RECOVERY FROM INTERRUPTIONS TO A DYNAMIC MONITORING TASK: THE BEGUILING UTILITY OF INSTANT REPLAY

Mark St. John, Harvey S. Smallman, & Daniel I. Manes
Pacific Science & Engineering Group, San Diego, CA

Detecting changes in complex monitoring tasks is important for situation awareness, yet surprisingly difficult. Interruptions exacerbate this problem. An intuitively appealing solution to this problem is Instant Replay. Users could replay interrupted periods at high speed to quickly perceive changes. Instant Replay's appeal seems to rest on its familiarity and realistic re-presentation of the temporal sequence of the interrupted situation. However, current theories of perception, including Naïve Realism (Smallman & St. John, 2005), predict this emphasis on realism to be misguided. We compared two versions of replay against three alternative tools in a naval air warfare simulation in which 35 participants monitored a busy airspace for significant changes. One alternative, CHEX, a situation awareness recovery concept we are developing, automatically detects and logs changes into an interactive table. CHEX provided an effective representation for quickly recovering situation awareness. In contrast, realistic Instant Replay proved worse than no support at all.

Many tasks, such as civil emergency operations, air traffic control, and military operations, require consistent and careful monitoring of complex situations and detecting of significant changes. For example, in air warfare, naval operators must maintain awareness of aircraft within an airspace, constantly evaluating their potential threat to friendly assets, and taking precautionary actions to keep their assets safe. Unfortunately, there are several ways in which situation awareness can be degraded and changes missed.

It has long been appreciated that visual attention acts to restrict the mental representation of scenes. For example, stress-induced "tunnel vision" narrows attention toward the center of gaze, leading to missed peripheral events (e.g., Baddeley, 1972; Williams, 1995). Moreover, Mack & Rock (1998) showed that observers performing even simple visual discriminations completely missed the presentation of salient visual features just away from fixation – they showed inattention blindness. Additionally, humans are likely to miss changes to even simple scenes if the transients that accompany the changes are masked or disrupted in any way. This change blindness is surprising severe and has been documented in many contexts (see Rensink, 2002 for a review), including air warfare (DiVita, Obermayer, Nugent, & Linville, 2004; Smallman & St. John, 2003) and is only just being recognized as a significant problem for human computer interaction (Varakin et al., 2004).

The nature of the task environment can also degrade situation awareness. Multi-tasking and other sources of interruptions can reduce awareness since changes occurring during an interruption will obviously go undetected. When users return, they may still go undetected since the telltale change transients occurred during the interruption. Users may also experience a "resumption lag" once they return to the primary task, as they work to re-acquire situation awareness, retrieve suspended task goals, and perform any pending actions (see Monsel, 2003). Resumption lag has been studied primarily in problem solving tasks in which users must recall the state and suspended goals of the primary task once they return from an interruption (e.g., Altmann & Trafton, 2002). A number of interface design concepts have been proposed to reduce this resumption lag, such as negotiating the timing of interruptions (e.g., McFarlane, 2002; Monk, Boehm-Davis, & Trafton, 2002) and cuing suspended goals (Trafton, Altmann, Brock, & Mintz, 2003).

These prospective interventions are appropriate for static task environments such as VCR programming because the environment does not change during the interruption. However, we have found that these interventions are less appropriate for dynamic task environments. In dynamic environments, the situation continues to evolve during the interruption, and unobserved changes that occur during an interruption must be inferred from the new, current state of the situation, forcing significant reassessment. These challenges have led us to focus on the period following an interruption in order to reduce the resumption lag. The solutions we propose should help overcome the ill effects of inattention blindness and change blindness during active monitoring, as well.

For these reasons, Smallman developed a Change History EXplicit (CHEX) tool for situation awareness recovery that automatically detects important changes to the situation and logs them in a sortable table. Importantly, changes in the table are linked back to the map so that selecting a change highlights the affected object on the map, and vice versa. Smallman & St. John (2003) showed that participants using CHEX identified changes much faster and more reliably than participants relying on their own powers of observation to detect changes or on less sophisticated automatic alerting aids that tended to distract and clutter the display.

