Neuroergonomics and Cognitive Engineering, Vol. 102, 2023, 58–65
https://doi.org/10.54941/ahfe1003005

Limitations on the Use of Eye-Tracking Data to Understand Operator Awareness

Randall J. Mumaw[1] and Dorrit Billman[2]

[1] San Jose State University Foundation, San Jose, CA 95112, USA
[2] NASA Ames Research Center, Moffett Field, CA 94035, USA

ABSTRACT

Safety events in commercial aviation have implicated failures of attention and awareness, and the need for improved attention and awareness has pushed trainers toward eye-tracking systems as a primary tool for evaluating monitoring performance. Further, in the last 15 years, eye-tracking technology has become easier to acquire and use in operationally realistic settings. Although we believe that eye-tracking can be a useful evaluation tool, this paper describes basic limitations on relying on eye-tracking as the sole means for evaluating attention and awareness. A set of other measures are offered that can better identify where performance breakdowns occur.

Keywords: Commercial aviation, Attention, Awareness, Eye-tracking, Performance measures

INTRODUCTION

In the last 25 years, a number of accidents and serious incidents in commercial aviation have implicated loss of flightcrew awareness of basic flight path parameters or autoflight state. Standard examples are Turkish Airways 1951 and Asiana 214 (loss of awareness of decreasing airspeed and increasing pitch on approach); Atlas Air 3591 and Armavia 967 (loss of awareness of excessive nose-down pitch); Flash Airlines 604 and Kenya Airways 507 (loss of awareness of large bank angle changes); and Bhoja Air 213 and the Thomsonfly incident at Bournemouth (loss of awareness of autoflight state). In these events, loss of awareness can lead to a surprise, inappropriate control inputs, and a loss of control in-flight (LOC-I). The Commercial Aviation Safety Team (CAST) performed an analysis of a set of 18 of these events as part of its Airplane State Awareness activity a decade ago (see CAST, 2014).

This cataloguing of airplane state awareness events led to a strong industry focus on improving pilots' attention allocation and airplane state awareness. In every one of the events analyzed by the CAST team, flightcrew attention had been shifted away from flight path management prior to loss of awareness. In some cases, this was due to a conscious decision about attention management (or task management) (e.g., doing paperwork), and in other cases, the loss of awareness was tied to human limitations (e.g., spatial disorientation).

Note, however, that both the initial CAST analysis and additional analysis (Mumaw et al., 2019) also identified a range of non-attentional factors that contributed to these safety events, reflecting the typical diverse inputs to accident causation. Examples are:

- Flightcrew impairment
- Airline safety culture
- Crew Resource Management (CRM)
- Ineffective airplane alerting schemes
- Inappropriate control actions
- Poor understanding of the autoflight system
- Invalid source data or loss of air data

These analyses have led to a significant industry response that—through mitigations such as display and alerting design, pilot training, and improvements to airplane systems—attempts to improve pilot awareness both broadly and more narrowly (i.e., around certain types of airplane state information). Evaluations of these mitigations require valid measures of attention and awareness. Eye-tracking, which is documenting the location, sequencing, and timing of eye fixations, is often used as a primary measure. In this paper, while we advocate for giving eye-tracking-related measures a role in these evaluations, we also discuss the limitations on these measures.

THE RANGE OF EYE-TRACKING MEASURES

First, we should describe the range of eye-tracking measures, taken from a broad review of eye-tracking studies in the aviation domain[1]. Certainly, other forms of capturing eye fixations are possible:

Percent Dwell Time (%DT) – One measure of where pilots are allocating their visual attention is a summary of how much time was spent in each Area of Interest (AoI), where AoIs are typically individual displays (e.g., Primary Flight Display (PFD)) or meaningful display elements (e.g., attitude indicator). This measure is typically presented as percent dwell time and indicates the percentage of total task time each AoI was fixated. This measure is averaged across a set of pilots, e.g., it might be reported that on average a pilot spent 30% of task time fixating the PFD during a flight.

Fixation Dwell Time – A related measure is the duration of each fixation. A few studies (see Wickens & Dehais, 2019) have suggested that short fixation dwell times may be a marker of expertise. Specifically, skilled pilots may be able to extract information more efficiently from each fixation (see also Moray, 1986). The potential advantage of this efficiency is that fewer attentional resources are spent on information extraction and are, therefore, available for higher-level cognitive functions.

Event-Triggered Fixations – %DT has the advantage of capturing central tendencies for a group of pilots over a period of time. However, %DT does not preserve the precise time at which those fixations occurred. There can be value in knowing when a fixation occurred, especially as it relates to timing

---

[1] Note that we are not interested (for this paper) in eye-tracking measures tied to operator sleepiness or alertness; we are assuming an alert pilot whose attention has shifted away from important indications.

of some operational event. For example, did the pilot fixate the autothrottle mode after it transitioned to HOLD and stopped managing airspeed?

