Presented at the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University. 2001.

ATTENTION TO SAFETY AND THE PSYCHOLOGY OF SURPRISE

Christopher D. Wickens
University of Illinois, Aviation Human Factors Division
Savoy, Illinois

ABSTRACT

Aviation accidents are rare and often triggered by surprising atypical events, often where important data do not get noticed or attended. Yet most psychological research on attention and human response deals with and statistically analyzes "typical" or "average" behavior, not the psychology of surprise. We emphasize the relevance of attention to safety, to automation, and to the psychology of surprise and briefly summarize some research on automation and free flight that highlights this linkage. Implications for design, training and modeling and research are then described.

INTRODUCTION

Designers of the national airspace, and that in Europe must deal with the issue of gridlock. Bottlenecks, delays and cancellations already plague commercial air travel, and projections are for continued growth of demand for air travel over the next decades. As in all high risk systems, such as medical care and ground transportation, so also in air travel, this pressure for "productivity" must be balanced against the demands for safety; and there is indeed a very real danger in tipping the scales in favor of productivity; this in spite of the FAA's clearly stated goals for an 80% reduction in fatal accidents over the next few years. As has often been pointed out, even if the <u>rate</u> of accidents remains constant, the increased number of flights, required to meet the increasing demands, will serve to <u>decrease</u>**,** rather than increase flight safety.

What then can psychologists contribute to addressing this tradeoff between safety and productivity, particularly those of us who come from the tradition of experimental psychology and human performance theory. The purpose of my talk today will be to suggest that there is quite a bit that we can contribute, particularly if we modify in some respects the way that we chose our paradigms, and treat our statistics.

What I will do first, is present what I see as a "conspiracy" of factors that inhibit these contributions, then describe some of the research that we have carried out, that tries to link the psychological study of attention and automation to the issues of safety; and finally to highlight five general guidelines and implications of such research.

The "Conspiracy"

Our conspiracy begins with the simple fact that in aviation, fortunately, accidents are rare. This is inherent in Reason's (1990) famous "Swiss cheese" (Amalberti, 2000) model of accidents in highly safe but complex systems: a lot of things have to go wrong, and the probability that all of them will, in a well designed system, is inherently very low. Represented statistically, we can say that only very unsafe circumstances will produce accidents, and hence we are dealing with the "tail" of some frequency the distribution of safe conditions (Figure 1). As the figure shows, such conditions might be represented by a combination of unusually bad weather, unusually high time pressure and, an atypically poorly trained or poorly skilled pilot (or controller). Finally, we may posit that many accidents are triggered by events that are themselves highly unusual, and therefore unexpected. Thus, we can infer that a reasonable source of variance in accident causation is inherent in the "psychology of surprise" or the human response to the unexpected.

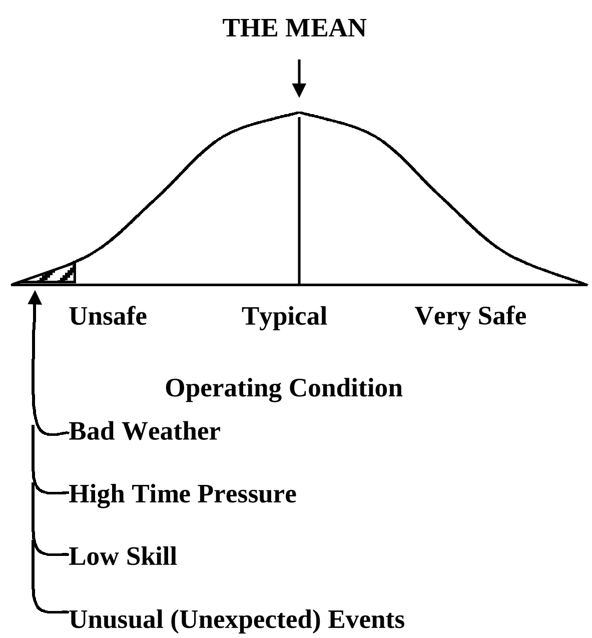

Figure 1.

The second element of our conspiracy relates to the statistics of surprise. Psychologists don't often do a very good job of studying the psychology of surprise – the response to extremely rare events. (I note as an exception, the wealth of data on vigilance performance; but most vigilance experiments do not involve truly surprising events; but only events that happen infrequently within the context of a simulated duty period.) As we as we know, researchers rely very heavily on the traditional statistics of the expected value or mean; (t-tests, ANOVAs and so forth); but as shown in Figure 1, accidents more typically involve the tail of the distribution: the least prepared, least skilled pilot, rather than the typical one. Unsafe behavior, in very safe systems is NOT at the mean, but out in the tail.

A second reason, somewhat related to the first pertains to statistical power. Surprise is, by definition, low N (in experimental settings, studying surprising events would have to be low N). But as all statisticians know, low N means low statistical power; and low power means that we are unlikely to get "statistically significant" effects out of our research on surprising events in experimental setting; that is, effects that meet the time-worn criterion of the "05 level" (or even the .10 level). And when less significant effects are obtained, there is <u>less</u> incentive to try to publish them, and there is often a strong reluctance by journal reviewers to get them through the publication process.