Here, we further evaluate the design space of situation awareness recovery tools by comparing CHEX against an alternative tool that has much intuitive appeal: Instant Replay. Upon returning to the monitoring task, users could replay the interrupted period at high speed to quickly search for any recent changes to the situation. To increase the meaningfulness of this comparison, we chose a fairly realistic simulation of air warfare tasking in order to evaluate the tools against the conditions that occur in an actual complex and dynamic task (see Figure 1).

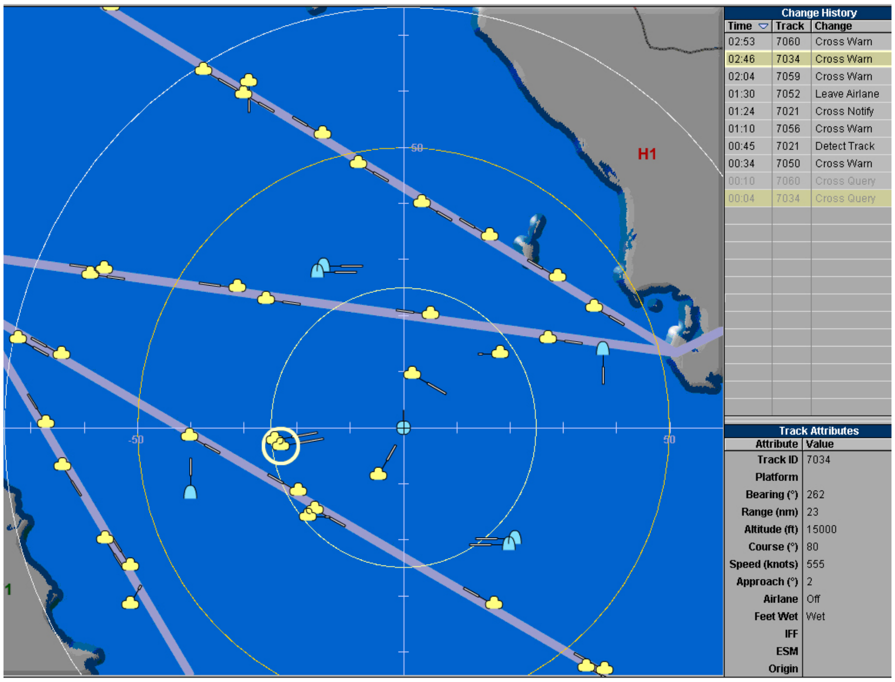

Figure 1. Screenshot showing the map, data display (lower right), and CHEX table (upper right).

The appeal of a replay tool seems to rest on two ideas: its familiarity from searching and cueing video clips, and its realism in that it involves the re-presentation of the actual sequence of events in their natural format, albeit at increased speed. It therefore involves the use of the same natural observational powers that users employ to find changes in real-time.

It has been our experience, albeit anecdotal, that this familiarity and realism beguile many users, including many Human Factors professionals, into predicting high utility for Instant Replay. Additionally, well known Human Factors principles, such as Pictorial Realism (Roscoe, 1968) and Direct Manipulation (Hutchins, Hollan, & Norman, 1986) predict value for a realistic tool like Instant Replay.

Reflection on the severity of inattention blindness and change blindness, however, suggests that the appeal of Instant Replay may be naïve. This naivety seems to rest on an inflated view of one's own observational powers. Participants have been shown to be quite unaware of their poor ability for detecting changes to a scene during a short video clip (Levin, Momen, Drivdahl, & Simons, 2000). In our previous study, too (Smallman & St. John, 2003), participants dramatically underestimated the difficulty of the change detection and identification task and the benefit that the CHEX tool could provide them. This inflated view of one's own abilities, coupled with a desire for familiarity, results in a general belief that realistic displays that support natural perceptual abilities are ideal for supporting good task performance. We term this belief Naïve Realism (Smallman & St. John, 2005).

To investigate replay and CHEX more deeply, we analyzed the change detection process into two parts, detecting which aircraft changed, and identifying the type of change. We also designed alternative versions of replay to support these different parts.

We predicted that, contrary to naïve belief, the Instant Replay tools would provide little benefit for either detecting or identifying changes due to the effects of inattention blindness and change blindness. Explicit representation of changes in the effectively organized CHEX tool, however, would again provide substantial benefits for detection and identification.

METHOD

Participants. The participants were 35 members of the general public recruited from CraigsList.com (an internet classified ads website), local universities, and Pacific Science & Engineering Group.