Fixation Sequences (Scanning) – There are several ways to capture the fixation sequence, or scanning pattern. In one study, Haslbeck and Zhang (2017) applied a transition-matrix analysis to capture fixation sequences within the core flight instruments. This allowed them to identify frequently occurring sequences. Other researchers have focused on the variability in fixation sequences. This measure—which is, in some sense, the opposite of a scanning pattern—indicates how uncertain or unpredictable the fixation sequence is. Finally, Dehais et al. (2015) applied another scanning-related measure: the explore/exploit ratio. This measure characterizes both the sequence and duration of fixations. Specifically, it compares the number of saccades and short-duration fixations (those around 100 ms) to the number of long-duration fixations (those greater than 240 ms). When the former pair of measures increases, they describe this behavior as exploring, which is connected to searching for information. When the long-duration fixations increase relatively, the behavior is called exploiting, which is associated with a deeper processing of information.

AoI Neglect Latency – Another measure of scanning is the amount of time between fixations on a specific AoI; we refer to this as AoI neglect latency. Because there is so much information across the flight deck interface, as well as a need to look out the window or at paper charts and procedures, visual attention can be over-committed to a few AoIs over short time periods. The potential downside is losing the awareness of an important change on an unmonitored AoI, a failure at the heart of a number of aviation accidents (e.g., Turkish 1951). Of course, an ideal rate of fixating a particular AoI can change across flight phase; more frequent fixations are important when indications are changing more rapidly.

AoI Relevance – A few studies have focused on the specific indications that are attended, suggesting that more experienced pilots have a better understanding of which information is relevant to the current task. One study, in particular, is Bellenkes et al. (1997), which found that the more experienced pilots fixated indications that were related to the primary indications used for task performance.

LIMITATIONS ON EYE FIXATIONS AS A MEASURE

Eye-tracking technology has advanced significantly in the last 15 years, and while it can be fairly easy to capture and represent a pilot's fixation behavior at a fine-grained level, there are clear limits on what eye fixations can tell us about attention and awareness.

Fixations May Not Reveal Attentional Focus

The mind may not follow the eyes. Psychological research has established (e.g., Warm et al., 2008) that sustaining attention—remaining vigilant—on a monitoring task is resource-intensive and stressful, and, therefore, attention cannot be sustained over a long period of time without a considerable performance decrement. Indeed, even 15 minutes on a sustained attention task can lead to a performance decrement. To cope with this, humans take breaks from periods of effortful sustained attention. Casner and Schooler (2015), in trying to understand lapses in monitoring and failures to make routine callouts, found that pilots sometimes engage in mind-wandering, which is thinking about something other than what their eyes are fixating. Mind-wandering is diverse; pilots will think about an upcoming vacation, a family situation, or just something that is not the task at hand. Their eyes will be directed at some slice of the immediate environment, but attention may be directed inward. However, note that Casner and Schooler did find that pilots may impose some control over when they engage in mind-wandering and may suppress it when they anticipate a short, important period for monitoring.

Another way in which fixations may not align with attention is when attention becomes tied to a non-visual input source, such as a radio transmission. That is, the pilot may be visually fixating a display, but attention may be on interpreting a complex change to a clearance.

Another potential mismatch between attention and fixations is tied to peripheral vision. Even when pilots are fixated on basic flight deck indications, they can pick up useful information from their peripheral vision, especially for more dynamic changes. Examples are rate of change of airspeed or altitude tapes and the larger actions of the other pilot. Often, when something meaningful is picked up peripherally, the pilot is likely to re-orient and fixate that information, once again aligning fixation and attention. However, that prompt for re-orienting will not be picked up by eye fixations.

Even when attention is focused on what is fixated, it can be difficult to identify how closely aligned the fixation is with attention. Simons & Chabris (1999) identified a phenomenon, called "inattentional blindness," that can occur when attention is strongly driven by a task-focus or from expectations about what information is present. In their original study, they found that when viewers were asked to follow a particular element of the action in a video, almost 50% of those viewers (across conditions) could fail to notice unexpected and non-task-relevant actions that were in the same location—for example, a "gorilla" walking through. Thus, a pilot may fixate near but fail to see unexpected values or changes on the interface—for example, a visual check to confirm that the flaps are extended for take-off can fail to see that the flap setting is not correct, as occurred in the Spanair 5022 accident.