For an event to be truly surprising, in an experimental simulation, it must typically occur no more than once per subject, so that power can only be obtained by running repeated subjects, with only one observation per subject. Since these events are surprising, their latency will be long, and usually quite variable, further thwarting the effort to gain "significance".

One elegant example of success in this endeavor, combating this element of the conspiracy was an experiment by Summala (1981), examining driver responses to very unexpected road hazards. He had a confederate in a parked car by the roadway unexpectedly opening the car door in the face of an oncoming vehicle; a camera recorded the time between the door opening and the initiation of an avoidance maneuver; an elegant and valid measure of truly surprising hazard response time that was obtained for several hundred drivers. Here again we note that the variability, and particularly the extreme (long) values are probably more meaningful to safety than are the mean values, since it is these extremes of inattention and delayed responses that are most likely to underlie the accidents.

Two other examples of these data, collected from aviation experiments, are provided by Beringer and Harris (1999), in his studies of automation failure response, and by studies that have examined pilots detecting surprising runway incursions, while flying with a HUD (Wickens & Long, 1995; Fischer, Haines, & Price, 1980; Fadden, Ververs, & Wickens, in press).

In our experiment on runway incursions, we observed that a mean response time of 12 seconds, while long in itself, was accompanied by a variability that extended out to as long as 18 seconds. Even more dramatically were the response times found by Beringer in pilots' reactions to totally unexpected failures of automation (e.g., a runaway trim setting). These were highly variable between pilots, and sometimes well over a minute in duration.

What is important in these studies then is not the statistical significance of the differences between conditions, but rather the absolute length and distribution of response times, including particularly the longest response time. In terms of its safety implications, such an extreme RT should not be considered as an "outlier" to be discarded, but a critically important data point.

Another challenge in aviation safety research is to identify the appropriate contexts in which such surprising events occur; and would be likely to link to accidents. This is the domain of the safety models, that are currently the focus of work at NASA Ames research center. Since accidents are rare (and hence do not avail themselves of good material for statistical inference), to what extent can we rely on incidents to represent the precursors of accidents? And therefore capture those incidents in laboratory/simulation paradigms, and examine the effect of interventions in reducing their number. As we know, the linkage between the incident data base from ASRS, and the accident analysis, revealed by the NTSB reports is critical here. But this linkage is also challenged by two impediments.

First, as we noted, the linkage between accidents, serious incidents, and less serious ones needs to be better established. As shown in Figure 2, the pyramid represents schematically the inferred relation between five types of "events". Accidents (midair collisions) are at the top, and incidents of decreasing levels of severity, but increasing frequency, are arrayed toward the bottom. We infer that there is a causal relation from bottom to top. Thus, psychologists who develop interventions based on laboratory simulation research would like to believe that those which reduce the frequency of events at the lower level, will have a

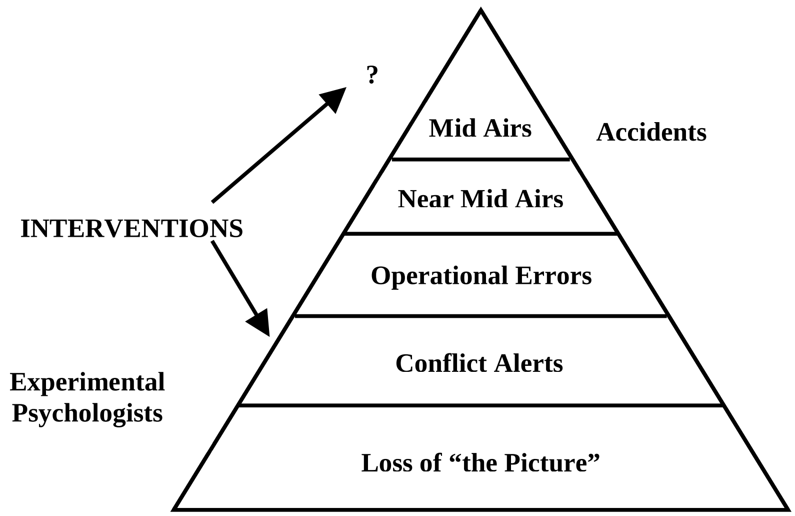

Figure 2

corresponding effect at higher levels. But such causal flow must be better established through research and safety models.

Second, valuable as the ASRS reports are, there is a potential concern that a subset of reports, illustrating a phenomenon of interest to the investigator, (and used to document the safety-relevant value of investigating the phenomenon), may not be as closely tied to safety as it seems (Amalberti, 2000). A typical use of ASRS data may start by asserting that: "40 reports on phenomenon X were located during the 5 year period of 19xx to 19yy."

While 40 may sound like a reasonably large number; it is probably a VERY small percentage of the 30,000 to 50,000 such reports that arrive each year (0.05%). Hence that phenomenon, fascinating as it may be to the applied psychologist, may not in fact be very safety relevant in operational circumstances; at least it cannot be guaranteed to be so on the basis of the claim that some number (e.g., 40) records were found (Amalberti, 2000).

