workload sessions.

While performing the UAV task, pilots monitored an adjacent display of raw terrain data with the goal of finding enemy tanks as quickly and accurately as possible (pilots responded by saying 'TANK' and pointing to the target). The terrain scrolled continuously across the screen at a rate of one complete refresh every 30 seconds, thus tanks, when present, were on-screen for a full 30 seconds. There were 20 tanks, which appeared pseudo-randomly within a range of 30 seconds and five minutes.

In some cases, pilots were provided with an automated target recognition aid to assist in the tank-detection task. The aid functioned at one of four levels (pilots experienced only a single level of automation): a non-automated baseline (BL), a 90% reliable aid that committed equal false alarms and misses (A90), and a 60% reliable aid with a 3:1 likelihood of committing either false-alarms over misses (FAP), or misses over false-alarms (MP). Correct alerts sounded no earlier than three seconds after the tank appeared, and no later than seven seconds.

Four dependent variables were recorded to measure compliance and reliance in the context of multiple UAV control. UAV-task performance was measured via accumulated 'idle' time. Compliance was measured by the pilots' response times following an automated alert. Reliance was measured by response times when the automated aid committed a rare miss, and was also measured by the proportion of responses made after the sounding of an alert. The latter measure provides an indirect measure of reliance, as pilots who frequently viewed the terrain monitor would be more likely to detect a tank within the 3 – 7 second pre-alert period.

RESULTS

UAV Performance

Figure 2 shows the performance costs associated with increasing workload. As workload increased from low to high, idle time increased reliably, $F(1,72) = 248.3$, $p < .001$. Doubling objective taskload resulted in a tripled cost to task efficiency: accumulated idle time increased from approximately 600 seconds to 1800 seconds. None of the automation conditions demonstrated effects on UAV performance that approached significance.

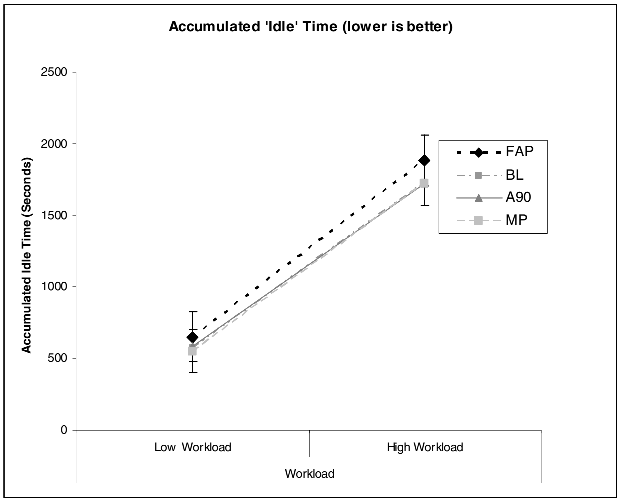

Figure 2: Accumulated 'idle' time, a measure of UAV performance. High workload resulted in a threefold increase in 'idle' time.

Tank Detection Accuracy

There was no effect of automation condition on the accuracy of tank detection, which was essentially perfect across all conditions. Instead, variability in tank detection performance was demonstrated in response times.

Measure of Compliance

Figure 3 shows a significant main effect of automation condition on post-alert response time (our primary measure of compliance), $F(2,56) = 27.7$, $p < .001$, indicating that the FAP condition yielded longer response times than either the 90% reliable condition or the MP condition, reflecting the 'cry wolf' effect. Though response times were longer for the 90% reliable condition than for the MP condition under high workload, this interaction did not approach significance.

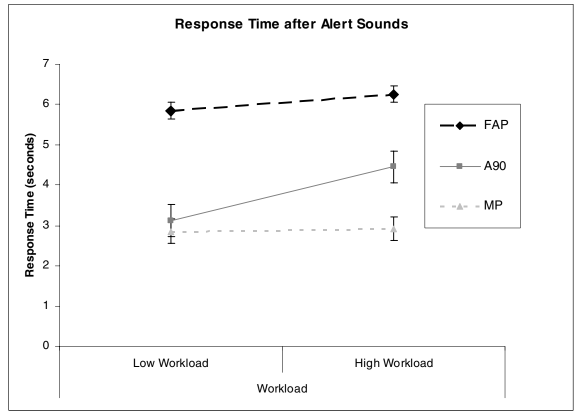

Figure 3: Compliance, as measured by the response time following an alert. Increased response-time indicates lower compliance

Tank Detection Reliance Measures

There was no significant main effect of automation condition on response times when the alert was silent, though there was a non-significant (and anticipated) trend indicating that pilots in the MP-condition were faster than those in the FAP-condition. Figure 4 shows the second measure of reliance, the proportion of responses made after the sounding of an alert. For this measure, there was a significant main effect of workload, $F(1,56) = 14.3$, $p < .001$, such that as workload increased, the proportion of responses made after an alert increased as well. There was a significant main effect of automation condition, $F(1,56) = 8.63$, $p < .002$, indicating that pilots in the FAP and 90% reliable conditions responded after an alert (i.e. were more reliant) more frequently than pilots in the MP condition.

