NASA/TM–20210000047

Analysis of Pilot Monitoring Skills and a Review of Training Effectiveness

Randall J. Mumaw
San Jose State University Foundation

Dorrit Billman
NASA Ames Research Center

Michael S. Feary
NASA Ames Research Center

December 2020

NASA STI Program…in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question via to help@sti.nasa.gov

- Phone the NASA STI Help Desk at (757) 864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM—20210000047

Analysis of Pilot Monitoring Skills and a Review of Training Effectiveness

Randall J. Mumaw
San Jose State University Foundation

Dorrit Billman
NASA Ames Research Center

Michael S. Feary
NASA Ames Research Center

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California

December 2020

Acknowledgements

This work was performed under a contract with the FAA (DTFAWA-11-A-80004). The authors gratefully acknowledge careful reviews and discussion from Kathy Abbott, Cristoph Ammann, Sabreena Azam, Immanuel Barshi, Julia Haber, Hans-Juergen Hoermann, Roberty Koteskey, Robert Mauro, Keith Ruskin, Lance Sherry, Mike Stewart, and Chris Wickens.

Available from:

NASA STI Program
STI Support Services
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

This report is also available in electronic form at http://www.sti.nasa.gov
or http://ntrs.nasa.gov/

Table of Contents

List of Figures and Tables

Acronyms and Definitions

%DT....................................percent dwell time
ACARS................................Aircraft Communications Addressing and Reporting System
ACRM.................................Advanced Crew Resource Management
ACT ARC.............................Air Carrier Training Aviation Rulemaking Committee
ADI......................................attitude direction indicator
ALT......................................altitude
ANTS..................................Anesthesia non-technical skills
AoI......................................area of interest
AoV.....................................Areas of Vulnerability
AP..autopilot
APP.....................................approach (autopilot mode)
AQP....................................Advanced Qualification Program
AS..airspeed
ASA....................................Airplane State Awareness
ASIAS.................................Aviation Safety Information Analysis and Sharing
ATCO.................................air traffic control operators
ATIS...................................automatic terminal information service
ATM...................................air traffic management
ATQP..................................Alternative Training Qualification Program
BDAR.................................Brief-Discuss-Advocate-Resolve
CAST..................................Commercial Aviation Safety Team
CBSD..................................Concepts-Briefing-Simulation-Debrief
CBT.....................................computer-based training
CDU....................................control data unit
CPDLC...............................controller pilot data link communications
CRM...................................Crew Resource Management
DT.......................................dwell time
DVD....................................digital versatile disc
EFB.....................................electronic flight bag
EFPM..................................extended flight plan message
EM......................................error management
ESSAI.................................Enhanced Safety through Situation Awareness Integration
ET.......................................eye tracking
FAA....................................Federal Aviation Administration
FAF.....................................final approach fix
FltDAWG............................Flight Deck Automation Working Group
FMA....................................Flight Mode Annunciation
FMS....................................flight management system
FO.......................................First Officer
FOM....................................Flight Operations Manual
FOQA.................................Flight Operations Quality Assurance
FPM....................................flight path management
FPMWG..............................Flight Path Management Working Group
G/S......................................glide slope

GA..general aviation
HDG......................................heading
HDG SEL...............................Heading Select (autopilot mode)
HDT ....heads down task
HSI..Horizontal Situation Indicator
HUD......................................head up display
HUT ....heads up task
IATA......................................International Air Transport Association (of ICAO)
ICAO.....................................International Civil Aviation Organization
ICU ....intensive care unit
IFR ....instrument flight rules
ILS ....instrument landing system
IMC.......................................instrument meteorological conditions
KSA ....knowledge, skills, and abilities
kts..knots (approximately 1.151 miles per hour)
LNAV ....lateral navigation (autopilot mode)
LOC ....loss of control
LOE.......................................line-oriented evaluations
LOFT ....line oriented flight training
MCP ....mode control panel
ms..millisecond
NASA ....National Aviation and Space Administration
ND...navigation display
NG...next generation
nm ....nautical miles
NOTECHS ....non-technical skills
NOTSS.................................non-technical skills for surgeons
NTS.......................................non-technical skills
NTS.......................................non-technical skills
NTSB ....National Transportation Safety Board
OTW ....out the window
PARC ....Performance–based operations Aviation Rulemaking Committee
PC ....personal computer
PF..pilot flying
PM...pilot monitoring
RAPT ....Risk Awareness and Perception Training
RCT.......................................randomized clinical trial
RTO ....rejected take-off
SA ....situation awareness
SAGAT ....situation awareness global assessment technique
SE..Safety Enhancement
SEEV ....Salience/Expectancy/Effort/Value
SOP ....standard operating procedure
SPAM ....Situation Present Assessment Method
SPOT.....................................special purpose operational training
SRM......................................Single-Pilot Resource Management
STAR ....Standard Terminal Arrival Route
TEM......................................threat and error management
UK CAA ....United Kingdom Civil Aviation Authority

UPRT ....Upset Prevention and Recovery Training
VDI ....vertical deviation indicator
VNAV...................................vertical navigation (autopilot mode)
VNAV PTH ....vertical navigation path (autopilot mode)
VS ....vertical speed
VSD ....Vertical Situation Display
VVM ....verbalize, verify, monitor
WHO.....................................World Health Organization

Analysis of Monitoring Skills and a Review of Training Effectiveness

Randall J. Mumaw, Dorrit Billman, and Michael S. Feary

Introduction

The commercial aviation industry world-wide has identified a need for improved pilot monitoring and awareness (e.g., FAA, 2013, ICAO, 2016). More specifically, aviation safety data indicate that failures in pilots' flight path management (FPM) monitoring and awareness have contributed to a range of undesired outcomes: accidents, major upsets, and non-compliance with air traffic control (ATC) guidance. The Federal Aviation Administration (FAA) has further stated that these types of FPM failures are likely to worsen with the increasingly complex air traffic control systems and FPM concepts proposed for NextGen (https://www.faa. gov/nextgen/what_is_nextgen/) operations (e.g., see Hah et al., 2017). Adding to this complexity is the introduction of increasingly automated aircraft systems that can increase monitoring burdens. One potential mitigation for this situation is to enhance pilot training for effective monitoring.

NASA Ames Research Center was asked to identify and evaluate *training approaches* that have the potential to enhance pilots' ability to effectively monitor for FPM (with the result of improved awareness). The focus of this work is to identify, develop or validate training guidance to improve pilot monitoring/awareness regarding FPM and mitigate the recent trend of accidents and incidents, especially loss of control (LOC) events. The result of this work should be input for improved industry standards and FAA guidance to reduce the risk of incidents and accidents due to inadequate pilot monitoring/awareness. This is the first of three reports that were developed for this project.

Pilot Flying/Pilot Monitoring vs Monitoring Activity

World-wide, airlines typically refer to the two pilots of the flightcrew as pilot flying (PF) and pilot monitoring (PM). This distinction is meant to emphasize that the most important role of the pilot who is not flying is to monitor actively, especially regarding FPM. Chapter 6 of Advisory Circular 120-71B, titled, "Standard operating procedures and pilot monitoring duties for flight deck crewmembers," offers further guidance on defining the PM's role:

- The PM is responsible for monitoring the current and projected flight path and energy of the aircraft at all times.
- The PM supports the PF at all times, staying abreast of aircraft state and ATC instructions and clearances.
- The PM monitors aircraft state and system status, calls out any perceived or potential deviations from the intended flight path, and intervenes if necessary.

The current project focuses on providing input for training for the PM role of the flight crew, especially regarding FPM. While the aim of the work is to provide inputs to help train the role of the PM, it is understood that both the PM and PF roles are engaged in the activity of monitoring. Also, since each pilot may assume the PM role, effective training has the potential to be effective for enhancing monitoring for both roles and for each pilot.

1. Executive Summary

The following are the primary points to take away from this report.

1.1. Monitoring for Flight Path Management: What is it? How does it Fail?

1. *We have defined monitoring for flight path management (FPM) broadly* (Sections 2A and 3*)*. Our discussion of monitoring extends beyond looking and listening to gather information from the operational environment. Our definition of the scope of monitoring includes the following related activities:
    - Planning for managing operational tasks so that there are ample attentional resources for monitoring during critical flight phases.
    - Operational knowledge and situation understanding ("sense making") that determine what information is relevant, how attention is allocated, and how information is understood.
    - Switching attention to be responsive to emerging events but also ensure flight path targets are being met.
    - Communicating a shared understanding of FPM objectives to coordinate monitoring with the PF.
    - Identifying deviations from the expected state to ensure that they are called out and managed.

A more detailed account of what are central monitoring knowledge and skills and what are supporting skills is presented in Section 3.

2. *We have identified the ways in which failures in monitoring for FPM affect operations*. Specifically, we identified these failure categories, which are expressed as operational outcomes:
    - *Imminent upset*: The pilot loses awareness that a core flight parameter has transitioned toward an upset, e.g., airspeed is low, pitch attitude is high, roll attitude has gone beyond 35°.
    - *Failing to meet flight path targets*: The pilot fails to confirm compliance with the assigned (current and future) flight path targets. When monitoring fails while under manual control, the airplane may, for example, drift off heading or airspeed. Or a monitoring failure can lead to a failure to meet a restriction at an upcoming waypoint.
    - *Incorrect airplane state or configuration*: The pilot fails to see that the airplane is not configured for the current situation or an upcoming transition; for example, ensuring that the airplane is configured for take-off or that autoflight is armed for re-intercepting the flight path.
    - *Failing to identify important changes in the flight environment*: The pilot fails to remain aware of changes that are occurring in the larger airspace. Failing to monitor the flight environment can leave the flight crew scrambling to adapt or perhaps lead to a higher likelihood for an unstabilized approach.
    - *Failing to maintain awareness of crew resources*: Each pilot should be aware of what the other pilot is currently engaged with; they should keep each other apprised when they are shifting their attention to a task that will involve them for an extended period.

- *Failing to call out a deviation*: When there is a clear deviation away from an FPM target, the PM fails to call the PF's attention to that deviation so it can be corrected or managed.
- *Failing to intervene when a deviation is not being managed*: Also, the PM can fail to intervene in order to ensure that appropriate corrective actions are initiated.

A more detailed account of these failure categories is presented in Section 2.1.

1.2. Barriers to Effective Monitoring

3. *Monitoring for FPM can be incredibly demanding during highly dynamic flight phases* (Section 2.3). In addition to updating how well current and future flight path targets are being met, monitoring for FPM must also maintain awareness of changes in the larger operational environment and the activities of other flight crew members. And these activities must be balanced with demands to monitor airplane systems and perform a range of other operational tasks. It is not possible for any pilot to achieve complete awareness of the operational situation, especially during the more dynamic flight phases.

4. *There are many barriers—tied to attentional resources—that can reduce the ability to monitor effectively for long periods of time.* Section 2.3 offers a detailed description of the set of factors that impede or interfere with how well attention is allocated. In addition to the high workload mentioned in item #3 above, the following should be considered:
   - *Interruptions and distractions*: Task performance is not completely under the control of the flight crew, and interruptions and distractions can take attention away from monitoring.
   - *Vigilance failures*: Psychological research has established that sustaining attention—remaining vigilant—on a monitoring task is resource-intensive and stressful. Attention cannot be sustained over a long period of time (sometimes even as short as 15 minutes) without a considerable performance decrement.
   - *Mind-wandering*: Research on pilots has shown that they sometimes engage in mind-wandering, which is thinking about something other than the task at hand.
   - *Stress and fatigue*: These not-uncommon conditions during flight operations have been strongly linked to a narrowing of attention.
   - *Channelized attention*: In the extreme, attention can become "channelized," and the pilot can lose awareness of even salient or central cues.
   - *Inattention blindness*: A pilot may fail to see unexpected values or changes on the interface.
   - *System automation*: There is an extensive literature on the ways in which autoflight and flight management add complexity to monitoring for pilots. Automation does not reduce the need to monitor; indeed, it increases the need to monitor.

The conclusion from this combination of a rapidly changing operational environment and well-established limits on attention is that *sustained attention on FPM or complete awareness is not possible throughout a pilot's duty time* and exhortations to remain vigilant will not be effective as mitigation.

1.3. Monitoring is Sense-making, not Scanning

5. *Monitoring expertise should not be characterized as a scanning pattern*. While the research literature on eye-tracking (ET) shows that there are a few recognizable scanning patterns within the "basic T" indications when performing a manual approach, no one has identified a meaningful scanning pattern across the full flight deck interface. Appendix B provides a review of the relevant ET literature.

6. *Monitoring, essentially, is "sense making" (Section 3.1).* It is systematic observation and interpretation of the current state of the airplane and its operational environment; it requires integration of current inputs with operational knowledge, which includes mental models, and the generation of expected values of flight path targets.

Further, monitoring depends on and is enabled by these skills:

- Task management, which includes:
  - planning for managing operational tasks so that there are ample attentional resources for monitoring during critical flight phases
  - switching among tasks to be responsive to emerging events but also ensure flight path targets are being met
- Communicating information to and from the other pilot, which includes:
  - updating a shared understanding of current state and FPM objectives
  - communicating needed actions and future monitoring targets and expectations

7. *We describe the important role of operational knowledge and a situation model in driving monitoring for FPM* (Section 3.1). Operational knowledge refers to a range of topics, such as airplane systems and indications, autoflight modes and behavior, and expectations about airplane performance. This foundational knowledge, derived from training and experience, allows a skilled pilot to develop a thorough understanding of the current situation and fill in some of the information that has not been presented. It also identifies relationships between the various indications. Thus, operational knowledge leads to techniques for confirming an understanding of what is happening in FPM, what is referred to as information relevance. A skilled pilot understands which indications are most relevant to the current situation.

   More importantly, this knowledge, when integrated into a situation model, allows a skilled pilot to generate expectations about the airplane's energy and trajectory in relation to the flight path targets. The situation model, which is a rich mental representation of the current situation, also allows projection into the future to support prediction and planning. Those expectations determine what information is monitored in the environment and supports judgments about how well FPM is progressing.

8. *The ability to identify potential threats to FPM also can drive judgments about information relevance* (Section 3). Threat identification—e.g., a tail wind—further enriches understanding, as captured in the situation model. When a threat to FPM occurs, a skilled pilot quickly recognizes it and understands what it means for changes to managing the flight path.

9. *Monitoring is tied to expectations* (Section 3.1*).* Monitoring—looking and listening to gather information about the operational environment—is not simply moving one's eyes to each indication and reading it. Typically, but not always, there is an expectation about what that

indication will be, perhaps based on a projection from the last glance. When the expectation is violated—that is, the indication is quite different from the expected value—the pilot will be surprised and will engage in a sensemaking process. The objective of sensemaking is to restore the understanding of the airplane's current trajectory and energy relative to the flight path targets. These expectations are driven by operational knowledge and the situation model.

1.4. Monitoring is Enabled through Task Management and Communication

10. *Task management is controlled by both strategic and tactical processes* (Section 3.2*)*. Task management determines how attention is allocated to tasks. There is a strategic process that generates a plan for performing operational tasks from the situation model. This plan, however, is dynamic and is modified frequently. There is also a tactical process that is driven by two questions:

    - *Am I done with this task or do I need to continue attending to it?* Monitoring for FPM requires that attention be distributed to "keep all the balls in the air." It is rarely wise to remain focused on the same task for an extended period of time; especially, in more dynamic flight environments. When there is a clear stopping point on a task, it is easy to determine that it is done. However, some tasks may not be easily resolved.
    - *What task gets attention next?* When following the plan, the next task is clear, but when there are multiple demands for attention, such as pop-up demands, or when conditions assumed by the plan change, then task management becomes more dynamic and difficult.

    Thus, tactical control requires an attention switching mechanism, which is driven by an awareness of how attention is being allocated and awareness of changes in the situation model.

11. *Monitoring is held together through flight crew communication* (Section 3.3*)*. Communication supports understanding when FPM objectives and targets are identified and discussed to develop a shared understanding of relevant indications, expectations about how indications will change, and setting thresholds for calling out deviations. Communication supports task management when both pilots talk about how they are allocating attention, especially when they move away from monitoring for FPM.