Task and apparatus. Participants monitored a geoplot of an airspace and reported any "significant" changes to aircraft. The display was shown on a 17" computer screen with a resolution of 1280 x 1024 pixels. The display showed a 170 x 120 nautical mile area reminiscent of the Persian Gulf, see Figure 1. Own ship appeared as a blue circle near the center of the display, violet colored airlanes criss-crossed the display, friendly aircraft were blue, and unknown, potentially threatening aircraft were yellow (MIL-STD-2525B symbology, Department of Defense, 1999). Speed and course were indicated by the length and direction of a gray line ("leader") emanating from each aircraft symbol. In addition to the map, users could access a variety of information about an aircraft, such as an identification number, range, speed, and altitude, by selecting the aircraft symbol on the map and then viewing a set of data that appeared in a window in the lower right corner of the screen. Additional displays and capabilities varied by condition (described below).

During each of three 15 minute scenarios, aircraft moved slowly about the display at realistic physical rates: the equivalent of 95 to 560 miles per hour (0.006 to 0.035 degrees of visual angle per second). There were approximately 50 aircraft on the display at all times, with aircraft occasionally entering or exiting the displayed area.

Each scenario contained an average of 36 significant changes. Fifty-two percent of the changes consisted of an aircraft crossing one of three operationally-relevant range rings at 75, 50, and 25 miles from own ship (only inbound ring crossings counted). These changes were the most obvious and easy changes to infer since the location of an aircraft just inside of a ring was good evidence that the aircraft had recently crossed the ring. Twenty-five percent of the changes were new aircraft appearing on the display, and the remaining significant changes involved sudden decreases in altitude, increases in speed, course changes from away from to toward own ship, departures from the path of an airlane, and crossing from over land to over water ("going feet wet"). Participants were instructed to ignore any changes made by friendly aircraft or aircraft flying along airlanes because these changes were not considered operationally "significant."

Participants reported significant changes by selecting the aircraft symbol on the map, then selecting the changed attribute in the data display in the lower right corner of the screen. For example, if an aircraft crossed a range ring, then the range attribute should be selected in the data display.

Three 30 second interruptions and two 120 second interruptions occurred, without warning, during each scenario. During the interruptions, aircraft continued to move and make changes. One or two changes occurred during each short interruption, and five or six changes occurred during each long interruption. When an interruption occurred, the screen was blanked, and participants were asked to rate their mental workload using the ratings portion of the NASA Task Load Index (TLX; NASA, n. d.). A warning signal sounded 10 seconds prior to the end of each interruption so that participants could prepare to resume the primary task.

There were five conditions. The Baseline condition showed the map plus the aircraft data display in the lower right corner of the screen. The map and aircraft symbols depicted the current status of all relevant aircraft attributes except altitude, and the data display presented the current status of all relevant aircraft attributes.

Table 1: Experiment Conditions

Condition	Detection Aid	Identification Aid
Baseline	Current situation	Current situation
Basic Replay	Temporal sequence	Temporal sequence
Explicit Replay	Change markers (during replay)	Temporal sequence
Explicit Markers	Change markers	Current situation
CHEX	Table entries	Table entries

The Basic Replay condition added a replay button in the lower left corner. Selecting the button caused the situation to jump backward to the beginning of the last interruption (or the beginning of the scenario, if there were no prior interruption), then replay at 20 times speed up to the current scenario time. Participants could use the replay tool whenever they wished.

The Explicit Replay condition automatically detected and marked significant changes by adding small red triangles to the aircraft symbols and a "pop" sound. The markers and pops appeared during replay at the time when the change occurred, and the markers moved with the aircraft symbols through the remainder of the replay (see Figure 2). They were removed when the replay ended. This feature was designed to minimized clutter on the map to just the period during which the user chose to invoke the replay tool. This enhanced tool supported change detection but still required users to rely on the temporal sequence to determine the identity of changes.

The Explicit Markers condition removed the replay function, but maintained the red triangles and pop sounds. Since these markers did not need to be preserved for replay, they were removed as the changes were reported. In addition to limiting clutter, this design also visually distinguished the aircraft having pending changes. This condition supported change detection, but required users to rely only on the current situation to infer previous changes.