A more extreme phenomenon is called "channelized attention," where the pilot becomes focused on one understanding of what is happening and can lose awareness of even salient or central cues in the field of view. There have been several airplane accidents (e.g., Tatarstan 363) in which the pilot, likely disoriented from a vestibular illusion/spatial disorientation, pitched the airplane down toward the ground and continued nose down inputs until crashing into the terrain. In these accidents, the ground proximity alerting system is calling out "terrain, terrain," in some cases for as many as 10 seconds, without a change to pilot nose-down control inputs. The pilot is seemingly unaware of the alert and the impending collision. Again, outward signs of attention, such as eye fixations, would fail to reveal how attention was focused.

Finally, stress and fatigue, which are not uncommon conditions during flight operations, can lead to a narrowing of attention. One of the most widely reported effects of stress on performance of cognitive tasks is that, in stressful conditions, attention becomes more narrowly focused on cues central to a task and less sensitive to more peripheral cues (Hockey & Hamilton, 1983; Hancock & Warm; 1989). Sleepiness from fatigue can affect attention in similar ways (e.g., Lim & Dinges, 2010; Roca et al., 2012)

Fixations May Not Reveal Awareness and Understanding

If we assume that, despite these limitations, the vast majority of eye fixations are a reliable indicator of where visual attention is focused in the moment, what can we say about pilot awareness and understanding of airplane state?

Unfortunately, there are concerns here as well. Sarter et al. (2007), in looking at autoflight system use, found that eye fixations did not guarantee awareness and understanding. Specifically, Sarter et al. created inappropriate behavior in the autoflight system, such as artificial flight mode annunciations and mode misconfigurations. The eye-tracking data indicated that, across the various manipulations, very few of the pilots who fixated this information followed through and corrected it, leading to undesired operational outcomes in some cases. In one scenario manipulation, the Boeing 747–400 was configured into a VNAV ALT mode while in cruise, which means that the airplane will not descend at the top of descent (T/D) point. Four of the 20 pilots ensured that they had the VNAV PTH mode prior to the T/D point. Nine of the remaining 16 pilots fixated the inappropriate VNAV ALT mode but took no action to change it, seemingly unaware of how that mode would affect airplane behavior. This might indicate lack of awareness of the specific VNAV submode or a lack of understanding regarding how the submodes affect airplane behavior. Many of the pilots who failed to descend as expected expressed surprise about VNAV behavior; they did not understand why it failed to descend.

INTEGRATION OF MEASURES TO DEVELOP A MORE VALID EVALUATION

These weaknesses in relying on eye-tracking measures to inform about pilot attention and awareness point to the need to recruit a range of performance measures to develop a more-complete analysis of performance breakdowns. It can be valuable to have eye-tracking data, which can indicate which flight deck indications were, indeed, fixated, either throughout a period of time or as a response to an important change. Obviously, if there is no fixation, then it seems unlikely that awareness and understanding will follow (although peripheral vision might aid awareness in some cases). However, the goal is to develop strong evidence about where performance breaks down, considering a set of options such as fixation, awareness, understanding, projection, and action. The following types of performance data can aid in confirming the succession through these stages:

Flightcrew control actions – Evaluation scenarios can be created that require flightcrew inputs in response to a changing operational situation. For example, if the pitch mode during cruise changes to VNAV ALT, we would expect an aware pilot to take actions to change it to VNAV PTH in order to have the airplane descend at the T/D point. Indeed, appropriate control actions may be the gold standard for evaluation of attention and awareness. Those actions reveal no breaks in the chain. Evaluation at this level (see also airplane performance) aids in validating the fixation data. When the appropriate action occurs, one can feel fairly confident that the precursors occurred (although it is also important to measure those precursors as well: Did the pilot really understand how the submode change would alter behavior?)

Airplane performance outcomes – This form of evaluation means that not only did the pilot take the appropriate action, but the resulting airplane state also informed action appropriateness. In many cases, evaluation of actions will be the ideal measure, but there can be situations in which several action paths are appropriate, and the objective is better expressed as flight path or airplane outcomes. An example is that the airplane met the waypoint constraints (altitude and airspeed), and thus, the pilot attended to and understood the influences on flight path management.

Verbal reports – Verbal reports are statements that a pilot makes to another pilot, or to ATC, or to him or herself that reveal awareness and perhaps understanding—calling out a mode or an airspeed or saying you will be high on path. To increase the likelihood or frequency of these, you can ask pilots to communicate as a crew or to communicate to the evaluator to describe what they are aware of and why they are taking actions. This requires them to volunteer what they are noticing and how they are understanding it. Ideally, this form of reporting is integrated with their typical crew duties and is not adding burdensome "non-operational" tasks.