1.5. Lessons from the Literature Review Regarding Training Monitoring

12. *We reviewed research literature in five domains on how to effectively train for monitoring and awareness* to try to identify training methods that could benefit pilots (Section 7). The domains were:
    - aviation
    - medical
    - air traffic management
    - driving
    - elite sports performance

    Each of these domains has something in common with the skill set required for aviation. However, the aviation, medical, and driving domains produced the most valuable insights.

13. *In the aviation domain (Section 7.1), we identified two research programs* that were particularly informative about potentially useful training methods at an airline level: Advanced Crew Resource Management (ACRM) (Holt, Boehm-Davis, & Hansberger, 2001) and Enhanced Safety through Situation Awareness Integration (ESSAI) (Hoermann, Banbury et al., 2003).

    - ACRM applied a Brief-Discuss-Advocate-Resolve (BDAR) sequence that shaped the interaction between the pilots. BDAR steps included actions supporting situation awareness, such as setting up a monitoring plan with expected verbalizations.
    - The ESSAI project developed resources for systematic approaches to training methods:
        - identification of concerns regarding situation awareness (SA) and of the important concepts underlying effectively maintaining good SA
        - development of a training program to teach the key elements identified, and measures to assess SA
        - execution of an evaluation study

    Each program developed realistic simulator scenarios for training and evaluating SA. A variety of additional studies contributed narrower data about training effectiveness.

14. *In the medical domain (primarily focused on surgery) (Section 7.2), training interventions focused on building understanding and sensemaking skills.* Researchers identified information that was most relevant to medical decision making and trained to improve noticing and use of that information to track critical, dynamic situations.

15. *In the driving domain (Section 7.4), there was a focus on hazard anticipation*. Again, the researchers identified information in the environment that drivers were not using effectively and creating techniques to train drivers to notice and project potential driving hazards.

16. *We described a promising training approach: Concepts-Briefing-Simulation-Debrief (CBSD).* Structuring training around activity cycles of teaching concepts and theory, providing a briefing, conducting a simulation, and doing a debrief (CBSD) is likely to be a powerful approach for organizing training. Improving the ability to build up a situation model and use it to understand the ongoing situation is a complicated learning goal. We think CBSD training cycles may be particularly effective for complex learning tasks such as this. Section 8 provides more details on this approach.

17. In terms of encouraging better communication around monitoring, a potentially useful training approach is to offer a structure for flight crew communication that covers:
    - current FPM objective(s)
    - current potential hazards or threats to FPM
    - relevant indications to monitor
    - expected behavior—e.g., for autoflight or airplane performance—and signs that things are not going as planned
    - trigger points, or the point at which the PM should ensure that the PF is aware of a deviation

1.6. Lessons from the Eye Tracking Literature

18. *The research literature on ET for commercial transport pilots offers a few lessons about monitoring.* Appendix B provides a review of this literature; here are the lessons we learned from that work:

- For common flight maneuvers, we have a general understanding of how visual attention is being allocated (in terms of percent time spent on each display).
- This research field has failed, in our opinion, to establish an ET-related description of monitoring expertise. While some studies find differences in ET behavior between skilled and less-skilled performers, these descriptions do not lead to a model of skilled performance.
- Studies that compare the PF and PM have identified strong differences in attention on the attitude direction indicator (ADI) but also find that, overall, the PM allocates visual attention in a manner similar to the PF. Ideally, the flightcrew should adopt a more complementary approach.

1.7. Report Organization

In addition to the Executive Summary (Section 1), the report is organized as follows:

Section 2. An account of *how and why monitoring for flight path management fails*. We describe what FPM is, how monitoring supports it, and the ways in which it can fail. We also describe the significant human limitations on attention.

Section 3. *A model of monitoring* that identifies important knowledge and skills underlying monitoring activities. We also compare our model to mathematical models of monitoring that have been used to predict monitoring behavior.

Section 4. *Explorations of skilled monitoring*. We review the eye-tracking literature to determine what can be learned by capturing visual attention with this method. For a different perspective of skilled monitoring, we also interviewed a number of pilots to analyze how several difficult monitoring situations might be managed.

Section 5. A brief historical perspective on *how pilots have been trained* to monitor.

Section 6. An overview of *how we identified and selected the relevant literature* on training monitoring and situation awareness.

Section 7. A literature review and summary of lessons to be drawn from five different domains regarding *effective methods for training monitoring* and situation awareness.

Section 8. Findings on essential knowledge and skills for monitoring and appropriate approaches to training.

In addition, this report includes two appendices:

Appendix A. A summary of previous reports that address monitoring and the training recommendations that came out of those reports.

Appendix B. A review of the literature on eye-tracking for commercial transport pilots.

2. Monitoring for Flight Path Management: How and Why it Fails

2.1. The Scope of Flight Path Management

Because the focus in this report is on monitoring to support FPM, we define FPM here to aid us in specifying how monitoring applies. According to a recommendation (16-4) from the Flight Path Management Working Group (FPMWG) to the Air Carrier Training Aviation Rulemaking Committee (ACT ARC), FPM is defined as "the planning, execution, and assurance of the guidance and control of aircraft trajectory and energy, in flight or on the ground." Similarly, according to the 2013 FAA report, "Operational use of flight path management systems" (which is called the FltDAWG report), the scope of FPM includes trajectory management and energy management.

For the present activity on monitoring for FPM, we focus on "execution and assurance," which refers to pilot activities used to achieve flight path targets and to keep the airplane operating within the flight envelope and away from external obstacles and threats. FPM also includes taxiing on the ground (although we do not address the ground portion in this report). We do not focus on advance, strategic planning of the flight path but do include feasibility assessments of ATC requests for revised flight path targets.

More specifically, for our analysis monitoring supports FPM in the following ways:

- Confirming that the airplane is within the flight envelope (i.e., the aircraft currently has appropriate airspeed, attitude, vertical speed, and energy to maintain a stable flight regime). This is monitoring relative to the expected flight envelope.
- Confirming that the current flight path indications—such as airspeed, altitude, or the flight director—are at the targets explicitly represented on the interface. For example, am I flying at the assigned airspeed?
- Confirming that the airplane and its systems are in the appropriate state/configuration in preparation for the current or upcoming segment of the flight path; autoflight modes are especially relevant here. For example, vertical navigation (VNAV) is engaged, approach (APP) is armed for approach, the mode control panel (MCP) altitude is set for the cleared altitude, the airplane is configured for the approach.
- Assessing the airplane's position, energy, and expected performance relative to FPM objectives. This can involve monitoring relative to targets that are implicit (not on the interface but in the pilot's mind, such as the 3-to-1 rule (i.e., planning 3000 ft of distance to descend 1000 feet)). For example, will I be able to make the altitude and airspeed restrictions at the waypoint? Or, will I be able to reach a stable airspeed in time to meet the stabilized approach criteria? Note that a skilled pilot is also able to determine when the airplane is unlikely to be able to comply with the flight path requirements and communicate this to ATC.
- Monitoring to obtain feedback on pilot actions, both on the flight controls and on flight deck system controls (e.g., airplane system inputs, such as mode selections).
- Monitoring the larger operational environment (e.g., scanning to detect and avoid weather, traffic and terrain hazards, monitoring radio traffic)
- Monitoring the activities and workload of the other member(s) of the flight crew to determine how well FPM is being supported.
- Detecting airplane flight path-related alerts, such as a low-airspeed alert.

For a skilled pilot, these basic monitoring activities are combined with control inputs to fly effectively and efficiently along the assigned flight path. Note that the role of monitoring for FPM, more broadly defined, also needs to include activities related to managing deviations; these will be further articulated in our model of monitoring (see Section 4).

2.2. Monitoring Failures related to FPM

Monitoring can fail to achieve all of these objectives, and frequently does, although most monitoring failures are minor and do not lead to an undesirable outcome. A failure in monitoring means that the pilot fails to be aware of and, therefore, to understand the airplane's state relative to the flight path (especially when there are deviations). Indeed, one driver for the current project are monitoring failures that have led to LOC accidents and incidents. The work by the Commercial Aviation Safety Team's (CAST, 2014) Airplane State Awareness (ASA) team focused on cases in which the pilot was not aware of a significant change to airspeed, energy, or attitude. Examples include the following:

- *Turkish 1951 (in 2009)* in which the flight crew was unaware that the airplane's airspeed decelerated significantly below the approach speed on approach and transitioned to a stick shaker event, and then, a crash with nine fatalities.
- *Ethiopian 409 (in 2010)* in which the airplane slowed down and pitched up. The flight crew seemed to be unaware of this for 27 seconds until the stick shaker came on. The resulting crash into the sea resulted in 90 fatalities.
- *Flash Airlines 604 (in 2004)* in which the pilot became distracted and allowed the airplane to roll 40° from where he meant to be. The Captain, who was PF, failed to manage this minor upset and continued rolling in the wrong direction. The aircraft crashed into the sea resulting in 148 fatalities.

Of course, there are also many less-tragic consequences of monitoring failures. After gathering examples of monitoring failures, we identified several general failure categories, which are expressed in terms of operational outcomes:

- *Imminent upset*, as represented by the CAST ASA accidents described above. In some cases, when monitoring fails, the pilot is unaware that core flight parameters have transitioned toward an upset, e.g., airspeed is low, pitch attitude is high, roll attitude has gone beyond 35°. When this failure occurs, the pilot is typically surprised that the airplane state has degraded to such an extent and appropriate actions need to be applied quickly to prevent further degradation into a more serious upset or LOC situation.

  One factor for the LOC cases is that the alerting system failed to make the flightcrew aware of a hazard (see Mumaw, Haworth, & Feary, 2019). In many LOC accidents, there was no alert for basic flight path hazards (e.g., low airspeed or bank angle), or the alert failed to get the pilot's attention, typically because it had no aural component.

  This type of monitoring failure, as the accidents show, can lead to a tragic outcome. Fortunately, these monitoring failures are more often caught and corrected. The CAST team looked at Aviation Safety Information Analysis and Sharing (ASIAS) data to determine how often these ASA-type situations occur in United States-based operations; they found evidence that numerous flights have had significant but short-term degradations—such as a stick shaker—that were recovered.

- *Failing to meet flight path targets*. A key element of FPM, as described above, is meeting the assigned flight path targets: level at the cleared altitude, fly the cleared airspeed or heading, stay on a glidepath, or cross a waypoint according to the altitude restriction in the flight plan. Monitoring is meant to confirm compliance with those assigned targets. When monitoring fails while under manual control, the airplane may drift off heading or airspeed. This activity is about "closing the loop"—using control actions to manage deviations from the target. When monitoring fails during autoflight, the failure is equivalent to the failure to confirm airplane state or configuration (see next item).

  Further, there is a strong interplay between control actions and the monitoring that determines whether the airplane will achieve the target. For example, if the flight plan has an altitude and airspeed restriction at a waypoint on descent, the pilot needs to determine what actions are required to descend and slow down to meet those restrictions. The pilot needs to project the airplane's progress toward those targets and then monitor whether that progress is as expected or requires additional inputs. A failure in this case can result in accident precursors, (e.g., a missed altitude).

- *Incorrect airplane state or configuration*. FPM is highly dynamic, and there are many times during a routine flight that the flight crew needs to configure the airplane for an upcoming transition (or ensure it is correct for the current flight phase). Examples are arming lateral navigation (LNAV) prior to intercepting the flight path, arming APP prior to intercepting the localizer and glideslope, ensuring that the correct ground signal has been tuned for an approach, ensuring that the vertical navigation path (VNAV PTH) mode is active, or ensuring that the airplane is configured for take-off. Monitoring is meant to confirm that the airplane is appropriately configured and has the correct targets. This assessment may require checking a spatially distributed set of indications, and, in some cases, it is guided by a procedure. When monitoring fails, the airplane proceeds in an inappropriate configuration, which can potentially lead to undesired outcomes.

  In one example of how this type of monitoring failure has led to an accident, the crew of Spanair 5022 failed to configure flaps for take-off and, without sufficient lift, crashed shortly after rotating away from the runway. For this flight, an airplane system failure removed the Take-off Configuration alert that could have prevented them from taking off. In one significant monitoring failure, the Spanair crew used the appropriate checklist but still failed to notice that the flaps were in the wrong setting.

- *Failing to identify important changes in the flight environment*. Another element of monitoring is to be sensitive to changes that are likely to occur. The flightcrew initially programs the flight management system (FMS) flight plan hours before the final phases of flight, and conditions at the destination airport—such as winds, visibility, storm activity, traffic, runway conditions—can change significantly. A skilled pilot can often anticipate and plan for likely flight plan changes. This planning can remove the need for FMS changes (e.g., programing an alternate runway) during a busy period or can create some space for adaptation to changing conditions and requirements.

  Failing to monitor the flight environment can leave the flight crew scrambling to adapt or perhaps lead to a higher likelihood for an unstabilized approach. The cues are often in the radio traffic, the changes to ATIS, or on the radar image. An example is radio traffic that reveals that many incoming flights are being asked to slow down shortly after top of descent. The pilot may notice that the current flight plan has a steep idle descent section to a high- altitude restriction. In this case, being slowed down could threaten the ability to make that altitude restriction. The flight crew might then request to start their descent early in order to be able to comply with the expected airspeed change and still make the altitude restriction.

- *Failing to maintain awareness of crew resources*. The typical (western-built) commercial transport has a flightcrew of two pilots. The flightcrew has many important tasks, and it is not possible for both pilots to always be engaged with FPM. However, ideally, one of these two pilots is always engaged with FPM. Each pilot should be aware of what the other pilot is currently engaged with; they should keep each other apprised when they are shifting their attention to a task that will involve them for an extended period; this is standard Crew Resource Management (CRM). When both pilots shift their attention away from FPM for an extended period, especially when they are in a more dynamic phase of flight, this is a failure in monitoring.

  There have been a number of accidents in which the entire flightcrew shifted attention away from FPM. Perhaps the most famous example is Eastern Airlines Flight 401 in 1972. In this case, three flightcrew members all became focused on troubleshooting a suspected landing gear extension failure. They had set the airplane in an altitude hold at a low altitude and then all three crewmembers focused on troubleshooting. During this time, one of the pilots inadvertently bumped the control column, knocking the autopilot out of the altitude hold mode, and the airplane started descending over several minutes, eventually crashing into the terrain.

From our discussions with pilots and airlines, we compiled specific cases for which monitoring can be difficult and may lead to a failure. We sorted these cases into the monitoring failure types introduced above:

- Incorrect airplane state or configuration:
  - Inappropriate autoflight mode selections; for example, the airplane is cleared to fly a heading to re-intercept the FMS path, but LNAV is not armed, and the airplane fails to intercept the path.
  - The vertical mode reverts to a vertical speed (VS) mode, and the airplane continues descending in VS, which may result in a missed FMS altitude restriction.
  - “Erroneous Glideslope.” There are situations where a malfunction, incorrect operating mode, or obstruction of the instrument landing system (ILS)equipment can lead to a "false" glideslope.
  - Incorrect barometric altimeter settings.
- Failing to meet flight path targets:
  - The “Slam Dunk.” The airplane is held at cruise altitude past the FMS-calculated top of descent point. If the airplane stays at that altitude too long, or if there is a strong tailwind, it may be difficult to meet a waypoint altitude crossing restriction on the flight plan.
  - “Short Approaches.” The airplane is assigned to fly direct to a waypoint that is at or close to the final approach fix (FAF). If this new clearance removes a

substantial number of track miles, it may be difficult to descend and slow down enough to meet the flight path objectives.

- Failing to identify important changes in the flight environment:
  - Late changes in tailwind or crosswind on arrival and approach.
  - Runway closures in degraded visual environments (e.g., at night), particularly at airports with multiple parallel runways

Other types of monitoring failures that are less central to the focus of this report include:

- *Failing to call out a deviation*. When a deviation from an FPM target occurs, the PM should call the PF's attention to that deviation so it can be corrected or managed. There have been a number of accidents in which the PM failed to inform, in a timely way, the PF about a significant flight path deviation (e.g., Asiana 214). Further, several airlines have told us that they still struggle with getting the PM to call out deviations in a timely way. Note that these failures may be tied to a failure to notice the deviation or a failure to call out a deviation that was noticed.
- *Failing to intervene when a deviation is not being managed*. After the PM calls out a deviation, the PF can still fail to correct or manage that deviation. In these cases, the PM needs to intervene in order to ensure that appropriate corrective actions are initiated. And, ideally, the intervention is coordinated; there have been cases in which the PM initiated control actions that were in conflict with the PF. For example, in the Flash Air 604 crash, both pilots were attempting to control the airplane as it dove into the sea.