At the end of this paper, I will summarize some of the implications of this conspiracy for psychological safety research. But now I would like to discuss in more depth, an emerging area where the mapping between psychological phenomena and safety issues is a very valid one and one in which the linkage to the psychology of surprise is very direct: the psychology of visual attention and its relation to automation.

Attention and Automation

The area of attention research has been close to my heart for over 30 years; and so (wishful thinking perhaps?) I can see its applicability to aviation safety issues. Indeed two recent accidents – the crash of the Singapore Airlines jet on takeoff in Taipei, and the sinking of the Japanese fishing boat Ehime Maru by the surfacing submarine Greeneville both illustrate breakdowns in operator attention; in one case the pilots apparently followed the attentional guidance of a prominent green line to a closed runway (while failing to attend to a less salient light signaling the open runway). In the other case, the operators failed to see the fishing boat on the surface through a periscope, despite the attentional guidance to look in that direction, which was offered by sonar.

We must start by establishing the importance and relevance of attention (or inattention) in aviation safety. To do so, we turn to Mica Endsley's work on situation awareness. Endsley (1995b) has classified situation awareness (and errors thereof) into 3 stages: failure to attend and perceive (i.e., notice), failure to integrate and understand, and failure to predict and project. While breakdowns in all 3 of these stages can have safety critical implications, a study that she carried out of NTSB accident reports verifies the heavy contribution of stage 1 errors to the accidents (Endsley, 1995a). Of 24 accidents analyzed, a majority (15) involved breakdowns in situation awareness, and of these, 72% of the errors were at level 1. A later study by Jones and Endsley (1996) of ASRS incidents revealed a remarkably similar percentage (75%) of SA-related incidents that were classified as breakdowns in stage 1.

One can of course add to this justification for the importance of attention research to transportation safety, the recent findings of distractions related accidents caused by cellular phones (Violanti & Marshall, 1996), and the statistics that inattentional errors are a leading cause of highway accidents (Malaterre, 1990). Finally, quite relevant here is the recent interest of psychologists in the phenomenon of “change blindness” or “inattentional blindness”, the very strong tendency to fail to notice prominent environmental changes if they are not directly in foveal vision at the time that the change occurs (Simons, 2000; Carpenter, 2001).

As I will argue below, attentional strategies are integrally related to human interaction with aviation automation, and increasing automation is a trend that will undoubtedly continue. Aviation researchers must struggle to deal with the human factors implications of new automation technologies designed to increase productivity and safety (Billings, 1996). Working with a National Research Council panel on Air Traffic Control Automation (Wickens, Mavor, Parasuraman, & McGee, 1998), we developed an information processing taxonomy of automation, which highlighted 4 stages of information processing operations, at which automation can assist human performance (Parasuraman, Sheridan, & Wickens, 2000).

Stage 1. Guiding attention by filtering and highlighting. A conflict alert system might do this, by highlighting a particular aircraft (or pair of aircraft) relevant to a future conflict.

Stage 2. Making intelligent inferences or diagnoses for the pilot or controller, by integrating information; for example an intelligent fault diagnostic system on board an aircraft, that can integrate symptoms of an engine failure (Hicks & de Brito, 1998).

Stage 3. Recommending a choice of action to the pilot; in the above, the system might provide a checklist of recommended actions based upon the inferred engine diagnosis

Stage 4. Executing actions. This is what an autopilot does. So too does an automated handoff for the air traffic controller, or an automated checklist that will automatically reconfigure parts of the aircraft that are not in their appropriate state, rather than waiting for the pilot to carry out the recommended actions manually.

It is worthwhile considering how a single system can incorporate automation at many stages. For example TCAS will call pilots’ attention to particular traffic aircraft by a color change on its traffic display (and filter traffic not assumed to be relevant; stage 1) TCAS will infer that certain traffic is likely to be in future conflict (stage 2), and it will recommend certain maneuvers through a resolution advisory (stage 3). While TCAS will not actually execute those maneuvers (stage 4), a parallel system for terrain avoidance, known as AUTOGCAS for fighter aircraft, accomplishes such automated response execution, by flying a “pullup” maneuver (Scott, 1999).

While there are many psychological issues associated with these different automation stages, my focus here will be directly on their attentional implications. To do so in Figure 3, I present a simplified model of human attention, in a complex visual world with many data sources shown at the bottom of the figure (the flight deck, the ATC workstation). Human attention, as we know, is attracted by <u>salient</u> events in the environment – the flashing light, the highlighted checklist item, or the prominent line on the pavement. And human attention will avoid non salient ones. As shown in Figure 3, these are referred to as bottom up effects.

At the same time, attention is guided by knowledge driven or top-down processes on the basis of two factors: <u>Expectancy</u> – we look where we expect to find information. This is key to the psychology of surprise. <u>Relevance</u> – we look to information sources relevant to the important tasks and goals at hand. Attention allocation is thus a balance between bottom up and top-down processes that is modulated by a fourth factor, the <u>effort</u> required to move attention from one location to another. Preliminary efforts are underway to derive comprehensive computational models of the net effects of these four variables in directing visual attention in complex environments (Wickens, Xu, Helleberg, & Marsh, 2001; Wickens, 2000).