The CHEX condition did not use the red triangles at all, again as a way of minimizing clutter. Instead, a table was added above the data display (see Figures 1 & 2). Each row of the table logged the time, aircraft identification number, and a short description of a significant change. When a change occurred, a pop sounded, and a new row was added to the top of the table. Selecting a row highlighted that row with a yellow fill and a yellow bounding box, selected the aircraft on the map with a yellow circle, and presented that aircraft's data in the data display. All other rows involving that aircraft were also highlighted in the table with the same yellow fill in order to provide additional context. This information was not relevant to the current task, but it would be valuable for actual air warfare. Reporting a change caused the row to gray out. For this experiment, the changes were sorted chronologically in order to suit the task. However, the CHEX concept allows for user control over the arrangement. For instance, the task used in Smallman and St. John (2003) was better suited by sorting changes by aircraft.

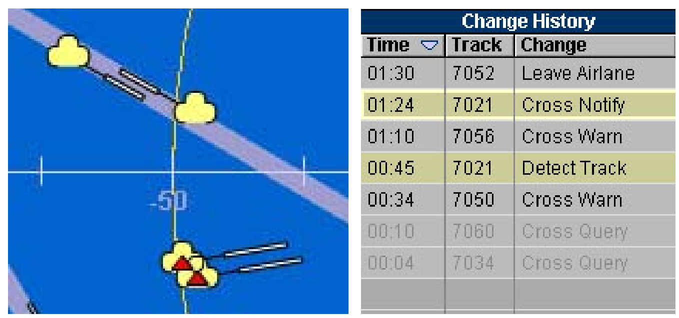

Figure 2. Close-up of triangle alerts from the Explicit Replay and Explicit Markers conditions (left), and a sample CHEX table (right). In the table, the change at time 01:24 is selected.

Procedure. Participants were randomly assigned to one of the five resumption tool conditions. They were instructed on the task, the definitions of significant changes, and their condition. They performed a five minute practice scenario, with guidance, three times. Participants then performed all three test scenarios. Finally, participants rated their mental workload one last time using the full TLX procedure.

RESULTS

Response times to correctly reported changes were computed for each scenario and then averaged to produce a time for each participant. Times were measured from the change event to the correct selection of the changed attribute, less the elapsed time of any interruptions. Therefore, only the time while a change was visible on the screen was included in the response time. Response times were submitted to a one-way between participants ANOVA of tool type (see Figure 3). The tools produced significantly different response times, $F(4, 30) = 10.5$, $p < .0001$. The CHEX tool was significantly faster than the other tools and 57% faster than Baseline (all post hoc tests by Student-Newman-Keuls, $p < .05$).

Misses were defined as significant changes that were not reported within three minutes or by the end of the scenario. Additionally, unreported ring crossings were considered missed when an aircraft crossed a closer ring. Errors were defined as incorrectly reporting an attribute that did not change. The number of misses and errors were computed for each scenario and then averaged to produce a miss and error count for each participant. These counts were then submitted to separate one-way between participants ANOVAs of tool type. The tools produced different numbers of misses, $F(4, 30) = 29.5$, $p < .0001$. CHEX and Explicit Markers produced fewer misses than the Baseline or either replay tool, and the Explicit Replay tool produced fewer misses than the Baseline or Basic Replay tool. The tools also produced different numbers of errors, $F(4, 30) = 12.1$, $p < .0001$. The CHEX tool produced fewer errors than any other tool. Explicit Markers produced more errors than any other tool (p's < .05).

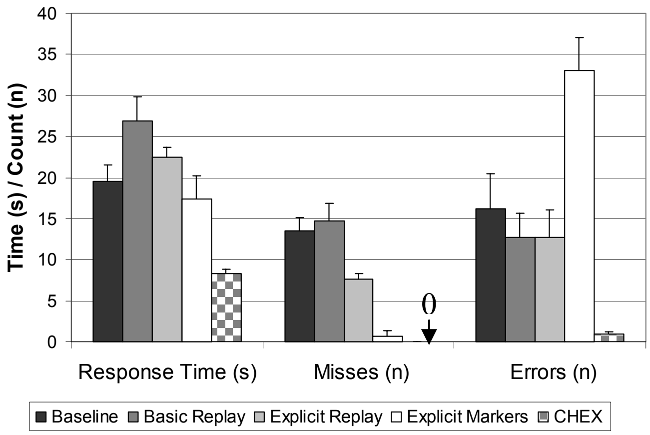

Figure 3. Response times, number of misses and number of errors per scenario for each tool.