The foregoing describes operational outcomes of failures in monitoring. However, our focus is on the psychological process underlying effective monitoring. In Section 3, we describe the full range of pilot skills and knowledge tied to monitoring for FPM. There are a number of "upstream" skills associated with attention management and situation assessment that can prevent the types of monitoring failures described here.

2.3. Human Limitations in Attention that can Reduce Effective Monitoring

As described in Section 3.1, task management is the allocation of attention to tasks or activities. Effective monitoring critically depends on effectively allocating attention and understanding human limitations in attention identifies potential threats to effective monitoring. During flight operations, there are multiple demands for attention and multiple tasks that must be executed. Potentially long duty times, rapidly changing conditions and demands, and unplanned events that compete for attention threaten effective task management. In this section, we briefly describe influences on attention and task management that may disrupt effective FPM during a flight; these are:

- high workload
- interruptions and distractions
- vigilance failures
- mind-wandering
- stress and fatigue
- channelized attention
- inattention blindness
- system automation

Monitoring is just one of the many tasks that are expected of the PM; others include:

- managing communications with ATC and company dispatch
- managing communications with the cabin crew
- running checklists at various points throughout the flight
- tracking and reporting flight progress
- managing any unexpected non-normal events or alerts
- updating the FMS, perhaps urgently, to reflect changes to the flight plan

Because of these various duties, the PM's workload can increase significantly during the flight, and there will be periods of time when the PM has no attentional resources for monitoring for FPM. Planning the distribution and timing of the expected tasks is a proactive part of task management.

Unexpected interruptions and distractions are even more difficult to manage. In their book on the demands for multi-tasking in real-world operations, Loukopoulos, Dismukes, & Barshi (2008) describe the range of assaults on the flight crew, which includes operational demands to keep moving, delays in getting flight information, changes to weather and, therefore, flight planning, disruptions due to passenger behavior or illness, and airplane system failures that need to be addressed. The potential list is long, and each item can require the flightcrew to stop what they are doing and attend to some emerging need.

The role of distractions in reducing awareness was prominent in the CAST ASA safety event analysis. In each of those 18 safety events, there was something that pulled flightcrew attention away from FPM, leading to a loss of energy or attitude awareness. In the example of Turkish Airways 1951, a radio altimeter failure generated unexpected airplane behavior during the approach. While the airplane was descending and slowing down, the two pilots started performing the Landing checklist. However, they were surprised by indications on the speedbrake system, which indicated both "speedbrake armed" and "speedbrake do not arm" at the same time. The cockpit voice recorder (CVR) revealed that their attention was on sorting out those incompatible indications and trying to complete the checklist as their airspeed decreased below their target airspeed, leading to a stick shaker and loss of control. In this case, the crew's limited autopilot understanding contributed to the distraction and inappropriate task management.

Another factor in limiting monitoring performance is the length of time on the job. Although shorter flights may only be in the range of an hour or two, the pilot's duty time in the United States can range up to 14 hours and be stretched across several short-haul flights. Other pilots will fly a single, much longer flight. Thus, a pilot can be required to be engaged with monitoring for FPM over many hours in a day. Psychological research has established (e.g., Warm et al., 2008) that sustaining attention—remaining vigilant—on a monitoring task is resource-intensive and stressful; that is, attention cannot be sustained over a long period of time without a considerable performance decrement. Warm et al. show that even 15 minutes on a sustained attention task can produce a performance decrement. Note that this performance decrement is not a product of distractions or interruptions; this research looked at situations in which the operator was uninterrupted.

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 the task of FPM-related monitoring. Mind-wandering is diverse; pilots will think about an upcoming vacation, a family situation, or just something that is not the task at hand. Thus, even when there are no other operational demands that disrupt what the PM is trying to do, pilots are not likely to be fully engaged in monitoring throughout the flight. Casner and Schooler did suggest that pilots may impose some control over when they engage in mind-wandering and may be able to suppress it when they anticipate a short, important period for monitoring.

Stress and fatigue, which are not uncommon conditions during flight operations, have been strongly linked 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).

In the extreme, attention can become “channelized,” and the pilot can lose awareness of even salient or central cues. There have been a number of airplane accidents (e.g., Tatarstan 363) in which the pilot, likely confused by 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 10 seconds or more, without a change to pilot nose-down control inputs. The pilot is seemingly unaware of the alert and the impending collision, likely due to “channelized” attention.

Another potential contributor, called “inattentional blindness” (Simons & Chabris, 1999), is a phenomenon that can occur when attention is strongly driven by a task-focus or from expectations about what information is present (i.e., confirmation or expectation bias). In this case, a pilot may fail to see unexpected values or changes on the interface, e.g., a visual check to confirm that the flaps are set as expected can fail to see that the flap setting is not correct, as we saw in the Spanair 5022 example. Thus, the pilot may not have awareness of information in the environment, even when it is salient or in the central field of vision.

A very different factor that can complicate monitoring is the degree and nature of automation in the airplane. There is an extensive literature on the ways in which autoflight and flight management add complexity to monitoring for pilots (e.g., Bainbridge, 1983). It does not reduce the need to monitor. Indeed, monitoring the autoflight system is typically difficult because pilots need to gather scattered indications that aid in determining its current state and behavior and use knowledge about how a specific mode will behave in the current situation. Also, in some cases, important mode information is not available on the flight deck interface.

In summary, there are various reasons for attention to be pulled away from FPM or for awareness to be diminished: attention can be narrowed and captured; there are limits on how long it can be sustained or focused on FPM; expectations can affect perception of what data are actually present, and information can be scattered and incomplete. We believe that the conclusion to be drawn from this long list of barriers is that *sustained attention on FPM or complete awareness is not possible throughout a pilot’s duty time* and exhortations to remain vigilant will not be effective as mitigation. While it might be possible to increase pilot motivation to engage with monitoring for FPM over short periods of time, we believe that enhancing foundational monitoring skills, which should include task management strategies, is a more appropriate path to better awareness.

3. Model of Monitoring for FPM

In Section 2, we described monitoring for FPM in the language of operational tasks. To understand the types of training that can potentially improve monitoring, we use this section to describe monitoring in psychological terms to capture underlying knowledge and skills (see also Billman et al. [2020] for a more-detailed description of monitoring and its supporting activities). The knowledge and skills identified in this section describe a high level of monitoring performance; there will be large individual differences across pilots.

Two important concepts are foundational to our model of monitoring: the situation model (and its cycle) and task management.

3.1. Situation Model and the Sensemaking Cycle

3.1.1 The Situation Model

Effective monitoring contributes to the pilot's understanding of the ongoing situation regarding the airplane state and its operational environment; this resulting understanding can be called a situation model. Specifically, a situation model represents the pilot's understanding of the current dynamic situation. This situation model will have different levels of detail and accuracy, depending on the pilot. A goal of training is to improve a pilot's ability to develop an effective, accurate situation model.

A mental model, which also refers to "what's in the head," is different from the situation model. A mental model captures general knowledge about "how things work," such as weather patterns or the airplane's autoflight system, while a situation model captures specifics about the current moment and projects them into the near future.

A situation model is built by selecting one (or more) mental models from memory and adapting them to the current situation. Knowledge from memory is either a mental model or operational experience, and relevant knowledge is integrated with data from the world. There are three types of inputs to the situation model (see also Hutchins, Holland, & Norman, 1985).

- *Data and information from the world*. Data and information are acquired through monitoring activities. For example, looking at the primary flight display (PFD) can update your awareness of the current airspeed, altitude, or attitude; or listening to radio chatter can update awareness of the conditions at your destination airport. Information might also include the control actions taken by the PF or the agreed-upon task allocations for the remainder of the arrival.
- *Relevant mental models*. The collections of facts and rules that help a pilot understand and anticipate the behavior of a system or device. A pilot will have multiple mental models and will draw from them to inform the current situation because the current situation links to the relevant knowledge. For example, the situation model may draw on general knowledge of VNAV descent or energy management, how to slow down during descent, and procedures for communicating with ATC. In turn, the pilot's experience can lead to mental model revisions when new information or relationships are discovered.
- *Knowledge the pilot has from operational experience*. This knowledge is more experiential and less of an integrated understanding of some system or device. For example, the pilot landed at the same airport a week ago and is aware that there is a confusing turn on the taxiway due to a missing sign.

3.1.2. The Sensemaking Cycle

Our model of monitoring is strongly tied to the process of understanding, or making sense of, the ongoing situation. Mental models of familiar systems and situations provide a ready-built framework and accompanying expectations that can be used to guide monitoring. For example, a mental model might identify variables whose current values are important to know. It might also set expectations about sensible values of these variables and how they normally change.

The process of monitoring can be depicted as a cycle that updates the situation model and is also guided by the model. This cycle, shown in Figure 1, has three key processes that connect with the situation model at multiple points:

- Identify gaps in understanding.
- Gather relevant data and information.
- Identify appropriate actions.

Figure 1 also shows the relationship, at a high level, between the situation model and monitoring activities. The cycle begins when the pilot identifies a gap, inconsistency, or puzzling aspect in the situation model, perhaps due to the need to update a key flight parameter or due to a new operational demand. Other prompts might be a potential threat to FPM, such as being held high during an approach, or a reminder to initiate a checklist. Further, the pilot needs to set priorities by identifying and focusing on what is the most important need at the moment.

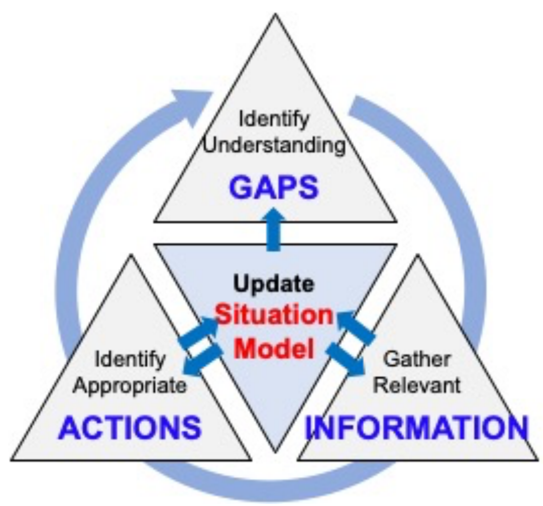

Figure 1. Situation model and sensemaking cycle.

The situation model is used to identify the data and information—e.g., flight deck indications—for addressing those gaps. The situation model identifies which information is relevant, determines where that information can be found, and, in many cases, generates expectations about the current value. Information is not simply gathered, but it serves as evidence evaluated for how it can reduce the gaps in understanding. In turn, that information and the knowledge that comes with it update the situation model. Looking for information to address a question or a gap is a "top-down" process, guided by the pilot's goals. In addition, information may be spontaneously noticed, as when an alert is detected.

Finally, the situation model identifies what actions are needed for FPM, such as increasing speed or adding drag with the speedbrakes. FPM is a dynamic task and there is likely to be variability around the flight path targets, e.g., airspeed fluctuations. One of the challenges of monitoring is separating the normally occurring variability from a true deviation that is not being resolved. This is an example of problem identification, and it has two elements:

- The ability to identify a true deviation from normal variation.
- The ability to determine that the current situation is different from what it was believed to be, or re-frame; for example, realizing that airspeed is continuing to decrease because the autothrottle mode is no longer managing airspeed.

Taking action sets expectations about what will change, and when taken by a pilot, they change the state of the world. Again, monitoring the new situation updates the situation model, and observations may be compared with expectations.

An important characteristic of this cycle is that it processes iteratively: the situation model is repeatedly updated as new data and information arrive through monitoring, and the understanding captured in the situation model further drives the need to address whatever gaps or discrepancies may arise.

The larger theme is that monitoring is driven by developing an understanding of airplane state and its operational environment. And, from the reverse perspective, monitoring occurs in an operational context. That is, monitoring is not merely looking at indications; looking and interpreting are guided by and informed by expectations of the current understanding.

Figure 2 provides a more-detailed view of what is involved in monitoring for FPM. On the left is a set of monitoring activities that are specific to FPM. On the right of the situation model are the outputs that shape the monitoring on the left.

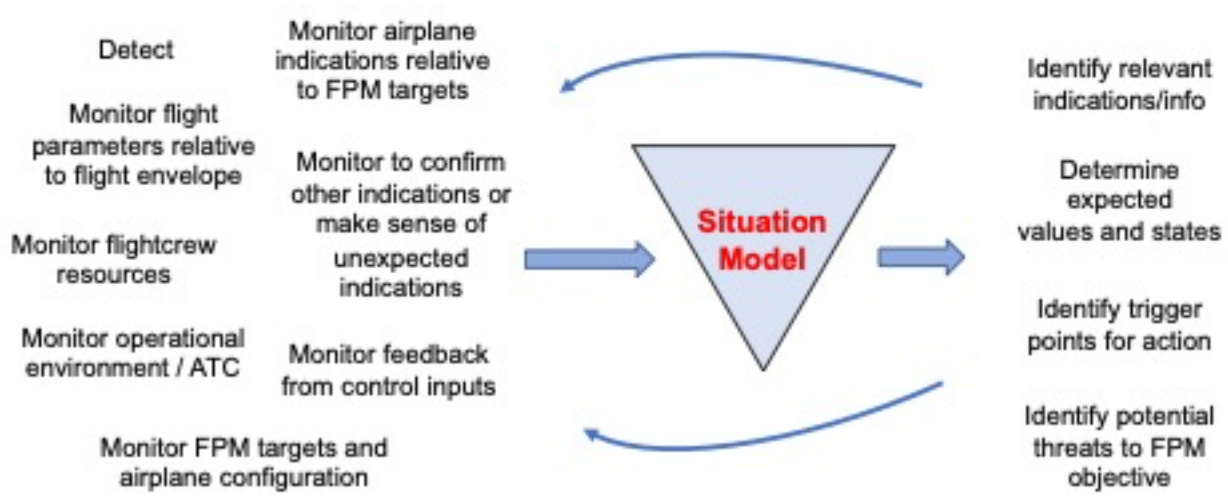

Figure 2. The ways in which the situation model influences data gathering.

Knowledge and experience not only inform the content of the situation model, but they also give guidance about how to monitor. The following is a breakout of the types of information-gathering used to address different questions. We believe these monitoring topics, collected largely from interviews with pilots, provide a fairly complete accounting of data and information gathering for FPM (and is also compatible with Chapter 6 of Advisory Circular 120-71B):

- *Monitor flight parameters relative to the flight envelope*. Confirm that the airplane currently has appropriate airspeed, attitude, vertical speed, and energy to maintain a flight regime, i.e., not entering an upset.
- *Monitor airplane indications relative to FPM targets*. Assess the airplane's position, energy, and expected performance relative to FPM objectives. FPM targets can either be explicitly represented on the flight deck interface, such as an airspeed bug, or can be implicit targets (not on the interface but in the pilot's mind). For example, am I meeting the stabilized approach criteria?
- *Monitor FPM targets and airplane configuration (including autoflight).* Confirm that the airplane and its systems are in the appropriate state/condition in preparation for the current or upcoming segment of the flight path. For example, APP is armed for approach, the MCP altitude is set for the cleared altitude, the airplane is configured for the approach.
- *Monitor feedback from control inputs*. Monitor to receive feedback on pilot actions, both on the flight controls and on flight deck system controls (e.g., airplane system inputs, such as mode selections).
- *Monitor the operational environment/ATC*. Monitor the larger operational environment through looking out the window, listening to radio traffic, and looking at flight deck interface displays regarding terrain, obstacles, weather, and traffic.
- *Monitor flightcrew resources*. Monitor the activities and workload of the other member(s) of the flight crew to determine how well FPM is being supported or what resources are available, if needed.
- *Monitor to confirm other indications or make sense of unexpected indications* .In many cases, monitoring is used to restore coherence or give confidence to the situation model. A pilot may be surprised to get a bank angle alert and will need to monitor to attempt to understand why the airplane is in a different state than was expected.
- *Detect*. Data gathering is, in some cases, not because the pilot intentionally allocated attention to a specific area of the interface. Information is "pushed" to the pilot, from an alert or from a communication from another person.