As shown in Figure 3, automation at its different stages has a direct influence in guiding or influencing attention. At stage 1, automation will highlight data sources deemed to be important (and may filter those inferred to be of no relevance). At stage 2, by advising that a certain condition or diagnosis exists, automation will implicitly direct attention to those cues, or sources of data that support the existing condition; this is a phenomenon that has been described by cognitive

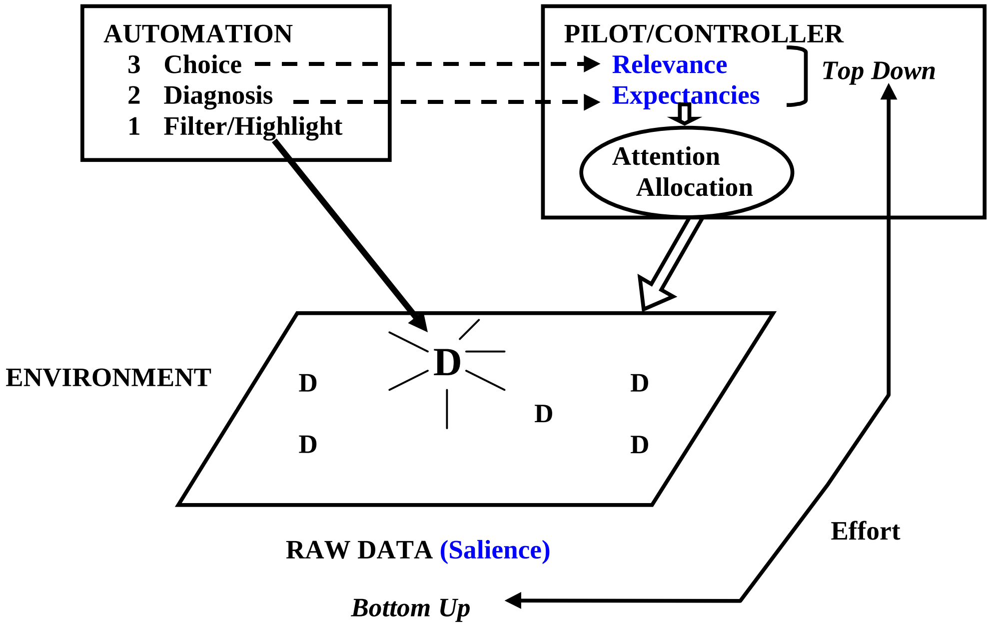

Figure 3

psychologists as the "confirmation bias" – the tendency to seek information consistent with prior beliefs or, in this case, the beliefs established by the automated diagnosis (Mosier, Skitka, Heers, & Burdick, 1998; Skitka, Mosier, & Burdick, 2000).

At stage 3, by recommending a particular course of action, automation will implicitly direct attention toward information relevant to that course of action; but may, by implication, divert attention away from other relevant data sources.

Finally, at stage 4; automation that fully accomplishes tasks including the execution of their responses, will leave attention free to wander to other activities, a possible form of neglect of the fully automated task that could be serious if the automation is faulty or imperfect – the so called "complacency effect" (Parasuraman, Molloy, & Singh, 1993; Parasuraman & Riley, 1997).

We turn now to a sampling of some of the laboratory research findings that reveal that operators may have problems in appropriately allocating attention when using such automation systems. Such problems may be particularly likely, under the very <u>unlikely</u> circumstances when the automation operations are imperfect or faulty (i.e., automation operates in ways other than those intended by the designer, and expected by the user; Wickens, 2000).

At stage 3 for example, we have found that compelling flight path guidance cues, such as the highway or pathway in the sky; while doing an excellent job of guiding the pilot to execute the appropriate commands to maintain the flight path, may also channel or focus the pilots' attention to the forward path, and thereby restrict sensitivity to other hazards, particularly those not directly visible in the forward view (Olmos, Wickens, & Chudy, 2000; Fadden, Ververs, & Wickens, in press).

At stage 2; I call attention to the series of studies that Mosier and her colleagues have conducted on intelligent fault diagnoses systems for the cockpit (Mosier et al., 1998; Skitka et al., 2000). On the rare (and therefore unexpected) occasions when such diagnoses may be incorrect, pilots were found, not only to follow the implicit guidance of the automation (which resulted in shutting down the wrong engine), but also to fail to attend to correct information that directly contradicted that diagnosis, as inferred from their later report.

We have recently completed a study in which automation inference of the intent (future flight path) of an aircraft was occasionally incorrect (i.e., generally reliable, but imperfect; Wickens, Gempler, & Morphew, 2000). Here too, we found strong evidence that pilots attended to that flight path intent information, as represented by the predictor symbol, and attended with less vigilance to the "raw data", in this case the actual flight path of the traffic aircraft as represented by its symbol on the traffic display. Hence, they encountered problems when the predictor was faulty.

Finally, a larger set of studies in our laboratory has looked directly at stage 1 automation; where intelligent agents explicitly direct the user's attention to events and objects assumed to be of importance (Yeh, Wickens, & Seagull, 1998; Yeh & Wickens, 2001, in press; Wickens, Conejo, & Gempler, 1999; Davison & Wickens, 2001). For the audience familiar with the basic research paradigms in visual attention, we have tried to model our effects within the framework of the cost-benefit analysis of attention proposed by Posner (1978) several years ago; and we have often used visual scanning measures to directly track the allocation of visual attention.