Situation Awareness measures (concurrent) – A specific form of verbal reporting is some form of situation awareness (SA) metric. Some SA measures try to capture concurrent awareness reports through prompts during performance. In one paradigm, the evaluator can obscure/hide the interface momentarily and ask the pilot to state the current value of a flight parameter. A variation is to ask the pilot to call out any indications that differ from what is expected.

Situation Awareness measures (retrospective) – Other SA metrics wait until an operational scenario has completed (so work is not disrupted by artificial tasks) to prompt about awareness. Unfortunately, these metrics add a memory component that may make certain types of awareness less likely to be reported. Ideally, SA metrics can be embedded in the operational scenario to avoid creating non-operational tasks that introduce performance incentives to monitor for the sake of responding to the non-operational task.

Conceptual knowledge – Another check on the links between fixations and action is the conceptual underpinning. As stated above, going from fixation to action requires understanding the significance of the indications being fixated. Does the pilot understand that the HOLD mode indicates that the autothrottle will not manage airspeed? Thus, sometimes performance failures may occur even when there is fixation and awareness, but the pilot does not fully understand the implications. In the Sarter et al. study mentioned above, they asked pilots about how each of the VNAV submodes behaves to see if there was sufficient understanding to go from fixation to action.

The preceding is not intended to be a "how to" on developing better measures for evaluating attention and awareness; it is meant to broaden the considerations for the types of measures to employ. These descriptions were short but, ideally, they can prompt evaluators to look beyond eye-tracking to find complementary measures that more directly measure awareness, understanding, and appropriate action. In turn, this may enable guidance on how measures of awareness contribute to understanding different aspects of learning and performance.

CONCLUSION

Understanding what information is attended, reaches awareness, and is understood can be valuable in identifying and assessing performance breakdowns in operational tasks or gaps in training. For example, it might help trainers identify where additional explanation or practice might be valuable. While the issue of monitoring was largely a response to accidents and incidents, effective monitoring is valuable for basic flight path management skill. In this paper, we described limitations on the value of eye-tracking measures to provide evidence of attention and awareness. After describing the range of eye-tracking measures and their limitations, we describe a set of other measures that can aid in pinpointing how operator performance broke down.

ACKNOWLEDGMENT

The authors would like to acknowledge funding for this work from the FAA's ANG-C1 human factors research division.

REFERENCES

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. *Aviat Space Environ Med., 68*(7), 569–579.

Casner, S. M., & Schooler. J. W. (2015). Vigilance impossible: Diligence, distraction, and daydreaming all lead to failures in a practical monitoring task. *Consciousness and Cognition, 35*, pp 33–41.

Commercial Aviation Safety Team (2014). Airplane State Awareness Joint Safety Analysis Team; Final report: Analysis and results. https://www.skybrary.aero/bookshelf/books/2999.pdf

Dehais, F., Peysakhovich, V., Scannella, S., Fongue, J., & Thibault, G. (2015). "Automation surprise" in aviation: Real-time solutions. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention, *Human Factors, 31*(5), 519–537.

Haslbeck A, Zhang B. (2017). I spy with my little eye: Analysis of airline pilots' gaze patterns in a manual instrument flight scenario. *Applied Ergonomics; 63*, 62–71.

Hockey, G. R. J., & Hamilton, P. A. (1983). The cognitive patterning of stress states. In G. R. J. Hockey (Ed.), Stress and fatigue in humanperformance (pp. 331–361). New York: Wiley.

Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact ofshort-term sleep deprivation on cognitive variables, *Psychological Bulletin, 136*(3), pp. 375–389.

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: Cognitive processes and performance. New York: John Wiley and Sons.

Mumaw, R. J., Billman, D., & Feary, M. (2019). Factors that influencedairplane state awareness accidents and incidents. CAST SE-210 Output 2. Report 2 of 6. NASA/TM-20205010985. Moffett Field, CA: NASA Ames Research Center.

Roca, J., Fuentes, L. J., Marotta, A., Lopez-Ramon, M., Castro, C., Lupianez, J., &Martella, D. (2012). The effects of sleep deprivation on the attentional functions and vigilance; *Acta Psychologica, 140*, pp. 164–176.

Sarter, N., Mumaw, R., & Wickens, C. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye tracking data. *Human Factors, 49*(3), 347–357.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*, 1059–1074.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*(3), 433–441.

Wickens, C. D., & Dehais, F. (2019). Aviation Expertise. In P. Ward, J. Schraagen, J. Gore, & E. Roth (eds.). *The Oxford Handbook of Expertise: Research and applications.* Oxford University Press.