These activities update the situation model, and, in turn, the situation model influences monitoring for FPM in at least four ways, as depicted in Figure 2:

- *Relevance*. It determines which airplane and operational information is most relevant to monitor at this moment. The operational knowledge in the situation model (generated from the mental models) specifies which indications or information you should be monitoring, from an understanding of how the indications are related. A simple example is to monitor thrust and airspeed, in addition to altitude, when climbing since airspeed will need to be maintained with more thrust during a climb.
- *Expectations*. The situation model also generates expectations about indication values, autopilot modes, airplane performance, position relative to the flight path, and many other aspects of FPM. This is a ***critical*** contribution of the situation model because monitoring relies largely on comparing the current state to the expected

state. That is, monitoring is not merely looking at and determining the current pitch attitude and power setting. A skilled pilot will compare current pitch and power values to what is expected in the current flight phase to determine if FPM is progressing as expected.

- *Trigger points*. The situation model provides guidance about decision points for control actions or for shifts in what needs to be monitored. For example, by developing a projection of the flight path tied to position and energy state, a skilled pilot can determine the latest point on the current trajectory when it is possible to descend for the waypoint restriction.
- *Threats*. The situation model also taps into knowledge about potential threats to FPM. Certain conditions or events are recognized as being a potential threat to maintaining FPM. For example, an unanticipated strong tailwind near the top of descent should prompt FPM-related monitoring relevant to maintaining the FMS-determined flight path. Or, if ATC asks for a 'direct to' a waypoint that will remove a number of track miles, the flight crew should begin assessing their ability to descend and slow down with the shortened path.

3.1.3. Elements of Monitoring

This monitoring process, captured in Figure 1, develops through the acquisition of multiple types of knowledge and skill that work together to support effective monitoring. We characterize these skills and knowledge as the following.

Operational knowledge (and mental models): Know how to manage trajectory and energy in the airspace, and what you can learn from the interface.

Background knowledge is critical to effective monitoring. As we said above, monitoring is not merely looking at an indication; it is, in most cases, comparing an indication with an expectation of what it should be. For example:

- Why hasn't airspeed decreased as much as I expected it would at this point?
- Why am I in VS mode?

Accurate expectations come from an understanding of the airplane, its systems (especially the autoflight system), the operational environment, and an understanding of the current state. Skilled pilots understand airplane performance, such as, how quickly the airplane can lose airspeed on descent. They understand the autoflight system and what modes are typical throughout the flight. They understand weather and how it can affect operations. They understand the geometry of an FMS-generated flight path and how the airplane uses thrust and pitch to fly that path. They also have knowledge of the interface and its indications.

For certain types of knowledge, we use the term "mental model," which is a collection of interrelated facts and rules that help a pilot understand and anticipate the behavior of a system or device. Mental models may vary in their detail and accuracy. A pilot has mental models for a variety of systems: the airplane hydraulic system, autoflight, weather, the airspace, aerodynamics, and many others. These mental models can help a pilot anticipate how weather or wind will change; how the autoflight system will meet an airspeed target; or how traffic flows into a large airport during the busiest time of day. Although mental models are typically incomplete and imperfect, they allow a pilot to reason about FPM and the constraints on it.

Situation model: Understand the airplane's current state relative to the operational environment.

A pilot also needs to have an understanding of the current state of the airplane and the operational environment around it. Whereas a mental model is used to capture general knowledge about weather or airplane systems, the situation model captures weather and airplane system specifics in the current moment and projects it into the near future. A skilled pilot knows whether he/she is on the path or above the path or is too fast to slow down for the next waypoint restriction. These descriptions capture what is true at a point in time. Although the various mental models are used to inform the situation model, it is the situation model that allows the pilot to generate expectations about the current indication or state.

It is also important to note that the flight deck interface presents an impoverished view of the airplane's situation. The interface provides individual elements: current airspeed or heading, next waypoint of the flight path, or the vertical mode. There are some displays, such as a Navigation Display or Vertical Situation Display, that provide a more integrated depiction. However, the pilot's situation model can provide a richer, more three- (or four-) dimensional representation of the airplane's current and expected energy and trajectory relative to the flight path.

The situation model is dynamic. In one sense, it is being updated continuously as new data and information are gathered. In another sense, the situation model is used to generate expectations and projections into the future. More specifically, the situation model is continuously updated as new data and information are gathered, new comparisons are made, and needed actions are identified. Also, it continuously guides the process of monitoring by generating expectations, providing guidance on where to locate information, or project what control actions may be needed.

Efficient information extraction: Rapidly capture operationally relevant information.

A basic process of monitoring is to efficiently comprehend relevant data and information in the operational environment, e.g., flight deck indications, paper charts, visual and aural alerts, and spoken communications. The skilled pilot has a familiarity with the arrangement and format of the flight deck displays, indications, and other materials and is able to efficiently locate and comprehend data and information.

Interface configuration: Set up the flightdeck interface to facilitate monitoring.

Monitoring can be facilitated by configuring the interface (see Vicente et al., 2001 for a process-control example). In newer airplanes, especially, information is layered and there are display modes and display hierarchies across the interface. Accessing the most useful information can mean having the DATA mode on the NAV Display. An understanding of what indications or information is most relevant and where those data reside allows a pilot to configure the interface for gathering information more efficiently.

Threat identification: Understand the FPM threats that need to be managed.

FPM is often straightforward, e.g., maintaining altitude and speed during cruise. However, at various points in the flight profile, FPM can become more complicated due to unexpected changes from ATC or from related "threats" to maintain the cleared flight path. A skilled pilot can, in many cases, anticipate the threats to FPM and take action to manage them. Examples of potential complexities concerning FPM are:

- transitions from full (VNAV/LNAV) autoflight to more tactical control or even manual control
- changing wind and weather that can influence airplane energy
- a very late runway change on the arrival

A proactive pilot will look for these types of threats as they relate to FPM, asking, "Are there threats to continuing to meet my FPM objectives?". This judgment is highly dependent on having a well-fleshed-out situation model.

3.2. Task Management

In addition to being driven by activities to enhance understanding, monitoring is also regulated through task management, which controls where attention is being allocated. Attention is limited to focus on a single activity (Moray, 1986), and there are many demands for the flightcrew's attention in addition to monitoring for FPM. Pilots need to switch attention between various operational tasks to manage them all effectively. Thus, monitoring activities must be appropriately prioritized and coordinated with other tasks.

Task management has elements of planning and deliberate allocation of attention to specific operational tasks, but attention allocation can also be a product of in-the-moment shifts between tasks. Holder et al. (2016) define task management this way: "Task management is a dynamic process that involves both strategic and tactical organization of pilot tasks over the course of a flight. Strategic activities involve proactive planning, task prioritization, task allocation, task resource management, managing the timing of tasks, and anticipating and assessing the flight situation. Tactical activities are to monitor and respond to real-time changes in the flight situation and include monitoring task performance, reprioritization, reallocation, making decisions, and managing emergent events and disruptions."

In Figures 3–5 we illustrate three ideas about task management:

- The strategic element from Holder et al. or planning for deliberate allocation of attention to a sequence of operational tasks, which will include monitoring for FPM; this planning is driven by the situation model. That is, the situation model identifies what questions need to be answered and what information needs to be gathered, which specifies a prioritization of tasks.
- The role of an attention switcher that controls where attention is directed. This is the tactical element from Holder et al. Attention switching is important because of the potential other demands on attention, including unexpected operational demands as well as non-operational demands, which can include distractions, interruptions and mind-wandering. These other demands can force a need to re-plan.

Figure 3 shows a sequence of planned operational tasks stretching forward in time. The situation model drives the planning of this set of operational tasks. Recall that monitoring is only a subset of the tasks to be performed.

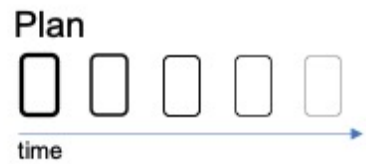

Figure 3. A plan for the set of tasks to be performed.

Figure 4 illustrates an array of other activities that compete for attention, namely re-planning, mind wandering, unexpected operational demands, distractions and interruptions. These last three are externally driven; that is, the pilot notices some event and turns attention to it. Re-planning and mind wandering are internally driven activities.

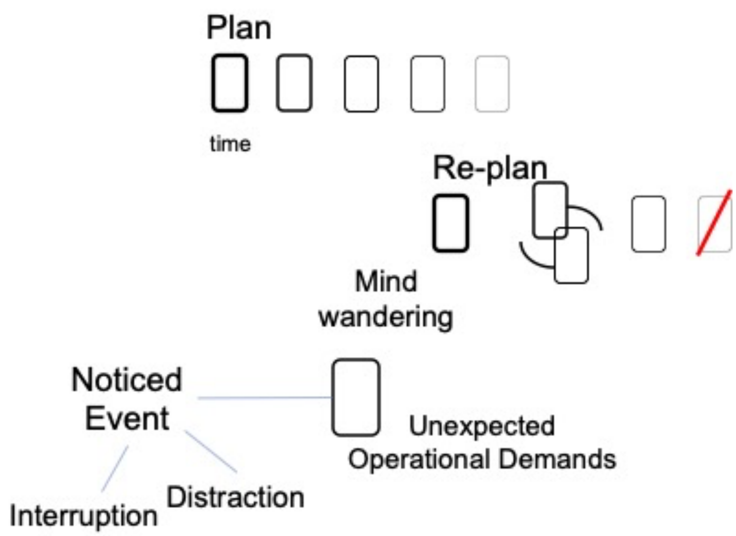

Figure 4. The larger set of activities that need to be managed.

Finally, in Figure 5 these come together with the concept of an attention switcher that must be responsive to changes in the operational environment about which task should be receiving attention. Attention is on a current task, and two questions are raised:

- *Am I done with this task or do I need to continue attending to it?* Monitoring for FPM requires that attention be distributed to "keep all the balls in the air." It is rarely wise to remain focused on the same task for an extended period of time; especially, in more dynamic flight environments. When there is a clear stopping point on a task, it is easy to determine that it is done, e.g., setting the cleared altitude on the MCP. However, some tasks may not be easily resolved, especially when trying to deal with an unexpected problem, such as making a complicated input to the flight plan. The question can also be overtaken; that is, in some cases, a new task pops up and demands attention, whether the current task is done or not.

• *What task gets attention next?* When following the plan, the next task is clear, but when there are multiple demands for attention, task management becomes more dynamic and difficult. The green “selector knob” is the metaphor for switching attention to a new task or activity. The situation model often plays a role in determining which task has highest priority, but task management is not always well thought out. An interruption can demand immediate attention.

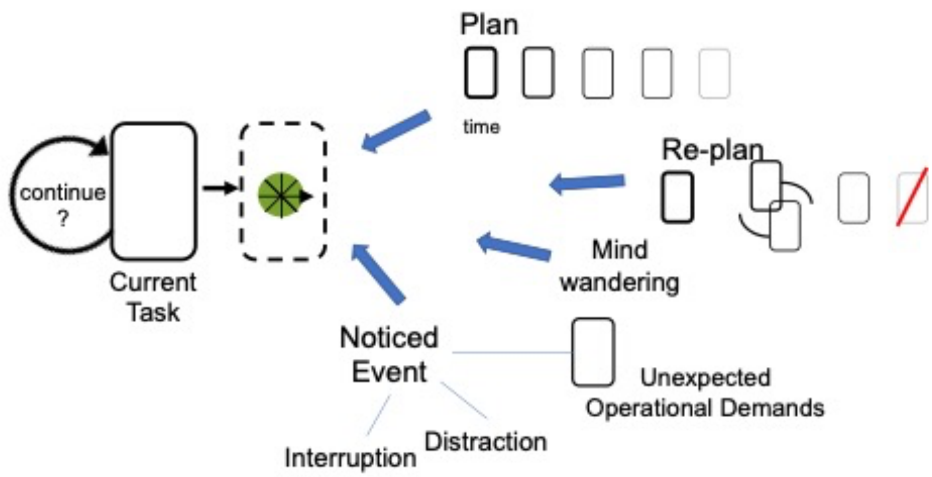

Figure 5. Schematic depiction of task management processes.

Generally, task management or attention switching needs to strike a balance between the need to stay focused on a single task to be able to complete it and the need to maintain an awareness of all relevant changes during a flight. Monitoring and the resulting changes to the situation model play a large role in this balancing; the skilled pilot adapts task management in real time in response to the evolving situation, but there are also limits on attention switching.

In particular, attention switching requires maintaining an awareness of how attention is being allocated and remaining responsive to changes in the situation model. Awareness of how cognition is operating is a meta-cognitive skill; and in this case, refers to attending to how your attention is being allocated. Meta-cognition is notoriously difficult to achieve, particularly in high workload or dynamic environments. This awareness can fail, however, and a concern here is failing to switch away from one task (perhaps some operational task, perhaps a non-operational distraction) to ensure that other important tasks are being managed at an appropriate frequency.

3.2.1. Enablers of Monitoring

As before, we identify the types of knowledge and skill that work together to support effective monitoring. We characterize these skills and knowledge as the following.

Strategic Task Management. Manage operational tasks (and, therefore, workload) through planning to ensure there are ample resources for monitoring.

In Section 2.3, we described how high workload, along with other barriers, can threaten the attentional resources needed for monitoring. In some phases of flight, such as descent, both pilots can become overwhelmed by non-monitoring activities that eliminate their ability to maintain awareness of basic flight indications. Indeed, even in routine situations, both pilots can allow themselves to shift their focus away from FPM-related tasks. However, skilled flight crews anticipate periods when FPM has high importance and shift non-monitoring tasks out of that period to the extent possible.

Attention Switching. Ensure that periodic monitoring for FPM does not lapse.

As we said, strategic task management can help ensure that attentional resources are not overly committed to non-monitoring tasks during critical periods of the flight. However, unanticipated tasks, interruptions, and distractions can also take attention away from monitoring. A skilled flightcrew would, ideally, be aware that they were neglecting monitoring for FPM and find a way to allocate attention back to FPM. Effective responses to short-term attentional shifts require awareness of how you are allocating your attention.

3.3. Broader Scope of Monitoring

In the previous two sections, we described the importance of sensemaking and task management in monitoring for FPM, and we described the relevant knowledge and skills that underlie monitoring. It is worth emphasizing that monitoring for FPM is not just about gathering data and information. More broadly, monitoring supports the Pilot Monitoring in determining whether FPM objectives are being met and then intervening when objectives are not being met. This more broadly defined scope for monitoring depends on other activities:

- task management (as described above)
- communication
- action implementation

If we turn to considering the role of PM, and not just the process of monitoring per se, the PM needs skills in task management, in communication, and in intervention and implementation of flight control actions.

3.3.1. Enablers of Monitoring

Therefore, the following are additional types of knowledge and skill that work together to support effective monitoring.

Communication: Develop a shared understanding of FPM objectives and threats.

Communication between flightcrew members is foundational for enabling skilled monitoring. As the flight path is defined and revised and performed, the PF and PM should be striving for a shared understanding of their FPM objectives and the potential threats or challenges to FPM. This communication aids in the following ways:

- Ensures that both pilots, PF and PM, understand and agree on short-term and/or long-terms FPM goals.
- Ensures that potential FPM threats are identified and understood in terms of how they will be managed.
- Establishes criteria (perhaps a threshold) for PM call-outs of deviations.
- Supports each pilot's role in maintaining awareness of the other pilot's ability to stay engaged with monitoring.

- Aids in coordination when there is a need to initiate a correction to a deviation or take other control actions.

Intervention: Take necessary actions to ensure that FPM targets are achieved.

When monitoring identifies a problem, there is a need to determine that the fight control actions being taken are sufficient, e.g., to resolve a deviation from a flight path target. It is central to the role of the PM to identify the threshold between normal variability and a deviation that needs to be addressed. The skilled pilot will have strong expectations about the degree of variability that is likely to occur and the level of variability that should get more-careful attention. Communication between PM and PF is frequently sufficient for the PF to take needed control actions and manage deviations or other concerns. However, if a critical problem is not being managed, the PM needs to intervene in some more forceful way. In some cases, this means increasingly assertive communication to ensure that the PF is taking the appropriate actions. In the worst cases, the PM may need to intervene on the controls. Indeed, the CAST ASA accidents and incidents provide examples in which the PM grabbed the controls to prevent an accident—but also failures in which the PM, although well aware of a worsening problem, failed to intervene and stop the crash.