As one example, we have found, in a rotorcraft simulation , that directing pilots' attention to important hazards like powerlines or terrain by cueing, while offering benefits to the detection of those hazards, will direct attention away from other uncued hazards in the area (Davison & Wickens, 2001).

We have also studied automation-based attentional guidance effects in a series of studies that have examined the implications of freeflight, and the cockpit display of traffic information (CDTI) for pilot visual attention (Wickens, Xu, Helleberg, & Marsh, 2001; Wickens, Xu, Helleberg, Carbonari, & Marsh, 2000). Using eye movements to infer the allocation of attention, we have found that approximately 25% of a pilot's visual attention is occupied by looking at the CDTI to monitor for traffic and derive conflict avoidance maneuvers. (At some times this percentage increases to nearly 40%.) Furthermore, this attention is "borrowed" (reallocated) primarily from visual attention to outside monitoring of the forward view, rather than from the instrument panel.

Our interest in this research is whether the pilots' reliance upon the CDTI to guide their attention to look outside for traffic, may in fact act as a focusing tool that will inhibit the detection of traffic which is **not** on the CDTI. That is, an aircraft which does not have the technology to be known by the CDTI data base, or whose necessary transponder equipment is inoperable. We have found that, as long as pilots are prepared for the existence of these infrequent "rogue aircraft", their detection is not particularly inhibited. That is, pilots appear to be able to "optimally" calibrate their attention allocation strategies, noticing these rare, but nevertheless possible aircraft in a timely fashion. In the context of the psychology of surprise, we say that these aircraft are "unexpected", but not "truly surprising".

In another experiment, conducted on a Boeing 747 simulator, we have examined the tradeoff between bottom up (salience) and top down (expectancy) guidance of visual attention as pilots monitor the flight management system of an automated flight deck (Mumaw, Sarter, & Wickens, 2001; Mumaw et al., 2000). Our interest was in the problems of mode awareness, as related to a series of accidents that occurred during the 1990s (Hughes, 1995; Hughes & Dornheim, 1995), and well documented in the research and analysis of Sarter and Woods (1994, 1997, 2000). Well trained pilots are not always adequately aware of the mode of automation that is active or "armed" within the FMS. Such breakdowns in mode awareness, under unexpected or unusual operating conditions have critically important safety implications.

Our interest particularly focused on the pilots' attention to the flight mode annunciators: small boxes above the primary flight display that provide critical information regarding the operating mode of the FMS. We used visual scanning measures to study the trade-off between top down and bottom up processing, in guiding attention to the FMAs upon their change. Using eye movement recording, we noted first, that these were only a very small part of the pilot's natural scan pattern, being fixated, on the average, less than 2% of the time, We found, secondly that even upon a change in FMA status, signaled by the appearance of a green box which was designed to capture attention, pilots scanned to the FMA only 60% of the time, and even when they did, their scan was often delayed by 5-10 seconds after the green box onset. Hence the bottom-up attention grabbing properties of the green box may have been inadequate. In the context of the model provided in Figure 3, pilots' attention and scanning appeared to be much more directly driven by top down, than by bottom up factors; so it becomes important to establish that these top down factors are accurate and are well calibrated to the frequency and importance of real world events (mode changes) as these events are signaled for example by the green box mode changes.

I would now like to offer a few general conclusions of this research, and much other work that I have not reported here, based upon the work of Carbonnell, Ward, and Senders (1968), Moray (1986), Senders (1980), Sheridan (1970), and others.

1. Pilots are reasonably accurate in calibrating their attention allocation strategies to the bandwidth (frequency) of events, perhaps slightly

oversampling channels with very rare (but experienced) events; but undersampling those with important events (but see 4. below).

2. Salient events such as onsets will sometimes capture attention, but not invariably so (Nikolic, Orr, & Sarter, 2001).

3. The cognitive or physical effort required to shift or move attention long distances (e.g., via head movement) will sometimes inhibit that shift.

4. There is a concern that sources that contain non-salient but important events that have **never been experienced** by the operator may be undersampled. That is, not just the unexpected event – for which we are well calibrated – but the truly surprising event that has never helped before to that operator.

Implications

I hope that I have established the importance of attention – how the pilot and controller monitors the complex array of spatial information availed by both old and new technology – for the safety of flight. I see implications of this linkage to five aspects of aviation psychology.

1. Design. The design of the flight deck or ATC work station has tremendous implications for the allocation of attention. How well can salient display events capture and direct attention to important information? Our analysis of the Boeing data indicate that the simple illumination of a green box may be insufficient to attract attention (see also Nikolic, Orr, & Sarter, 2001), and Sarter's work on alternative modality cueing offers some promise here (Sklar & Sarter, 1999). But how much salience is "too much" (the loud annoying alarm, or flashing light; Woods, 1995)? These events may channel attention so that important raw data remain outside of the focus of attention (Yeh et al., 1998). Models of pilot expectancy can reveal the conditions under which more, rather than less attention guidance should be offered by more, rather than less salient cueing.