To assess the effect of easy versus difficult to detect changes, response times were submitted to a two-way mixed effects ANOVA of tool type and change type (ring crossings vs. the more difficult to detect non-ring crossings). The tools again produced different response times, $F(4, 30) = 15.5$, $p < .0001$. The CHEX tool was better than all other tools, and the Basic Replay tool was worse than every other tool, including Baseline! The delay imposed by replaying the interrupted period actually interfered with finding and reporting changes.

Non-ring crossings were reported more slowly than ring crossings, $F(1, 30) = 60.4$, $p < .0001$, and there was a significant interaction indicating that the differences between conditions were exacerbated for non-ring crossings, $F(4, 30) = 7.0$, $p < .001$. However, for the CHEX tool, there was no effect of change type since all changes were equally explicit in the table.

We also tested to see if the interruptions produced a resumption lag. Namely, would changes that occurred during an interruption be responded to more slowly than changes occurring while the user was watching the map due to a delay in "coming back up to speed" on the monitoring task following an interruption? Response times were split according to whether they occurred during an interruption, and submitted to a two-way mixed effects ANOVA of tool type and change time (interrupted or real-time). Confirming the prediction, response times were 34% longer for changes occurring during interruptions, $F(1, 30) = 39.0$, $p < .0001$.

Even the CHEX tool showed a small resumption lag, taking 10 seconds to report each interrupted change, but only one second to report each visible change on average. The delay was most likely due to the fact that several changes needed to be reported following each interruption. The Basic Replay tool, on the other hand, took 32 seconds to report each interrupted change (vs. 22 seconds for each real-time change), and only half of the changes present were reported. These lags are much longer than those found in static tasks (e.g. Trafton et al., 2003), presumably because of the difficulty spotting changes.

Finally, an analysis of the TLX subjective workload scores indicated that participants rated the Explicit Markers and CHEX tools as lower workload than the other tools, $F(4, 30) = 7.90$, $p = .0002$.

To summarize, the CHEX tool produced faster response times and fewer errors than any other tool, and it produced fewer misses (zero) than all but the Explicit Markers tool. The CHEX tool offered a highly efficient combination of change detection and change identification. The Explicit Markers tool produced few misses, but mediocre times and a high number of errors. This pattern of data suggests that the tool offered good change detection, but participants had to guess the identity of many changes.

Instant replay, however, imposed a delay for detecting changes, and no help with identification. Further, embedding change detection markers within the replay tool (Explicit Replay) produced more misses than making the markers consistently visible (Explicit Marker).

Remarkably, given the replay tools' poor performance, there was no decrease in participants' use of either replay tool throughout the experiment. Participants apparently continued to place their faith naively in the replay tools' utility. The Basic Replay tool was used an average of 31, 34, and 31 times during each scenario, and the Explicit Replay tool was used an average of 42, 40, 46 times during each scenario.

DISCUSSION

Change awareness is a critical component of situation awareness in monitoring tasks, and a good situation awareness interface should support both the detection and identification of significant changes with minimal clutter or distraction from other, potentially more critical tasks.

The Baseline display, showing just the current situation, offered very poor support for change awareness. Rather, it produced high miss rates, high error rates, and generally slow response times: 17 seconds on average during real-time monitoring and 23 seconds on average after an interruption.

Instant Replay, despite its intuitive appeal, offered little support for detecting or identifying changes over the Baseline condition. In fact, the Basic Replay tool was worse than nothing, since it added a delay as it replayed the temporal sequence. Nonetheless, participants continued to use both replay tools throughout the experiment. It is true that the replay feels compelling, and it changes the task from having to infer changes to being able to watch them occur. However, the results indicate that this feeling belies the actual utility of the tool. In particular, the less visible non-ring crossing changes frequently remained undetected and unidentified.

Change detection is difficult enough, in fact, that automatic detection of significant events is warranted, in spite of the potential reliability and trust issues (but see St. John, Smallman, Manes, Feher, & Morrison, in press, for one method of reducing the cost of false alarms). Indeed, adding explicit markers to the changed aircraft improved detection. Embedding the markers within the replay tool, however, limited their utility. The Explicit Markers condition, though, allowed participants to become aware of the changes as they occurred, rather than when they explicitly requested a replay, and it kept the markers visible until the changes were reported.