3.4. Mathematical Models of Monitoring

The utility of the monitoring model described above is to characterize the range of flightcrew knowledge and skills related to monitoring for FPM. There is another approach to modeling that offers a different perspective on what influences monitoring: mathematical models.

Mathematical models were initially developed in the 1960s (e.g., Carbonell, 1966) to predict the performance of a human operator performing a control task such as would be encountered in process control (e.g., a nuclear power plant control room); more specifically, modeling a human who is monitoring a small set of indicators for deviations. The modeling attempted to predict eye-tracking behavior based largely on the characteristics of the display elements (e.g., Senders, 1983). The rate of change of an indicator (called bandwidth) proved to be a good predictor of how frequently a human fixated it. Some models also took into account the "cost" of looking away from an indication, where cost refers to the momentary loss of awareness of an indication that is important for the operator's task.

While these models were effective at predicting eye-fixation patterns, we believe that the monitoring task they were modeling is not a good match to the job of an airline pilot. Certainly, there are periods of performance when an airline pilot is making manual inputs to the flight controls responding to deviations in the "basic T" instruments, such as flying a manual approach. However, because of the larger set of displays to be monitored and the use of the autopilot for most flight phases, airline pilot monitoring is typically not bounded by a small set of dynamic indicators. Also, unlike more-limited laboratory tasks, airline pilots must attend to various other demands outside of the flight deck interface.

In an updated version of the early mathematical models of monitoring, Wickens (e.g., Wickens et al., 2003) expanded the factors considered by the model and applied his model to a more realistic aviation context. Wickens proposed four factors that drive monitoring behavior:

- The **S**alience of the indication.
- The pilot's **E**xpectancy of the indication, which is the same as the bandwidth concept (that is, a measure of the rate of change).

- The **E**ffort needed to shift attention to an indication.
- The task-relevance or **V**alue of the indication (which can capture cost).

Wickens uses the acronym SEEV to represent these four factors. The SEEV model, like the older models, is able to make fairly accurate predictions of eye-tracking behavior.

In the modern commercial jet transport, there is a strong link between two of the SEEV factors: expectancy and effort. That is, for the flight deck interface, the indications with the highest rate of change are largely also the “basic T” indications that require the least effort to monitor since they are directly in front of each pilot. Regarding salience, the flight deck interface does not have strong changes in salience across displays or over time, where salience is determined by bright colors, flashing, or other rapid visual change. There are a few alerting schemes that use flashing and color to increase visual salience, but these occur infrequently.

The remaining factor is Value, and in our model, we identified value, as derived from the situation model, as the most critical driver of monitoring performance. Note, also, that Value is the only SEEV factor that is determined by the human operator; the other factors are characteristics of the interface or of aircraft dynamics (in the case of expectancy). Our model describes the role of monitoring (attention) in understanding the current and future states of the airplane, which, in turn, supports performing operational tasks. From this perspective, the objective of monitoring is to update and fill in information that is essential to that understanding. Other research that is more tied to how attention is used in comprehension and task performance also emphasize the role of understanding in monitoring (e.g., see Rothkopf et al., 2007).

In summary, although the mathematical models capture monitoring behavior to a substantial degree by focusing on “bandwidth,” they do little to articulate “value,” which, we believe, plays a large role in monitoring for FPM. Value comes from the pilot, and, therefore, is the one element that can be addressed by training. A complete specification of value should, therefore, identify appropriate objectives for training.

4. Explorations of Skilled Monitoring

A monitoring model allows us to identify the range of knowledge and skills involved in monitoring for FPM. Further work is required to determine which of these elements separate highly skilled performers from low-skill or average performers. That is, some knowledge or skill may be acquired by the vast majority of pilots very early in training while other knowledge or skill may be acquired only by the most adept or experienced pilots. The role of training is to identify discriminators and aid pilots in developing the knowledge or skill that is important and that they are unlikely to have developed on their own. In this section, we look at two different approaches for describing skilled monitoring: eye-tracking and cognitive task analysis.

4.1. Eye-tracking Studies of Pilot Monitoring

To determine if we could gain insights about skilled monitoring, we reviewed the literature on ET, which is the tracking of eye fixations in the surrounding environment. This approach to studying attention allocation has been used to study pilot monitoring and performance for more than 60 years. Measures of ET can reveal where experienced/skilled pilots look—how their visual attention is allocated. Ideally, this knowledge can, in turn, identify gaps for training less-experienced/less-skilled pilots.

Appendix B provides a review of the relevant literature with the objective of determining what is known about how commercial transport pilots monitor. That review describes specific ET measures and descriptions of what is known about where pilots look. A few studies also attempted to identify how ET measures discriminated between the more-skilled and less-skilled pilots, with a focus on percent dwell time (%DT). That is, are some pilots dedicating more visual attention to indications or displays that are particularly relevant to the task they are being asked to perform? Generally, these studies of how visual attention is allocated show considerable variability across pilots. The average %DT across pilots hides the range of variability within the pilots in a single study, and in some cases, individual pilots differ by a factor of two or more.

In one study, Lefrancois et al. (2016) measured fixations for twenty A320 pilots during a manual approach. They separated out the four pilots who performed best on the approach (in terms of flight error) and then identified differences between those pilots and the other 16 in terms of %DT. [Note that data from individual pilots is often not reported in these studies.] The four best-performing pilots (averaged as a group) spent significantly more time on attitude, the localizer indication, and the glideslope indication.

In another study, Reynal et al. (2017) had 10 737NG pilots fly an approach using autoflight. They influenced the difficulty of the approach, forcing pilots into a situation that required a go-around. However, five of the pilots inappropriately continued the approach. When Reynal et al. compared the two groups (land vs go-around), they found that, unlike in Lefrancois et al. (2016), a higher %DT for attitude was associated with significantly worse performance. The "land" pilots also had a significantly lower %DT for the Navigation display.

Dehais et al. (2017) had 12 flight crews perform an approach followed by an unexpected go-around. They identified the six best-performing flightcrews (based on how well they managed the go-around) and compared them to the other six flightcrews. The strongest difference in this case, interestingly, was associated with the PM. The PM from the best-performing crews, on average, had a %DT of 19% on the airspeed area of interest (AoI), versus less than 12% for the other crews.

The intent of these studies is to understand how skilled pilots or flightcrews are attending to the flight deck interface when performing common flight tasks. These researchers have identified differences between pilots in terms of where they are looking and how well they perform the flight task. However, these findings fall short of a prescription for training the application of visual attention. Ideally, the results would lead to a more detailed description of the links between fixations and control actions. Is there a critical time, for example, for checking airspeed or attitude in a particular maneuver? Or is it truly just a higher fixation frequency that leads to better awareness of deviations from flight path management targets? The results, at least, show there can be strong differences in how visual attention is allocated and that those differences can be associated with flight performance.

The review in Appendix B offers broader ideas about differences in visual attention between skilled and less-skilled pilots. Specifically, researchers have investigated the value of the following measures:

- Fixation dwell time, which is the duration of a fixation. Skilled performers may have shorter dwell times overall.
- Data relevance, which addresses which indications or displays are fixated. Skilled performers may be better at allocating attention to task-relevant information.

4.2. Skilled Monitoring for Difficult FPM Situations

A second approach to uncovering the elements of skilled performance is cognitive task analysis (see Crandall, Klein, & Hoffman, 2006), which we applied to explore monitoring strategies as described by skilled pilots. We identified a small set of difficult monitoring situations (see a longer list in Section 3) and interviewed skilled pilots to understand how they use their situation model to identify and monitor relevant indications. One of the primary ideas in our monitoring model is that the situation model is a strong driver of monitoring, and one of the benefits of the situation model is generating expectations about the current and future states of the airplane and the operational environment.

To further explore this idea, we interviewed ten highly experienced commercial jet transport pilots about how they monitor when faced with difficult FPM situations. We tried to identify links between their thinking (situation model) and their activities, such as display selection and monitoring. These discussions led to the creation of representations that help characterize how a pilot's understanding of the current situation can lead to generating expectations and then sampling from the flight deck interface to confirm (or not confirm) those expectations.

The following pages show these representations for a set of four cases in which monitoring FPM can be difficult. Each representation has four columns:

1. *Situation model*. Recall that a skilled pilot has a number of mental models of elements of the larger system—such as weather and autoflight—combined with knowledge gained from experience and training. Appropriate knowledge is brought together to form a more complete mental representation of the current state of the airplane in the operational environment. This representation is the situation model. [Note that the four cases below show the link, with an arrow, between specific understanding and expected values.] For this exercise, we focused on identifying the following types of knowledge:
    - current position relative to trajectory/path and relative to energy
    - understanding of FMS flight path, particularly position or energy threats downstream
    - understanding of other FPM constraints, such as weather, traffic or terrain
    - understanding of typical airplane performance, such as how quickly it can slow down
    - what behavior you will get from the current autoflight mode
    - identification of locations or values that define decision gates
    - how to configure the flight deck interface to get important FPM information
2. *Expected values and decision gates*. The situation model should aid the pilot in developing expectations about the values of indications, such as airspeed; when an autoflight mode will engage; how the airplane will perform; and in generating decision gates. Decision gates are points in space or time that define a need to shift FPM strategies.
3. *Actions*. This column captures the range of actions the pilot takes to execute FPM. This includes monitoring and control actions. Control actions and verbalizations are easily observed. Harder to observe are the eye fixations—the places where visual

attention is allocated. [Note that the four cases below show a link, with a dashed line, between monitoring actions and the displays needed for that monitoring.]

4. *Display elements*. This column captures which display or indication the pilot is monitoring. It is not a complete account of what a pilot will fixate but identifies important information for FPM.

The representations on the following six pages show the links across these four columns; understanding leads to generating an expectation, and monitoring is used to confirm that expectation, or the expectation provides context for monitoring. Each FPM monitoring situation is described and an example is identified.[1] We then lay out skilled behavior in small chunks, separated by some description. The bold, underlined statements identify trigger points in the scenario.

These four case studies emphasize the following important points:

- Monitoring is not merely looking at core flight instruments (not just instrument scan); monitoring requires comparing those indications to the expected value. Skilled pilots either understand how the airplane should perform and/or are gathering data over the course of the flight to generate expectations about what they will see. (How does the pilot perceive the situation?)
- Skilled pilots do not look continuously at the full suite of flight indications; the context or situation determines which indications are relevant for monitoring.
- A skilled pilot maintains a situation model mentally that provides a richer, or at least different, representation of the airplane's current situation than does the collection of flight deck indications. Certainly, newer, more-integrated displays, such as the NAV display and Vertical Situation Display (VSD), are moving in the right direction, but skilled pilots seem to develop an understanding that is separate from the collection of indications in front of them.

[1] These examples are largely based on Boeing airplanes, and monitoring details will change as the airplane model changes. Also, the first three examples assume that the airplane does not have a VSD although the last case does use that interface tool.

4.2.1. Case Study #1

Case 1: The airplane is held at cruise altitude past the Top of Descent point, and there is a waypoint altitude constraint that may be hard to make.

Example: Held in CRZ at FL310; using the BDEGA3 arrival (LEGGS transition) into SFO. The waypoint LOZIT has an 'at or below' restriction at 16,000

Situation Model	Expected Values & Decision Gates	Actions	Display Elements
<u>Nearing ToD; prior to new clearance</u> Flight path is programmed in the FMS; A/P will descend aircraft on the arrival as programmed. LOZIT to BDEGA leg is steep (known from experience)	Should cross LOZIT below 16,000 due to geometry of FMS flight path	Verbalize current plan and any inconsistencies. Monitor winds and update FMS Identify difficult FPM segments Identify traffic that might cause a late descent	 NAV Display wind vector FMS LEGS page NAV display, Radio chatter

At this point, the flight crew is expecting a normal descent on the FMS flight path. However, the flight crew is also looking ahead for potential threats to the plan, such as changes to the wind or traffic. Also, the flight crew has determined that one of the flight plan legs is pretty steep, and they know it can be flown more easily if they cross LOZIT well below 16,000 ft (rather than simply making the restriction at 16,000 ft).

<u>ATC requests delaying descent until notified</u> Airplane will now go above FMS flight path; Eventual flown path will need to be steeper; Use 3-to-1 to determine: What is latest position where I can still make LOZIT at 16,000? Generate other options in case you reach the gate	Generate a GATE on current path at FL310 where it is too late to get down to LOZIT at 16,000	Estimate how long you will be held high CNTRL: slow down, as possible, to preserve options and decrease energy. Monitor winds Monitor position relative to gate	Radio Traffic NAV Display wind vector NAV Display

The flight crew understands that as they continue flying at FL310, they will get further above the FMS flight path and that they will need a steeper descent, and that there is a limit in terms of how long they can remain at FL310 and still make the restriction at LOZIT. They use the 3-to-1 heuristic to generate where the decision gate is that determines if they can make LOZIT at 16,000. They also begin to plan for what to tell ATC if they reach that decision gate.

Situation Model	Expected Values & Decision Gates	Actions	Display Elements
ATC requests start descent (prior to reaching decision gate) Starting down will transition from VNAV ALT to VNAV SPD, and stay in VNAV SPD until you get near the flight path.	Vertical mode = VNAV SPD	Monitor vertical autopilot mode	FMA
Because the airplane is above FP to LOZIT, need a steeper descent rate; speedbrakes and higher airspeed can help.	Appropriately high rate of change to height above path	CNTRL: extend speedbrakes CNTRL: Speed up to descend faster Monitor height above path	VDI (icons and digital values)
Vertical mode will transition to VNAV PTH when you are close to the FMS flight path	Vertical mode = VNAV PTH	Monitor vertical autopilot mode	FMA VDI

The flightcrew, after starting down, is monitoring to determine how well they are progressing toward making LOZIT at or below 16,000. They will expect to see appropriate mode changes and expect to see fairly rapid change on the digital readout of the vertical deviation indicator (VDI).

They may also use the green arc, which is highlighted in a similar situation in the next case study.

4.2.2. Case Study #2

Case 2: On descent, ATC requests a 'direct to' that eliminates some track length, and there is a difficult-to-make waypoint altitude restriction.

Example: BDEGA3 arrival (LEGGS transition) into SFO. From PYLLE cleared direct CORKK, comply with airspeed and altitude restrictions.

Situation Model	Expected Values & Decision Gates	Actions	Display Elements
<u>On descent; prior to new clearance</u> Flight path is programmed in the FMS; Autopilot will descend aircraft on the arrival as programmed.	end of arrival is CORKK and BRIXX at 11,000	Verbalize current plan and any inconsistencies. Monitor for potential changes to flight plan	Radio Traffic Traffic on NAV Display
<u>ATC requests 'Fly direct CORKK,' which shortens path</u> This change means a loss of track miles and the airplane will now have to fly a steeper path (fewer miles but same descent distance).		Change FP on FMS LEGS page: Direct to CORKK	CDU: LEGS
		Monitor winds	NAV Display Wind Vector
	CORKK at 11,000 and 250 kts	Estimate extent of change to vertical path by looking at size of flight path arc that has been removed with the 'direct to'	NAV Display

The flightcrew is initially flying the arrival as planned but also monitoring to determine if changes are likely. At some point on the descent, ATC requests a direct route to the waypoint CORKK, which is several waypoints ahead on the FMS flight plan. There is an airspeed and altitude restriction at CORKK. The flightcrew enters the new clearance into the FMS and also uses the NAV display to judge how big a change this will introduce to their flight plan track. A big change in track will create more of a challenge to get down and slow down.

For an aircraft with no VSD, the MCP-altitude green arc is the quickest gross indication of energy state. This arc predicts when the airplane will reach MCP altitude. The green arc should be in vicinity of CORKK when 11,000 is in the MCP. This altitude is also the most restrictive for this arrival.	Green arc should be near CORKK	Monitor position relative to restriction	NAV display / green arc
Other descent modes can also provide more direct control over descent rate. For FLCH, you need to ensure throttles are idle (not in HOLD at a higher setting)	Decelerating from 280 to 250, allow at least 3 miles prior to CORKK to reach 11,000 ft		

The flightcrew is trying to determine how to judge their ability to descend to 11,000 and slow down to 250 kts with the track miles they have. They have identified an indicator that can allow them to get a good estimate of their success.