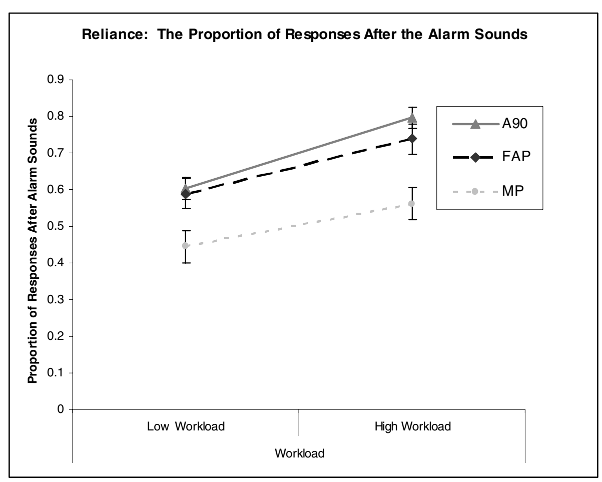

Figure 4: The proportion of responses made after the alarm sounds. A higher proportion (e.g. A90 and FAP conditions) indicates lower reliance.

DISCUSSION

While pilots were able to manage multiple UAVs, there was a substantial cost to performance as the number of vehicles under their control increased. This result follows from hypothesis 1), in which we expected workload to reduce overall operator performance. It appears that even when control of multiple UAVs requires minimal physical demands, the cost of increasing mental workload can be disastrous.

A second goal of this project was to examine the viability of the FA-hurts model vs. the independence hypothesis proposed by Meyer (2001). While previous studies (Wickens, Dixon, Goh, and Hammer, 2005; Wickens, Dixon, and Johnson, 2005) have shown that FAP aids are associated with reductions in measures of both reliance and compliance, this was not borne out in our paradigm. Contrary to hypothesis 2), our results favored the independence hypothesis: the MP condition was associated with an increase in our measure of compliance and a decrease in the measures of reliance, while the opposite was true of the FAP condition.

While none of the automated conditions appeared to affect performance in the concurrent UAV task (as would have been expected given our understanding of reliance and compliance), this may be explained by qualitative differences in UAV-management demand in our study versus others. In both Wickens, Dixon, Goh, and Hammer (2005), and Wickens, Dixon, and Johnson (2005), successful UAV flight required continuous physically-demanding manual control. Therefore, reallocation of attention away from the concurrent task necessarily invoked a cost to performance (i.e. reduced spare capacity). In our paradigm, control of each UAV was discrete and allowed pilots the opportunity to strategically allocate attention between the automated and concurrent tasks. Since optimal task switching is likely to be demanding, pilots might have chosen this strategy only when it seemed necessary (i.e. the MP condition).

We can draw several practical implications from our experiment. First, considering our finding that FAP aids were associated with delayed response times compared to both the MP or 90% reliable aids, it appears that FAP aids are more disruptive to performance than MP aids. Second, the independence hypothesis appears to garner support when task switching costs are high. Unless there is sufficient motivation to shift attention away from a concurrent task, as in the high-threshold MP aid condition, pilots may be willing to exhibit overly-high reliance on imperfect aids in order to reduce overall workload. Finally, while multi-UAV management may be practical in the future, our present results should provide a warning: even when concurrent tasks were supported by highly reliable automation, pilots were unable to avoid performance penalties while managing several vehicles. For multi-UAV management to become a reality, it may be necessary to reduce the demands of each individual vehicle and to support the strategic allocation of attention amongst several tasks.

REFERENCES

Bliss, J. and Acton, S. A. (2003) Alarm mistrust in automobiles: How collision alarm reliability affects driving, *Applied Ergonomics*¸ 34 (6), 499-509.

Cummings, M.L., and Guerlain, S. (2004) Human Performance Issues in Supervisory Control of Autonomous Airborne Vehicles. AUVSI's Unmanned Systems North America 2004 Symposium, August 2004.

Dixon, S. R., & Wickens, C. D. (in press). Automation reliability in unmanned aerial vehicle flight control: Evaluating a reliance-compliance model of automation dependence in high workload. *Human Factors*.

Dixon, S. R., Wickens, C. D., & Chang, D. (2005). Mission control of unmanned air vehicles: A workload analysis. *Human Factors, 47.*

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.

Parasuraman, R. M., Molloy, R., & Singh, I. L. (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*(2), 230-253.

Wickens, C. D., Dixon, S., Goh, J., & Hammer, B. (2005). Pilot dependence on imperfect diagnostic automation in simulated UAV flights: An attentional visual scanning analysis. *Proceedings of the 13th International Symposium on Aviation Psychology*. Dayton, OH.

Wickens, C. D., Dixon, S., & Johnson, N. (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.