Simple detection, however, was not sufficient. Participants often attempted to guess the identity of the changes, since the current state of the situation provided few clues, particularly toward the identity of the non-ring crossing changes.

Furthermore, marking changes directly on the map can create its own problems by distracting users from other, potentially more important tasks and by added clutter to an already busy display. There is a well known trade-off between information availability and distraction (see Smallman & St. John, 2003). Instant availability can be distracting and cluttering, but reduced availability, for example by requiring the user to press a key in order to see the markings, runs the risk of delaying important alerts. The Explicit Replay tool suffered this problem.

CHEX provides a hybrid solution to this trade-off by immediately presenting changes in the CHEX table but linking them to the map only at users' request when they select a change from the table. CHEX also offers support for both detection and identification. In terms of the Proximity Compatibility Principle (Wickens & Carswell, 1995), the CHEX table is separated from the map display, but the linking by highlighting mitigates this problem. Nonetheless, it may be useful to indicate the most important changes directly on the map while indicating less important changes only in the CHEX table.

CONCLUSIONS

Change detection, even during active monitoring, is more difficult than frequently assumed, and displays of just the current situation lead to many misses. Instant Replay, despite its beguiling intuitive appeal, offers no functional advantage. Explicit automatic detection of significant events is a superior approach. However, there remain many interface issues to resolve. CHEX provides an effective solution to the information access/clutter trade-off.

As in other domains, the emphasis of Naïve Realism on making an interface realistic without careful consideration of functionality can lead the designer astray (see Smallman & St. John, 2005).

ACKNOWLEDGEMENT

This work was supported by the Office of Naval Research.

REFERENCES

Altmann, E. M. & Trafton, J. G. (2002). Memory for goals: an activation-based model. *Cognitive Science, 26,* 39-83.

Baddeley, A. D. (1972). Selective attention and performance in dangerous environments. *British Journal of Psychology, 63,* 537-546.

Department of Defense (1999). *Department of defense, interface standard, common warfighting symbology: MIL-STD-2525B.* Washington DC: Department of Defense, Defense Information Systems Agency. Available at http://symbology.disa.mil/symbol/index.html.

DiVita, J. Obermayer, R. Nugent, W., & Linville, J. M. (2004). Verification of the change blindness phenomenon while managing critical events on a combat information display. *Human Factors, 46,* 205-218.

Hutchins, E., Hollan. J., & Norman, D. (1986). Direct Manipulation Interfaces. In D. Norman and S. Draper, (Eds.), *User Centered System Design.* Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 87-124.

Levin, D.T., Momen, N., Drivdahl, S.B., & Simons, D.J. (2000) The metacognitive error of overestimating change-detection ability. *Visual Cognition*, *7,* 397-412.

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction, 17,* 63-139.

Monk, C. A., Boehm-Davis, D A., & Trafton, J. G. (2002). The attentional costs of interrupting task performance at various stages. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting.* Santa Monica, CA: HFES.

Monsell, S. (2003). Task switching. *Trends in Cognitive Science, 7,* 134-140.

NASA (no date). *Task Load Index [TLX] Version 1.0, Users' Manual.* Available at http://iac.dtic.mil/hsiac/Products.htm#TLX

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53,* 245-277.

Roscoe, S. N. (1968). Airborne displays for flight and navigation. *Human Factors, 10,* 321-332.

Smallman, H. S. & St. John, M. (2003). CHEX (Change History EXplicit): New HCI concepts for change awareness. In *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 528-532). Santa Monica, CA: Human Factors and Ergonomics Society.

Smallman, H. S. & St. John, M. (2005). *Naïve Realism:* Misplaced faith in the utility of realistic displays. *Ergonomics in Design*, in press.

St. John, M., Smallman, H. S., Manes, D. I., Feher, B. A., & Morrison, J. G. (2005). Heuristic automation for decluttering tactical displays. *Human Factors.* In press.

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies, 58,* 583-603.

Varakin, D. A., Levin, D. T., & Fidler, R. (2004). Unseen and unaware: Implications of recent research on failures of visual awareness for human-computer interface design. *Human-Computer Interaction, 19,* 389-422.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: its psychological foundation and relevance to display design. *Human Factors, 37*, 473-494.

Williams, L. J. (1995). Peripheral target recognition and visual field narrowing in aviators and nonaviators. *International Journal of Aviation Psychology, 5,* 215-232.