Situation Model	Expected Values & Decision Gates	Actions	Display Elements
It is possible to increase descent rate by deploying speed brakes and/or increasing airspeed temporarily.	Green arc should move closer to your current position and be before CORKK	CNTRL: extend speedbrakes. CNTRL: increase airspeed temporarily, as possible, to increase rate of descent CNTRL: Engage FLCH or VNAV Speed (pitch for target airspeed)	Speedbrake handle position Airspeed; Vert Speed FMA
Other descent modes can also provide more direct control over descent rate. For FLCH, you need to ensure throttles are idle (not in HOLD at a higher setting)	Vertical speed will increase. VDI will show good progress toward getting on VNAV flight path; eventually transitioning to VNAV PTH	Monitor height above path Monitor vertical autopilot mode Monitor position relative to restriction	VDI FMA NAV Display and Green arc, Altitude, Vert Speed

The flightcrew needs to take action to aid them in descending and slowing down, and as they initiate these control actions, they will also need to monitor to determine that they are making adequate progress toward their goal. Further control actions may be needed if there is a judgment that progress is too slow.

4.2.3. Case Study #3

Case 3: An incorrect barometric pressure is entered, which leads to an incorrect altimeter during descent.

Example: The PM inadvertently enters 30.22 into the altimeter instead of 29.22, putting the airplane 1000 ft lower that what the altimeter says.

Situation Model	Expected Values & Decision Gates	Actions	Display Elements
<u>Throughout the flight</u>			Record altimeter values on paper Or Use the standby baro altimeter. This serves as a scratch pad for the initial value and allows monitoring of changes as well as displaying the difference in altitude in standard versus local altimeter setting.
Desire to generate an idea of what altimeter setting will be at destination		Get ATIS from datalink several times. Check on destination weather and whether it is changing.	
Use understanding of weather and pressure to generate an expected value	Expected value using current ATIS value and weather trend	Write down / record expected pressure for destination	

For this case, generating an appropriate expectation, through early monitoring of automatic terminal information service (ATIS), is as important as monitoring the altimeter on descent. The flight crew is generating a small range of expected values based on their understanding of the weather.

<u>Receive altimeter setting on descent at transition level</u> Compare the new value to the expected value to determine if it is outside the bounds of what is expected		Pull arrival ATIS from datalink to evaluate WX, which includes altimeter setting.	
If the value you entered is out of bounds, confirm with ATC	Revise expected value as appropriate	monitor altimeter	PFD altimeter

Ideally, the flight crew is careful to set the altimeter with the correct value. When there is a wrong value, looking at it by itself may not be a sufficient cue to realize it is incorrect. This monitoring task relies on catching errors due to having a strong expectation about what a reasonable value will be at arrival.

4.2.4. Case Study #4

Case 4: Interference with ground-based equipment leads to a false glideslope signal that has a steeper angle.

Example: On approach to KSJC in instrument conditions; aircraft on ground blocking G/S antenna and distorting signal, presenting steeper G/S.

Situation Model	Expected Values & Decision Gates	Actions	Display Elements
Prior to being cleared for the ILS approach Develop a general idea where you will intercept the ILS glideslope and localizer. The ILS is ground-based and very reliable; failures are not expected.		Brief expected rate of descent. Brief expected ground speed Brief G/S angle	PFD Approach Plate EFB
Cleared for the ILS approach After Approach mode is armed, mode changes will occur when the G/S and LOC are captured.	G/S and LOC deviation should be reasonable before intercept position. VNAV should indicate close to on path even though it is not controlling the path	Arm the approach mode Monitor current G/S and LOC deviation and trends Monitor descent indications (confirmatory)	MCP, FMA PFD, G/S LOC CDI VDI VSI Altimeter

In this initial phase, the flight crew is capturing the ILS and monitoring to confirm that capture has occurred, and the G/S and LOC deviations are about what was expected. A false glideslope has the same appearance as a true glideslope, and any evidence of a false glideslope is in the descent indications.

After capturing the ILS approach After ILS capture, there are a number of ways to confirm that the approach is going as expected; specifically, expect the following	Pitch should be reasonable and flight path angle should be around 3 deg; VS should be normal between 800 and 1000 fpm; Airspeed should vary no more than 10 kts from the target speed; If there is a VSD, its noodle should terminate at the end of the runway; GS should be +- 1 dot; VDI should look like it is close to being on target. Green arc should be near TD zone while TDZE is set	Monitor various confirmatory indications	PFD VDI VSI Altimeter Green Arc

In the case that the airplane is descending on a false G/S, the flight crew should see pitch and the FPV below 3 degrees; VS is likely to be higher than expected; the VDI may show more deviation than expected; airspeed may be higher than expected, and, if there is a VSD, its noodle may intercept the ground prior to the runway symbology.

5. A Historical Perspective: How Monitoring has been Trained

Historically, training of monitoring took a narrow view of monitoring, focusing on a scan pattern of where pilots should look. The training, such as it was, specified a sequence of fixations on different instruments. More importantly, most pilots first learned to fly in a small airplane with very basic instruments. Figure 6 shows the standard layout of the core flight instruments. Across the top from left to right are airspeed, attitude (commonly referred to as the attitude direction indicator or ADI), and the altimeter. The bottom row, from left to right, shows a turn coordinator, heading, and a vertical speed indicator. This configuration, which is commonly referred to as the "six pack," was based on patterns discovered in a study of pilot scanning by Fitts, Jones, & Milton (1950). The label "basic T" was also used to highlight the importance of airspeed, attitude, altitude, and heading.

Figure 6. The "six pack:" basic flight instruments.

The vast majority of pilots were taught to scan these instruments using a hub-and-spoke, or radial, model with attitude as the hub: attitude, airspeed, attitude, altitude, attitude, heading, attitude, etc. In this approach, the attitude indicator is the most important indicator and should be checked after looking at any other indication. Some pilots seem to use an alternative approach to scan these instruments: a more circular pattern that starts with airspeed or attitude and scans around the six pack in a clockwise fashion to check all the indications.

The larger point of this training was to give the pilot a simple routine for checking each of the core flight instruments in a way that aided understanding. In modern commercial jet transports, these core instruments have mostly been converted from round dial indicators to a different format and are integrated into the PFD (see Figure 7). Airspeed is presented on the left, attitude in the middle, altitude to the right of that, and vertical speed is all the way to the right. The heading indicator is at the bottom. The turn coordinator has not been preserved in its original form. Other indications are integrated here as well.

Figure 7. Primary flight display.

Further, for the modern commercial jet transport, the PFD is just one of many displays on the flight deck interface. Figure 8 shows the flight deck interface for the 737 MAX, including dual head up displays (HUD). There are displays here to support navigation, flight planning, autoflight engagement and management, and other functions critical to a commercial flight.

Figure 8. Full 737 MAX flight deck.

The monitoring training that every pilot received to maintain awareness of the six core instruments offers little guidance for effectively monitoring this more complex interface. In our discussions with pilots and airline training departments, we have yet to hear a pilot who says he or she was trained on the best way to monitor a modern flight deck interface. We do not believe that there is a well-defined approach to monitoring that captures the expertise—however that expertise can be characterized—that develops over years of flying these airplanes.

6. Method for Identification of Relevant Training Literature

6.1. Conceptual Scope

Our review of the research literature, which was performed in 2019, is anchored in the much broader concept of monitoring described in Section 3, in which monitoring spans the processes of:

- gathering information from the world
- integrating the information about the current situation with background knowledge (mental model) to build a situation model
- using the situation model to assess and predict currently unfolding events

These sensemaking activities do not have clear boundaries and include activities that are sometimes characterized as "good airmanship" or "flying ahead of the plane."

The focus of the literature review was on training, and an effective training program is driven by several types of information, as illustrated in Figure 9. Any training program makes explicit or implicit assumptions about the components in Figure 9: the needed skills and knowledge to be trained, the training methods, and pragmatic constraints on development and delivery. Ideally, we would be able to consult reports about a variety of training programs for airline pilots to improve their monitoring and situation awareness, perhaps generally or even specifically targeting FPM; comparing the evaluations of such reports could show which training content and methods are most effective. Unfortunately, there are very few reports on the design and evaluation of such training programs. However, the broad types of skills and of knowledge, and their effective training methods, are likely to share similarities across other dynamic work domains. Therefore, we organized the literature review around work domains, selecting those we judged to have some overlap with aviation.

We looked for research on training of monitoring and situation awareness in these five work domains:

- aviation
- medical
- air traffic control or management
- driving
- elite sports

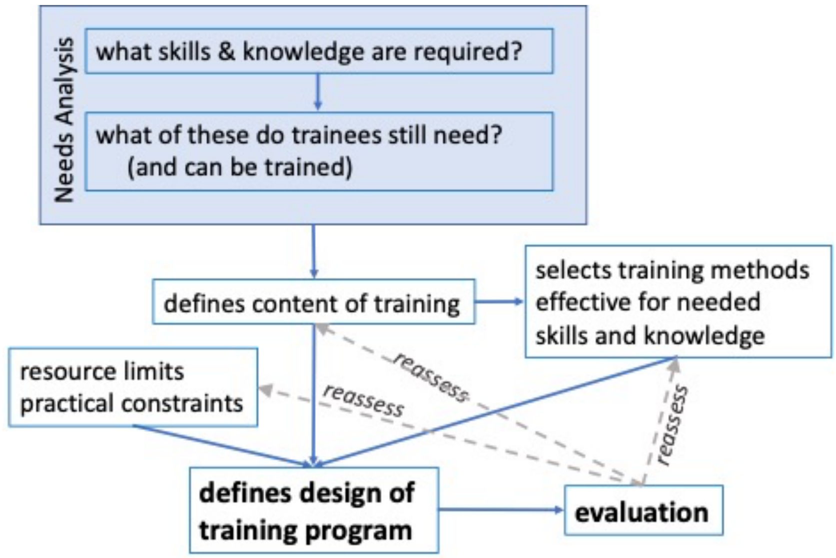

Figure 9. Training program inputs.

Monitoring and awareness of a dynamically changing situation is critical in each of these domains. We took a wide view of aviation but are primarily interested in commercial transport pilots flying airplanes with a modern flightdeck. We focused on the surgery team within the medical domain. For ATC, we address civilian flight control, again with modern instrumentation. For driving, we investigated training of non-professional drivers. We touched on a range of individual and team sports. These domains differ from aviation along several dimensions and we thought comparison across them might help clarify the conditions for effective training. We note seven domain characteristics along which comparisons can be made.

- *Dynamic pace*. In some domains, decisions and actions must be made at a much faster pace; in other domains, at a similar or slower pace, though all involve acting in a dynamic environment.
- In some domains, *performance-limiting factors* by the person are primarily cognitive, while, in others, perceptual-motor skills are primary. Note that in manual flight in visual conditions, perceptual-motor control loop skills are important, but we judged these rarely to be critical for current airline pilots.
- *Training intensity* is a rough indicator of the length of initial training and the intensity of recurrent training.
- We note the prevalence of *individual versus team performance*; in team domains, adoption of training methods based on Crew Resource Management is frequent.
- We also note the type of *information environment* that the person relies on. This may be primarily an Engineered environment of instruments and displays with abstracted information, or a Natural one, where events and changes may be more directly perceived.
- Finally, different fields have different drivers for *conducting and publishing research*. Medicine and Driving have: a) strong public concerns about safety; b) funding to support research; and c) researchers primarily rewarded for publication of results rather than development of materials held within a company.

This characterization is informal but looking across these similar domains should produce a broader understanding of the research and of what does or does not work in training monitoring.

Table 1. Informal Characterization of Dynamic Work Domains Reviewed

	Aviation	*Medical*	*ATC*	*Driving*	*Sports*
Dynamic pace	modrt-fast	moderate	moderate	fast	very fast
Sensory-motor limits	low	moderate	low	moderate	very high
Cognitive complexity	high	very high	high	low	moderate
Training intensity	high	very high	high	low	very high
Individual or team	team	team	team	individual	varies
Info: E or N - 1st	E(flightdeck)	N (humans)	E (radar)	N (streets)	N(ball, player)
Info: E or N - 2nd	N(window)	E(monitors)	N (window)	E(dashboard)	
Research & share	weak	strong	weak	strong	moderate

Key: info: E or N is whether the information source needed for operation is Engineered or Natural.

6.2. Search Tactics

Initially, we attempted a systematic review, but our goals did not match the organization in the literature, making it impractical. Instead, we used heuristic strategies, searching within each domain first with a search engine, accompanied by exploration of references and review papers.
We used Dimensions as the search engine (https://app.dimensions.ai/). This is a free, publicly accessible search tool that includes research publications and policy documents. It has effective Boolean search, good export capabilities, and forward and backward citations.

For each domain we searched for the combination of:

- "training"
- search terms describing the domain
- search terms for the type of activity

The search was almost always a restricted search of title and abstract. Terms were broadened or narrowed to find a manageable number of hits and adapted to terminology in the domain. For example, "medical" was too broad, and based on a preliminary exploration, we used this expression: ("operating room" OR anesthesiology OR surgery OR laparoscopic OR "emergency room" OR pediatric). Iterating on descriptions for the target aviation domain resulted in use of (aviation OR "flight deck" OR cockpit OR pilot). Concerning the type of activity, we relied heavily on "situation awareness," as the term "monitoring" often produced hits on other meanings of the term; "CRM" was also useful. "Hazard perception" was useful in the driving domain. Sometimes we conducted more specific searches; e.g., to find all the work of a particular research group. The literature search was conducted in 2019.

6.3. Selection Strategy and Kirkpatrick's Four Levels of Evaluation

We produced at least one general search for each domain. For each, we read the titles, excluding those clearly irrelevant to our topic. For the remainder, we read abstracts to determine relevance.

Inclusion criteria here were heuristic, in part dependent on how much research had been done in the domain.

We included for further reading any paper we identified (from its retrieved abstract or from mention in another paper or from our knowledge of the paper) that included any provision of a training intervention and included any measure of operational behavior or of impact at an operational level. We greatly valued impact on operationally relevant behavior, which is often the conceptual target of training, even when it is not measured.

These two types of evaluation (operational behavior and operational impact) align with the top levels of Kirkpatrick's (1959) classification of training evaluation methods. We characterize these levels as follows:

- The first level assesses a training intervention by asking participants to rate the program, typically for how much they liked it or how useful they thought it would be.
- The second level assesses learning within the context of training, typically, performance on a written test. This level was referred to as "learning." Note that when these levels were described, simulator-based training was not common so the context of training usually meant a classroom.
- The third level assesses change in work-relevant behavior. We included here assessment in simulators or in an actual operational context.
- The fourth level assesses impact at an organizational, not individual, level, such as reduced rates of retraining, reduced rates of safety violations, or economic impact.

In all domains, we strived to capture every study that reported a training intervention and made use of all data showing the outcome of an intervention. We included other research in a domain as well. In particular, we often included studies that provided well-motivated training plans even if the outcomes were not deeply evaluated. More detailed descriptions accompany each section.

6.4. Limitations

A number of potentially relevant work domains including shipping, plant process control, nuclear power plant operation, and security monitoring were not reviewed. The very large literature across medical fields was only sampled, with a focus on recent research and surgical settings.

We also did not review research on training and learning outside these operational settings. Work in cognitive psychology, education, and training has identified many principles important for effective learning. Our approach was not to focus on identifying principles; rather, we focused on training interventions using actual operational tasks or their analogs and looked at the overall intervention or comparison among interventions.

7. Literature Review: Training of Monitoring, Awareness, and Understanding in Dynamic Work Domains

In this section, we review relevant studies from the research literature (Section 6 describes the identification and selection of studies). For each work domain—aviation, medicine, air traffic management, driving, and sports performance—we describe what is known about training interventions that have or have not been effective for training monitoring/awareness.

In exploring the research literature on monitoring, one quickly encounters the use of SA as a concept and as a measure. This is not surprising given our descriptions of monitoring above and its role in making sense of the world. Situation awareness (for a characterization see Endsley, 1995b) is the product of monitoring and is often how monitoring effectiveness is defined. Thus, although this section often refers to improvements in awareness or SA, these should be considered as proxies for improvements in monitoring.