Correspondingly, good design should address the effort required to re-direct attention from one point of interest to another (Wickens & Carswell, 1995; Wickens, Vincow, Schopper, & Lincoln, 1997). This is, after all, the valuable asset that HUDs have achieved, in reducing (although NOT eliminating) the effort required to transition attention between the instruments and the world beyond (Wickens, 1997; Fadden, Wickens, & Ververs, 2000).

2. Training. New designs must be introduced hand in hand with training. Indeed much of the recent and valuable work on FMS training has focused on trying to remediate some of the deficiencies of the original FMS design, teaching pilots to better understand the logic of the FMS, and the meaning of its modes (Polson, Irving, & Irving, 1994; Feary, McCrobie, Alkin, Sherry, Polson, Palmer, & McQuinn, 1998; Sherry & Polson, 1999). Such training has direct implications for attention. A better mental model will provide more accurate top down guidance of where to look (and when) in order to anticipate mode changes.

Outside of the automated aircraft, recent research by Funk (1991; Chou, Madhavan, & Funk, 1996; Suroteguh & Funk, 2001; Colvin & Funk, 2001), and Dismukes (2001) have documented the breakdowns in attention management skills of pilots, leading to incidents in the ASRS data base, and in some cases to accidents. As Dismukes have noted, such breakdowns have direct implications for attentional training. The research by Gopher, Weil, and Bareket (1994) with the Israeli air force has suggested that such training in attention management can lead to valuable outcomes. In our own work (Wickens, Xu, Helleberg, & Marsh, 2001), we have observed the strong correlation between the amount of out the window scanning, and the success at visually sighting traffic, suggesting that successful monitoring strategies can be trained.

There is, however, an important caveat to training of attention allocation skills to deal with "truly surprising events" (those never experienced before). This is that people generally adopt "optimal" scanning strategies in which they are found to look more frequently at sources of greater frequency information. However people who have never experienced an event – the imperfection of an extremely reliable system, may never sample the information relevant to that event. This strategy is indeed "optimal" according to formal prescriptions (Moray, 2000). However, this form of behavior is, in fact, neither optimal nor wise. A challenge for training continues to be to develop attention allocation strategies to "expect the impossible".

3. Human in-the-loop simulation. As several investigators have noted, aviation systems are both complex and highly reliable (Amalberti, 2000). Hence on the rare occasions when they DO fail, they can fail in numerous complex, and unpredictable ways, often involving unpredictable human errors. Thus, for investigators, examining new technology such as that to be involved in distributed air-ground traffic management for free flight, it is essential to establish the nature of the operator response to such surprising

events (Wickens et al., 1998). Failure modes analysis, is a critical tool to identify some of the unexpected and "worst case" failure scenarios, so that simulation experiments can be conducted with fully measured performance of pilots and controllers in the loop. Those experimental participants should not be forewarned of the unexpected failures that might occur. That is, those events should be "truly surprising", and not just infrequent. Then, in the analysis of such performance, as I noted at the outset, it becomes important for the analyst to go beyond the "statistics of the mean" and pay attention to those statistics of the extreme, that can identify the worst case response, in the worst case scenario. For example, this might be a pilot who silently fails to comply with an ATC instruction (or automation recommendation), at the worst possible moment. These data are not "typical", but then again, neither are the accidents for which they are most relevant. If safety is at issue, than these data should not be discarded as "outliers" in order to normalize the data for conventional statistical analysis.

Unfortunately, the very complexity of such systems, and their relation with unpredictable human response, guarantees that it will be impossible to collect human performance data across the wide range of failure mode scenarios. This is particularly the case, given that the valid response to the truly surprising event is often captured by only a single data point from one subject in a complex experiment. Given however that such data are badly needed to examine the safety impact of new scenarios this leads to the importance of computer simulation modeling.

4. Modeling. A solution to the problem of complex time consuming human-in-the-loop simulations lies in valid pilot and controller performance models, such as those being developed at NASA Ames (Corker, Pisanich, & Bunzo, 1997; Tyler, Neukom, Logan, & Shively, 1998). Such models must be populated by <u>valid</u> estimates of the mean and extreme values of human performance parameters. If so, they can provide a rapid technique for estimating response time and accuracy for truly surprising events in new aviation technology; that is, the relevance of the psychology of surprise.

5. Research and data analysis. To return to the point that I made at the outset, in order to populate such models with valid data, it becomes very important for applied psychologists, working at all levels of simulation fidelity, to try to obtain data on the psychology of surprise; and share statistics of the mean and the extreme values of RT, accuracy and attention allocation measures, with those engaged in the modeling efforts described above. Extreme values should be reported in any document describing research whose such a manipulation of "surprise" has been carried out.