7.1. Aviation Domain

The importance of pilot awareness and understanding in a dynamically changing situation is widely recognized, as is the potential cost of loss of awareness by airline pilots. Although the potential value of training to improve awareness and understanding has been recognized (e.g., see Appendix A), we found few studies on training the relevant competencies. A separate report provided a review on current airline activities around monitoring.

The overall goal of the literature review is to identify research findings that might inform training approaches for commercial transport pilots. The studies we identified often differ from this target on several dimensions; most notably, the type of aircraft being flown, the level of expertise of the trainees, and the operational context. Some of the training studies have taken a “bottom-up” approach, focusing on where pilots look, and some a “top-down” approach, focusing on the process of integrating and understanding information. Interestingly, for each approach, considerable research has addressed measurement, including eye-tracking to estimate perception, and behavioral and verbal indicators to estimate SA (Flin et al., 2003). The development of measures has been particularly intense in the medical domain (as we describe in Section 7.2).

7.1.1. “Bottom-up” Approaches to Training

Several attempts have been made to characterize where expert pilots look (Appendix B describes this literature more extensively), which might then be used as a training target, with the aim of inducing less-experienced pilots to match the expert’s scan pattern (Bellenkes, 1999; Bellenkes, Wickens, & Kramer, 1997). Most of this research has been done using something akin to a “6-pack” layout instead of a glass cockpit, and in the context of manual rather than automated control. Gaze-pattern differences between more- and less-experienced pilots have been identified in some studies. Specifically, more-experienced pilots sometimes have shorter duration fixations on a single indicator with correspondingly more frequent shifts across the interface. They may also be more likely to sample less-important indicators, sometimes referred to as “minding the store;” e.g., checking heading even when the vertical path is changing (Bellenkes, 1999; Bellenkes et al., 1997). Experts also show more variation in scanning, depending on the type of maneuver; that is, experts are more sensitive to what is relevant to a particular task. Through these studies, it has been possible to construct scan-path examples for specific situations that embody properties of expert scanning and to use them to attempt to train novices to scan.

One systematic project (Bellenkes, 1999) that compared multiple training conditions tried to teach participants to follow an expert scan, given cueing where to look. However, they found that this did not result in better aircraft control performance, and trained scanning was not maintained when the cueing was removed. They observed that low-experience pilots seemed to lack a good understanding of 2nd- and 3rd-order control effects. In a subsequent study, they trained participants with no flight experience about control dynamics in an effort to provide them with a richer mental model to help them understand connections between indications. The researchers found, when comparing between experiments, that the no-flight-experience participants with the mental model training did much better than the considerably more-experienced pilots in two prior experiments.

Although the impact of the mental model training was determined from a between-experiment comparison, this rather dramatic outcome hints that scan-pattern differences between high- and low-time pilots may be better characterized as a result of better understanding. That is, it may be neither relevant nor feasible to directly train where pilots should look. Rather, it may be much more productive to train the mental model of control laws and the information needed to link up the underlying model and the current situation.

A second approach attempted to train gaze distribution at a general level: reducing participants' dwell time on a single area by training them to monitor time on task (Froger, Blättler, Dubois, Camachon, & Bonnardel, 2018). Participants in the experimental conditions were trained on a time-estimation task, learning to estimate a two-second interval and not fixate for more than two seconds on a 'heads down' task (HDT). Two independent tasks were constructed: one target-detection task presented "heads up" (HUT), and the other, an indicator-monitoring and control HDT. Two conditions included six two-minute training sessions. One used a permissive cue, allowing participants to continue fixating past two seconds; the other used a non-permissive cue, which covered over the area to prevent seeing it. They found no differences between conditions concerning performance on the target tasks. They also found no difference initially on number of violations of the two-second rule when the cue (permissive or non-permissive) was removed. They did find, 24 hours later, that the non-permissive cue condition produced fewer two-second violations than did the control condition. The supposed mechanism for time-estimation tasks is integrating and automating the estimation with the task (Froger et al., 2018; Taatgen, vanRijn, and Anderson, 2007, cited in Froger et al., 2018). Because this experiment embedded time-estimation in a single task, it raises questions about whether or how well time-estimation could be trained in the highly dynamic task sets of commercial jet transport. As the authors point out, assessment of transfer is needed.

As with many of the training studies, the situations used in the Bellenkes (1999) and the Froger et al. (2018) studies differ from in-flight situations in a commercial jet transport. These part-task studies used participants with less knowledge, skill, and experience; the flight displays were older "analog" displays instead of the modern "glass cockpit"; and the operational context was extremely simplified.

7.1.2. "Top-down" Approaches to Training

The "top-down" approach, instead of focusing on how gazes are controlled, trains the process of integrating and understanding information. We found several, somewhat interrelated, research projects pursuing top-down approaches to training to improve SA. We grouped the projects as follows:

- Targets airline pilots and embeds SA training within the broader training of non-technical skills (NTS) or CRM.
- Targets airline (or helicopter) pilots and focuses specifically on training SA.
- Targets general aviation pilots and focuses on SA.

Almost all of the directly relevant research was done in the 1990s or early 2000s. Hoermann (personal communication, 2019) notes that one of the largest projects was shut down due to a refocusing of aviation priorities following September 11, 2001. We found two research programs and one unpublished thesis focused on SA training in the last decade (Kearns, 2010, 2011; Muehlethaler & Knecht, 2016; Potter, 2011) that used general aviation pilots and military helicopter pilots. Research was included in this review if the research addressed evaluation of the effectiveness of a training program.

7.1.2.1. Training SA for Airline Pilots within a CRM/NTS Training Context

NTS and CRM are highly overlapping concepts (NTS is the label favored in Europe and CRM is favored in the United States); SA is a component skill for each. Indeed, SA was introduced to airline training through these broader concepts. Research in this section focuses on the impact of these broader programs on SA. Further, situation awareness/assessment and crew communication are closely linked in these studies. Communication, for example, from PM to PF, reveals PM's awareness, at least partially, and the expected need to communicate particular information can guide and motivate monitoring. Although the impact on SA is not always assessed or reported separately in these CRM studies, this research provides hints concerning SA training. We focused on studies where training or performance on SA or assessment was broken out from more general CRM performance.

A 2001 systematic review (Salas, Burke, Bowers, & Wilson, 2001) provides an overview of training research on CRM from 57 studies and we analyzed descriptive tables in this report to identify studies relevant to SA. Of the 57 studies covered, 15 specified SA or assessment as a training goal, and ten of those 15 assessed training outcomes by looking at operational pilot performance or organizational outcomes; that is, not just by assessing attitudes or in-classroom tests. For these performance-oriented training measures, the interventions with briefing and planning led to improvements across multiple studies. CRM training that measures SA can produce general training benefits.

One of the most relevant studies assessed the ACRM program (Holt, Boehm-Davis, & Hansberger, 2001), which was introduced and assessed initially at a regional carrier and subsequently at a major airline. The goal of the ACRM program was to make it easier for pilots to follow good CRM practices by providing instructions in procedural form and possibly reducing how much effort it took pilots to apply CRM principles in context. The core CRM activities were a BDAR sequence that shaped the interaction between the pilots. The BDAR steps included actions supporting SA, such as setting up a monitoring plan with expected verbalizations. The specific BDAR activities were interleaved into existing operational procedures and were included in flight deck reference manuals. One regional carrier fleet was trained on and provided with the proceduralized ACRM, and another fleet served as the control. In the intervention condition, training using the ACRM procedures was added to the regular CRM training; the control condition used only the standard CRM training. The intervention fleet used ACRM procedures for a year, followed by a final evaluation. The ACRM fleet was compared to a fleet using the existing form of CRM training on a variety of measures, including assessments of pilot performance managing a variety of unexpected problems. Most strikingly, the ACRM-trained fleet performed better than the traditionally trained fleet on multiple items in line-oriented evaluations (LOE) and Line Checks after a year of implementation. Thus, behavior at the individual level impacted fleet-level metrics, which are Kirkpatrick level 4 measures.

This study provides the best evidence we found that training, in combination with procedure change, can produce improved pilot SA behaviors at a fleet level. Likely bases for improvement were:

- providing general principles/methods for improving SA and the BDAR cycle
- linking the general method to specific procedures for how CRM actions should be applied across a wide range of situations
- providing the integrated format in the procedure manuals
- close coordination with the regional airline to get content details correct and appropriate

While this produced good benefits, it had high development costs. Preparing to embed the CRM actions into a large set of operational procedures was time-consuming, and execution of the first study required four years.

The follow-on study (at the major carrier) had a weaker intervention and assessment, all within the usual three-day recurrent training. This study compared traditional CRM to the integrated ACRM used in the regional carrier and added a third, "separate ACRM" condition. In this separate ACRM condition, crew were taught a separate BDAR procedure, but the operational procedures were not revised to interleave the BDAR actions. The format was much easier to prepare, and the hope was that integrated procedures would not be needed. This study showed some benefit of the integrated ACRM condition over the normal control condition and also that the "separate ACRM" method did worst.

Across other studies mentioned in the Salas et al. (2001) review, there were a variety of positive training benefits for measures that included SA. The Holt et al. (2001) studies provided the best data that a training plus procedures program with a focus on SA resulted in improved SA-related performance at individual and fleet levels. Note that while training was a part of the intervention, it also included important, integrated changes to the procedures in the Flight Operations Manual (FOM). Improved performance depended on the accompanying procedure changes (from Study 2).

7.1.2.2. Training directly Addressing SA of Airline Pilots

We identified only two research programs that trained and evaluated SA for airline pilots. We also identified a master's thesis on training helicopter pilots. Several researchers have suggested that airlines have been internally developing and assessing programs to improve awareness. As evidence of this, one early report from the 1990s states that many airlines adopted altitude awareness programs to good effect (Sumwalt, 1995), but we did not find reports on these programs. There are a large number of reports that offer advice about how training should be designed, often motivated by an analysis of accidents or incidents (Sumwalt, 1995), but these fell outside our evidence-based focus.

ESSAI (Enhanced safety through situation awareness integration) was a European Union-funded research project specifically focused on training and evaluating SA. It also included training of Threat Management, Situation Control, and Clues for Loss of SA—training intended to support and extend SA training. This project is described in several reports (Hoermann, Banbury et al., 2003 #5 and #6) and papers (Banbury, Dudfield, Hoermann, & Soll, 2007; Hoermann, Soll, Banbury, & Dudfield, 2004; Hoermann & Soll, 2004; Soll et al., 2003) that provide a considerable degree of detail about the project. The project developed resources for systematic:

• identification of concerns regarding SA and of the important concepts underlying effectively maintaining good SA

• development of a training program to teach the key elements identified, and measures to assess SA

• execution of an evaluation study

Figure 10 shows the core concepts ESSAI identified for SA. Note these are broad in scope and well aligned to our characterization in our monitoring model above. A broad set of processes are involved, and each is characterized in terms of sensemaking, for example “seek info” rather than “fixate this location.”

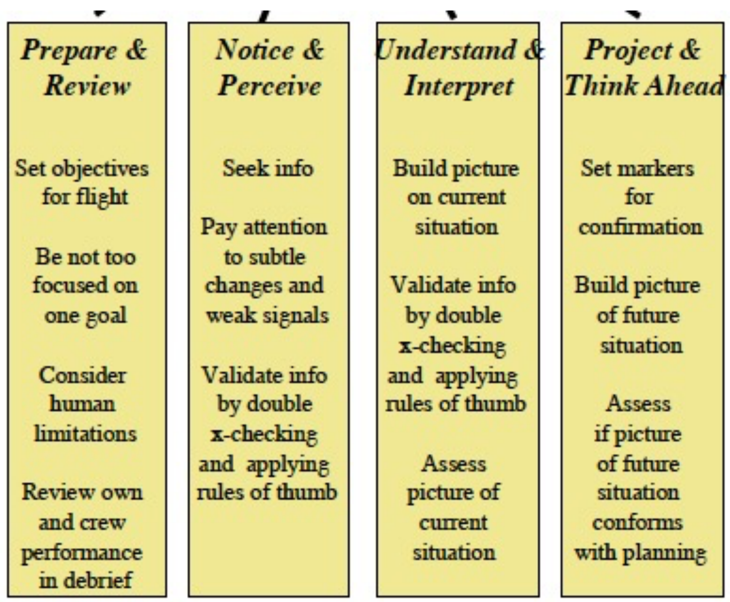

Figure 10. ESSAI model of the components process to maintain situation awareness (p. 25, Soll et al., 2003).

Thirty-two British, Italian, and German airline pilots participated in a training study that compared within individuals (pre- vs post-training) and between groups (ESSAI vs standard training). Eight two-pilot crews were assigned to either the standard training (control) or the ESSAI training condition. Pilot groups were similar in relevant experience, including prior exposure to CRM and human factors. The study design is illustrated in Figure 11. The ESSAI training provided explanations, examples, and exercises for the core concepts. Training was provided via:

• a DVD with key concepts and video examples

• a low-tech game exercise that required using the concepts to successfully “navigate” to a goal location in the training room

• two simulation training scenarios followed by debriefing

The control group received the usual LOFT training with no special emphasis on SA or Threat Management (see Figure 11).

The training simulation scenarios had challenging situations to manage, though not technically demanding in terms of flight maneuvers. Before and during the training simulations, in responding to the challenging conditions, the flightcrew also performed assessment tasks relevant to SA, and these may also have contributed to learning (i.e., directing attention appropriately). A detailed, positive debriefing with video review followed the training scenarios for each crew.

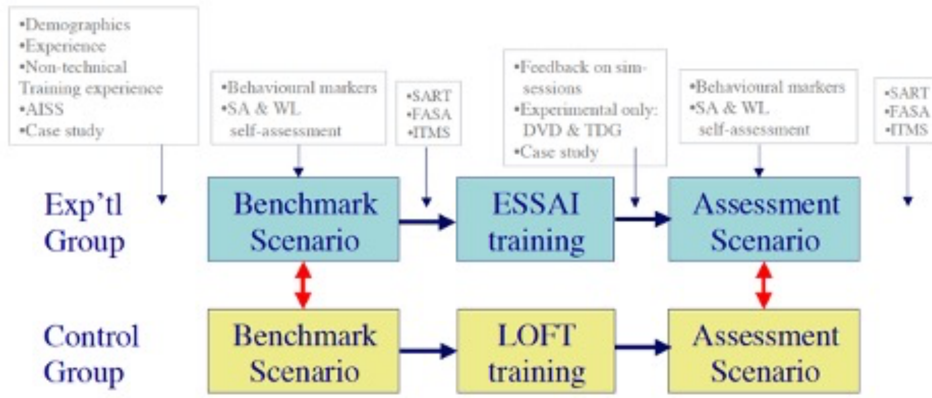

Figure 11. Design of the ESSAI evaluation (p. 35, Soll et al., 2003).

Pilots in each group had the same behavioral evaluation, which measured SA during the benchmark simulation before and the assessment simulation after a training session and compared performance between the ESSAI and control training groups. A range of measures was used, and scores on behavioral markers were of particular interest. For each of the 18 or 21 key events in the simulation scenarios, both crews were scored on a 1–3 scale for behaviors indicating noticing, understanding, and thinking ahead. Figure 12 shows an example.

Event:	Situation Awareness			
		Notice	Understand	Think Ahead
Descent Brief: Arrival weather close to VOR minima		Not notice at all OR cursory mention	Notice and talk about it / no specific discussion	Discuss plan of action of VOR approach unsuccessful
	Capt	1	2	3
	F/O	1	2	3
	Crew	1	2	3

Figure 12. Behavior scoring card (App B, Soll et al., 2003).

Pretraining SA scores were very similar between groups, but after training: a) the ESSAI group had improved, but the control group had not based on the within-subject comparison, and b) the ESSAI group performed better than the control group post-training in the between-subject comparison. The within and between design allowed a powerful analysis strategy. The events were divided by phase of flight as well: pre-flight briefing, departure and climb, descent briefing, and approach and landing. Of considerable interest, performance differences between conditions on the pre-flight briefing were due to a drop-in quality in the control condition. This provides a hint that SA training might offer some protection from a drop-in interest or alertness with greater familiarity or increased routine, as well as improvement based on additional learning.