Conclusion

In conclusion, I hope that I have shown some of the contributions that psychologists can make to safety research. I have emphasized the psychology of attention, because of my own interests, its amenability to computational models, its documented importance in aviation accidents, and its close relation to the psychology of surprise. However, this is certainly far from the only (or even the most important) contribution that psychologists can make to safety. We can point to the study of risk perception and its effects on choice behavior, and to the study of communications, as two additional safety critical areas of research. These, and many others are all mediated by the growing influence of automation. We believe that valid psychological research with sensitivity to statistical issues described above, has a great deal to offer in reducing the frequency of accidents and their incident pre-cursors, and hence in increasing the safety of air travel.

ACKNOWLEDGMENT

This paper was supported by research support NASA Ames Research Center NASA NAG 2-1120. Dr. David Foyle was the scientific monitor, and the Federal Aviation Administration DTFA 98-G-022: Dr. Dennis Beringer was the scientific technical monitor.

REFERENCES

Amalberti, R. (2000). The paradoxes of almost totally safe transportation systems. <u>Safety Science, 1</u>, 1-16.

Beringer, D. B., & Harris, H. C., Jr. (1999). Automation in general aviation: Two studies of pilot responses to autopilot malfunctions. <u>The International Journal of Aviation Psychology, 9</u>(2), 155-174.

Billings, C. (1996). <u>Toward a human centered approach to automation</u>. Englewood Cliffs, NJ: Lawrence Erlbaum.

Carbonnell, J. F., Ward, J. L., & Senders, J. W. (1968). A queuing model of visual sampling: Experimental validation. <u>IEEE Transactions on Man Machine Systems, MMS-9</u>, 82-87.

Carpenter, S. (April, 2001). Sights unseen. The Monitor on Psychology, 32(4). American Psychological Association.

Chou, C., Madhavan, D., & Funk, K. (1996). Studies of cockpit task management errors. The International Journal of Aviation Psychology, 6(4), 307-320.

Colvin, K., & Funk, K. (2001). Investigating the basis for task prioritization. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Corker, K., Pisanich, G., & Bunzo, M. (1997). A cognitive system model for human/automation dynamics in airspace management. 13th International Ergonomics Association Triennial Congress. Tampere, Finland.

Davison, H. J., & Wickens, C. D. (2001). Rotorcraft hazard cueing: The effects on attention and trust. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Dismukes, K. (2001). The challenge of managing interruptions, distractions, and deferred tasks. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Endsley, M. R. (1995a). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston, & N. McDonald (Eds.), Human factors in aviation operations (pp. 287-292). Aldershot, England: Avebury Aviation, Ashgate Publishing Ltd.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 85-104.

Fadden, S., Wickens, C. D., & Ververs, P. M. (2000). Costs and benefits of head-up displays: An attention perspective and a meta analysis (Paper # 2000-01-5542). 2000 World Aviation Congress. Warrendale, PA: Society of Automotive Engineers.

Fadden, S., Ververs, P. M., & Wickens, C. D. (in press). Pathway HUDS: Are they viable? Human Factors.

Feary, M., McCrobie, D., Alkin, M., Sherry, L., Polson, P., Palmer, E., & McQuinn, N. (1998). Aiding vertical guidance understanding (NASA/TM-1998-112217). Moffett Field, CA: NASA Ames Research Center.

Fischer, E., Haines, R. F., & Price, T. A. (1980). Cognitive issues in head-up displays (NASA Technical Paper 1711). Moffett Field, CA: NASA Ames Research Center.

Funk, K. H. (1991). Cockpit task management: Preliminary definitions, normative theory, error taxonomy, and design recommendations. The International Journal of Aviation Psychology, 1(4), 271-285.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. Human Factors, 36(4), 387-405.

Hicks, M., & de Brito, G. (1998). Civil aircraft warning systems: Who's calling the shots? In G. Boy, C. Graeber, & J. M. Robert (Eds.), International Conference on Human Computer Interaction in Aeronautics (HCI-Aero'98) (pp. 205-212), Montreal, Canada: Ecole Polytechnique de Montreal.

Hughes, D. (1995, January 30). Incidents reveal mode confusion. Aviation Week and Space Technology, 56.

Hughes, D., & Dornheim, M. A. (1995, January 30). Accidents direct focus on cockpit automation. Aviation Week and Space Technology, 52-54.

Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. Aviation, Space, & Environmental Medicine, 67(6), 507-512.

Malaterre, G. (1990). Error analysis and in-depth accident studies. Ergonomics, 33, 1403-1421.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and performance, Vol II (pp. 40-1-40-51). New York: Wiley & Sons.

Moray, N. D. (2000). Are observers ever really complacent when monitoring automated systems? Proceedings of the International Ergonomics Association and 44th Annual Meeting of the Human Factors & Ergonomics Society. Santa Monica, CA: Human Factors & Ergonomics.

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.

Mumaw, R., Sarter, N., & Wickens, C. (2001). Pilot monitoring and performance in an automated flight deck. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Mumaw, R., Sarter, N., Wickens, C. D., Kimball, S., Nikolic, M., Marsh, R., Xu, W., & Xu, X. (2000). Analysis of pilots' monitoring and performance on highly automated flight decks (Final NASA Project Report). Moffett Field, CA: NASA Ames Research Center.

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2001). Supporting pilot-automation communication through display design: Do visual onsets always capture attention. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Olmos, O., Wickens, C. D., & Chudy, A. (2000). Tactical displays for combat awareness: An examination of dimensionality and frame of reference concepts and the application of cognitive engineering International Journal of Aviation Psychology, 10(3), 247-271.

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.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D (2000, May). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, & Cybernetics, 30(3), 286-297.

Polson, P. G., Irving, S., & Irving, J. E. (1994). Applications of formal models of human-computer interaction to training and use of the control and display unit (Final Report). Federal Aviation Administration.

Posner, M. (1978). Chronological explanations of mind. Erlbaum.

Reason, J. (1990). Human error. New York: Cambridge University Press.

Sarter, N. B., & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. The International Journal of Aviation Psychology, 4(1), 1-28.

Sarter, N. B., & Woods, D. D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. Human Factors & Ergonomics, 39, 553-569.

Sarter, N., & Woods, D. D. (2000). Team play with a powerful and independent agent: A full-mission simulation study. Human Factors, 42(3), 390-402.

Scott, W. B. (1999, February 1). Automatic GCAS: You can't fly any lower. Aviation Week & Space Technology, 76-80.

Senders, J. W. (1980). Visual scanning processes. Unpublished doctoral thesis. The Netherlands: University of Tilburg.

Sheridan, T. (1970). On how often the supervisor should sample. IEEE Transactions on Systems Science and Cybernetics, SSC-6(2), 140-145.

Sherry, L., & Polson, P. G. (1999). Shared models of flight management system vertical guidance. The International Journal of Aviation Psychology, 9(2), 139-154.

Simons, D. J. (2000). Current approaches to change blindness. Visual Cognition, 7(1/2/3), 1-15.

Skitka, L. J., Mosier, K. L., & Burdick, M. D. (2000). Accountability and automation bias. International Journal of Aviation Psychology, 52, 701-717.

Sklar, A. E., & Sarter, N. B. (1999). Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. Human Factors, 41(4), 543-552.

Summala, H. (1981). Driver/vehicle steering response latencies. Human Factors, 23, 683-692.

Suroteguh, C., & Funk, K. (2001). The effect of flight deck automation and automation proficiency on cockpit task management performance. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Tyler, S. W., Neukom, C., Logan, M., & Shively, J. (1998). The MIDAS human performance model. Proceedings of the 42nd Annual Meeting of the Human Factors & Ergonomics Society. Santa Monica, CA: Human Factors Society.

Violanti, J. M., & Marshall, J. R. (1996). Cellular phones and traffic accidents: An epidemiological approach. Accident Analysis and Prevention, 28(2), 265-270.

Wickens, C. D. (1997). Attentional issues in head-up displays. In D. Harris (Ed.), Engineering psychology and cognitive ergonomics: Transportation systems, Vol. 1 (pp. 3-21). Aldershot, UK: Ashgate Publishing Ltd.

Wickens, C. D. (2000). Imperfect and unreliable automation and its implications for attention allocation, information access and situation awareness (Final Technical Report ARL-00-10/NASA-00-2). Savoy, IL: University of Illinois, Aviation Res. Lab.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and its relevance to display design. Human Factors, 37(3), 473-494.

Wickens, C. D., Conejo, R., & Gempler, K. (1999). Unreliable automated attention cueing for air-ground targeting and traffic maneuvering. Proceedings of the 43rd Annual Meeting of the Human Factors & Ergonomics Society. Santa Monica, CA: Human Factors & Ergonomics Society.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. Transportation Human Factors Journal, 2(2), 99-126.

Wickens, C. D., & Long, J. (1995). Object versus space-based models of visual attention: Implications for the design of head-up displays. Journal of Experimental Psychology: Applied, 1(3), 179-193.

Wickens, C. D., Mavor, A. S., Parasuraman, R., & McGee, J. P. (Eds.). (1998). The future of air traffic control: Human operators and automation. Washington, DC: National Academy Press.

Wickens, C. D., Vincow, M. A., Schopper, A. W., & Lincoln, J. E. (1997). Computational models of human performance in the design and layout of controls and displays (CSERIAC SOAR Report 97-22). Wright Patterson AFB, OH: Crew System Ergonomics Information Analysis Center.

Wickens, C. D., Xu, X., Helleberg, J. R., Carbonari R., & Marsh, R. (2000). The allocation of visual attention for aircraft traffic monitoring and avoidance: Baseline measures and implications for free flight (Technical Report ARL-00-2/FAA-00-2). Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C. D., Xu, X., Helleberg, J., & Marsh, R. (2001). Pilot visual workload and task management in freeflight: A model of visual scanning. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. Ergonomics, 38(11), 2371-2393.

Yeh, M., & Wickens, C. D. (2001, in press). Explicit and implicit display signaling in augmented reality: The effects of cue reliability, image realism, and interactivity on attention allocation and trust calibration. Human Factors.

Yeh, M., Wickens, C. D., & Seagull F. J. (1998). Effects of frame of reference and viewing condition on attentional issues with helmet-mounted displays. Proceedings of the 2nd Annual FedLab Symposium: Advanced displays & interactive displays (pp. 107-114). College Park, MD: U.S. Army Research Federated Laboratory Consortium.