ESSAI was an integrated project that developed and evaluated SA training, and it is an important model for future work. The details of the scenarios and the behaviors to assess will shift but identifying and developing a pool of challenging situations is a valuable extension. Further assessment of how SA training can identify and remediate particular vulnerabilities will be important. The training session relied on use of a full simulator and associated debriefing. An important question is whether significant improvement can be gained from a briefing format alone, and whether and how practice needs to be incorporated to produce benefit. Although considerable method description is provided in the reports, the debriefing is not detailed. Debriefing is a powerful training tool and understanding best practices for briefing remains to be investigated.

Robinson (2000) provided integrated training of SA and error management (EM), focusing on what he considered the cognitive rather than the team elements of CRM. This training emphasized that:

- good SA enables avoiding threats early on
- projecting future states and sharing this through briefings is important and best done at points of low workload
- together these activities free-up capacity to trap and mitigate errors and threats

The training delivered theory and practical implications in a one-day, classroom-based training session that emphasized briefings. Training was assessed in a line oriented flight training (LOFT) scenario that included three events used to evaluate three areas for SA: aircraft systems, the environment outside, and interaction of the aircraft with the world, particularly mode understanding. The last interaction area corresponds broadly to flight path awareness. Eight two-pilot crews who had been trained were compared to 35 crews flying LOFTs without any training on SA or EM. While data aggregation is a bit unclear in the report, level of awareness for Projection (the highest level of SA) was scored at 96% in the Trained vs 76% in the Control condition. Level of EM was also scored higher in the Trained condition. In sum, Robinson provided a single day of classroom training emphasizing the value of briefings, planning, vigilance, and their relation to error management. This training was sufficient to improve SA in a later LOFT setting, based on raters' scoring of behavior. Unfortunately, details about the research are limited.

This study suggests that classroom-based training, without reliance on simulated flights, may be sufficient to improve SA. Robinson (personal communication, 2019) states that key material was incorporated into training at British Airways and remains in training today.

Another relevant study (Potter, 2011) trained helicopter pilots to make call-outs to reduce altitude busts. This study focused on a "5-2-1" rule, which specifies that the PM should call out 500 feet, 200 feet, and 100 feet from a target altitude. The web-displayed, PowerPoint-designed training package provided general concepts, the 5-2-1 rule, and practice applying the 5-2-1 rule to pictured situations of descent or climb. An experimental and a control condition were run with 11 pilots each. The experimental group in the post-training simulator test made more of the altitude callouts, showing that a web-delivered training package can impact immediate performance on the targeted behavior (call-outs). However, the groups did not differ in the number or magnitude of altitude busts.

7.1.2.3. SA Training Targeting General Aviation Pilots

Three research projects have investigated training SA for general aviation (GA) pilots. An older body of work trained several underlying competencies expected to support SA, and it did so without recourse to a simulator (Bolstad, Endsley, Howell, & Costello, 2002; Bolstad, Ensdley, Costello, &

Howell, 2010). More recently, researchers in Canada and Switzerland have explored alternative, low-tech or high-tech training methods for SA training. Kearns (2011) compared computer-based training with active control in a simple simulator. Ryffel, Muehlethaler and collaborating researchers (Bektas, Knecht, Ferrari, Spillmann, & Muehlethaler, 2018; Muehlethaler & Knecht, 2016; Ryffel, Muehlethaler, Huber, & Elfering, 2019) investigated the use of eye-tracking in training in both simulated and actual flight.

Bolstad and colleagues developed a six-module training program to improve GA pilots' SA (Bolstad et al., 2002; 2010). The training concept was to teach skills underlying the ability to maintain good SA. The targeted skills were checklists, comprehension of ATC clearances, and perceptual-motor skills (trained in Experiment 1), task-switching (in Experiment 2), and contingency planning and pre-flight planning (in Experiment 3). Each experiment took place over multiple sessions. Time for Experiment 1 was about 7 hours. Training methods varied across modules, including reading materials and computer-game practice for the task-switching skill. Each of three experiments targeted the training of a particular skill set by providing information or practice on those skills. Each experiment included a control group that engaged in a learning activity not specifically targeting the core skills. Performance on the training tasks was measured before and after training, sometimes in isolation and sometimes as part of flying a scenario (e.g., a checklist questionnaire and performance on the checklist in the simulator). SA was assessed using a SAGAT-like procedure (Endsley, 1995a). Thus, the experiments tested whether:

- the targeted skill improved directly
- related performance improved in the context of the simulation
- SA improved

The results are complex, and data were only reported on measures where there was a significant difference between conditions. Considering flight performance in the simulator, there were no differences in Experiment 1 or 3, and one measure favored the Control Condition in Experiment 2. Concerning situation awareness global assessment technique (SAGAT) scores, each experiment presented 15 items, and, of the 45 opportunities, performance improved significantly more in the experimental condition on five and in the control condition on two opportunities. Analyses were only presented by-item with no overall assessment for each experiment. Concerning performance on the identical, trained task, performance sometimes did and sometimes did not improve more in the experimental condition where these had been directly trained. In short, of the many comparisons made, few differences between conditions found improvement on flight skills or on SA.

Lack of training effects can, of course, be caused by many factors. Initial skills may have been insufficiently learned to show improvement even on the same task; context or task similarities between the training tasks and the flight (transfer) tasks may have been insufficient for transfer; or the targeted skills might not be those limiting performance (on flight skills or SA measures). In summary, this type of complex training study can be difficult to perform.

Of interest, improvement in attention-switching was assessed based on improved performance on individual tasks. In the control condition, participants did not have experience doing the tasks together, and they had no practice on the individual tasks. Thus, differences in improvement might not indicate better attention-switching but simply more practice on the constituent tasks. Improving attention-switching sometimes requires large amount of practice.

More recently, researchers in Canada and Switzerland have explored alternative, low-tech or high-tech training methods for SA training of GA pilots. Kearns (2011) compared computer-based training with active control in a simulator. An extended group of collaborators at Zurich and Bern have been investigating curriculum and eye-tracking technology for curriculum implementation (Bektas et al., 2018; Muehlethaler & Knecht, 2016).

Kearns notes that for GA pilots, there are few analogs of CRM training; i.e., training for Single-Pilot Resource Management (SRM). Kearns developed a training program for workload management and SA to assess the feasibility of computer-based training rather than simulator-based training. The two 90-minute training conditions each included core concepts for SA and workload management, plus practice exercises. In one training condition, students used rudder pedals and yoke to control flying in several training scenarios. In the other training condition, students were asked to imagine they were the PF and watched videos of the same scenarios. Thus, these two conditions differed only in whether participants actually flew or imagined themselves flying (no information was provided about the quality of the example flight presented in the video). A no-training control was included. Participants were private pilots with a mean of 165 flight hours, assigned randomly, 12 per condition. Finally, all three conditions were assessed on a high-fidelity flight simulator. Workload was assessed with a secondary task: detecting whenever any of five letters held in memory were detected; SA was measured with SAGAT.

Strikingly, both training conditions performed better than the no-training condition on the SAGAT measure of SA, with large effect sizes ($d = 1.5$ for the Mental Practice Condition and slightly less for the Hands-on Practice). Not only did the Mental Practice improve SA, but it did so as much as did the Hands-on. Concerning workload management, there were no condition differences and the biggest contrast accounted for very little variance ($r^2 = .02$). Thus, the training benefit was specific to awareness.

This study led to several interesting points for follow-up. First, for this SA measure, this airplane, and these pilots, a "mental practice" training exercise in which pilots imagined themselves flying the trajectory shown in a video produced as much improvement as did flying the simulator. Second, this equivalence of a 'weaker and cheaper' training intervention to simulator flying contrasts with both broad expectations and other situations where simulator training has been compared to a 'lesser' method (Lee Chang et al., 2017). Note that Lee Chang et al. (2017) compared a longer, simulator-based training that included debriefing with a shorter, lecture-based training. Third, neither training condition produced any benefit to the secondary task measuring workload management. Despite the differences in aircraft and pilot skill level, this study raises a number of interesting questions, which are addressed below.

The Swiss researchers have two related lines of research that are on-going. One focuses on developing a training curriculum that uses head-mounted eye-tracking to aid both real-time feedback and debriefing; the second develops the technology to support this vision. The curriculum was designed to support active learning, self-explanation, high realism and, thus, increased transfer to actual use (Muehlethaler & Knecht, 2016). The curriculum was structured to teach SA and provide the following:

- a theory concerning SA, including mental models and scanning behavior
- a goal-setting briefing
- a simulator flight with real-time feedback from an instructor
- a second theoretical part including a detailed flight analysis
- a briefing including goal-setting
- the second iteration of the same simulator flight
- debriefing and formulation of take-home messages

Scenarios can be designed to focus on different situations, but training may be particularly helpful for situations that both require high SA and are high risk, such as Upset Prevention and Recovery Training (UPRT).

Ryffel et al. (2019) conducted two formative evaluations investigating how to exploit eye-tracking technology to train UPRT. SA is a critical part of prevention, and the goal was to train awareness and response, not gaze. They used simulated (n = 22) and actual flight (n = 6). This was not a training study; performance data were not presented, and the emphasis was on evaluating usefulness of the eye-tracking technology for training. However, the authors, and their participants commented on several possible roles for eye-tracking, including additional information for instructors during the simulation and for the pilots afterward. Both participants and authors thought a third use would be important: identifying and teaching the "correct" gaze, though we have no evidence that a correct gaze has been defined.

These possible uses vary in how much eye-tracking data are being used to train "bottom-up" and "top-down" processes. On the one hand, the goal of training might be that the trainees point their eyes along a specified trajectory. On the other hand, the goal of training might be to maintain a relevant, updated situation model with eye-tracking data providing information about what the user currently considers relevant and whether they know efficient "tactics" for finding that information. We believe it would be valuable to help the pilot to:

- integrate information in a situation model
- understand what information is relevant to sample or update
- apply tactical "tricks" for efficiently gaining and checking information

We did not, however, find any training studies that assess whether use of eye-tracking in fact improves training outcomes. (See Appendix B for related discussion of eye-tracking for airline pilots).

7.1.3. Themes from Aviation Training

We offer the following high-level points as a summary of these varied studies:

1. *There are only a few strong studies that assess the effectiveness of methods for training pilots to monitor and assess, whether for FPM or for other aspects of flight.* SA is often taught within the broader context of CRM, or in one case, with threat management. The two most comprehensive studies are from the 1990s (Hoermann, Soll, Banbury, & Dudfield, 2004; Holt, Boehm-Davis, & Hansberger, 2001) and may not reflect current pilot demographics, baseline training, operations, and aircraft. While the training principles are likely the same, the most appropriate topics and content may have changed.
2. *We found no evidence that it is possible to train gaze directly*, either a specific scan pattern or time to alternate between two fixation areas, and we found one failure to train gaze despite extensive, informed efforts (Bellenkes, 1999). Indeed, in the Bellenkes study, they 'accidentally' (in a between-experiment comparison) found that teaching a mental model was associated with large improvements in performance. Also of interest, in a training intervention designed to improve both SA and attention management, effects on SA but not attention management were produced. Further, the studies attempting to train regulation of eye movements used a much simpler set of displays than is found on a modern flightdeck. For the more complex flight deck interface, it seems even less likely to be able to train an effective scan pattern. Thus,

the limited evidence from aviation suggests that it is more promising to train higher-level understanding that can in turn be used to set goals and regulate information uptake than to train eye-movements or attention control directly. We will look to other domains for further evidence.

3. The aviation studies suggest other more-specific conclusions about training. Again, evidence from aviation for each of these conclusions is limited and a wider base of evidence is needed.
    a. *Integrating general with specific information is difficult and helpful to address in training*. This was done at high cost in ACRM by embedding the CRM principles (including SA) into specific procedures. This approach simplifies the integration into operations for pilots. An important question is whether and how pilots can be trained to do this more independently. Training that links general principles to specific situations in scenarios is promising, as in the ESSAI work. One "expert-novice" study (i.e., more and less experienced pilots) (Doane, Sohn, & Jodlowski, 2004) suggests that making the link between general knowledge and the specific situation is a vulnerability for novices. More- and less-experienced pilots did equally well making predictions that could be based solely on expectations from a general understanding of the situation. Less-experienced pilots did more poorly when the performance predictions had to be integrated with the details of the specific situation. Building a specific situation model may be difficult.
    b. *One finding suggests that conceptual training for SA may protect against a drop in SA over scenarios.* In the ESSAI study, benefit for the ESSAI training showed up as stable performance from first to second session while performance in the control condition dropped.
    c. *Simple observation may improve SA skills*. In one study, instruction plus watching videos on a desktop computer had significant and equal benefit to SA skill as did controlling flight with a rudder and yoke. The cases where less-interactive learning produced good benefits are important to understand because such training is more easily developed and delivered.
    d. *An important practical question is determining the magnitude of a training effort that will be sufficient to produce a valued outcome.* For example, a half-hour, web-based course produced immediate impact of the targeted behavior (call-outs) but did not impact altitude busts (Potter, 2011). Kearns (2010) was also motivated by the feasibility of moving from simulators with controls to computer-based training (CBT). Of course, training content is usually the most important factor, but ideally research allows us to identify the critical content as well as how it can be delivered (and evaluated).

7.2. Medical Domain

The potentially relevant literature in the medical domain is extensive, so we prioritized studies that used a training intervention, captured performance-based measures, and allowed a comparison between performance with and without the training intervention. We did not include exploratory studies or case studies. We did not focus on studies comparing experts and novices (or solely comparing practitioners of different skill levels) although these are widely used in studies assessing the validity of NTS. The literature is methodologically diverse, e.g., including interviews of what teaching anesthesiologists think about teaching SA (Haber, Ellaway, Chun, & Lockyer, 2017). We sampled primarily from training of surgeons and surgical teams but also included some studies of nursing, anesthesiology, and pediatrics. A recent review of SA in surgery (Schulz, Endsley, Kochs, Gelb, & Wagner, 2013) reports very little on training.

7.2.1. Primary Measures of Monitoring/Situation Awareness

Medicine has borrowed heavily from aviation in recognizing and training NTS, which include monitoring and situation awareness. The analogy between taking a plane through a flight and taking a patient through surgery has been widely recognized, and training programs in medicine, particularly anesthesiology and surgery, were initially based on CRM training in aviation. Indeed, there is a strong similarity between the concept of NTS in medicine and CRM concepts (see Helmreich et al., 1999), which are also sometimes referred to as non-technical skills. The importance of NTS has been recognized in medicine for decades (Gaba et al., 1998).

As noted above, monitoring (in the narrow sense) informs SA, which captures noticing, understanding, and anticipating the state of the patient and the operational context. Ideally, through effective monitoring, the practitioner develops an accurate, actionable situation model that captures the evolving situation awareness and assessment and supports decision making.

The specifics of NTS vary across studies but often include SA, decision making, and sharing information. SA terminology and methods have been used in medicine to account for practitioner activities such as reporting on patient status and history (e.g., blood pressure, drugs administered); responding to team and equipment status; accurately anticipating likely future events (e.g., anesthetized patient crashing); and recognizing the need for help. The development of, and particularly assessment of, training programs that target these skills has emerged more recently. The studies described here report either SA or NTS as a broader measure (without breaking out SA as a component).

Many of the research efforts to improve NTS have attempted to adapt and apply CRM training from aviation to medical and, especially, operating room settings. Training programs may be provided by commercial vendors or by research teams with medical and human factors experts, may target a broad set of NTS or a more specific skill, may be conducted over hours or months, and can vary greatly in how they are evaluated. Training programs are often initially evaluated by participant ratings of usefulness, rather than by impact on performance (Flin, 2004; Flin et al., 2007). However, there is now a large literature in medicine on efforts to improve and measure NTS, including SA. A great deal of research has been directed at developing good evaluation measures and several reviews focus on measures (Wood et al., 2017).

Resources are needed for informative study designs as well as informative measurement; some published studies use both. For example, a 2013 systematic review (Dedy, Bonrath, Zevin, & Grantcharov, 2013) of teaching a wide range of NTS for surgical residents found that only four of 23 studies used randomized clinical trials (RCT). The overall large amount of medical research produced can now contribute findings relevant to monitoring in aviation and specifically to SA. A recent systematic review of training SA in surgery (Graafland et al., 2014) found nine studies that attempted to measure or change SA. Also, Jung, Borkhoff, Jüni, & Grantcharov, 2018 completed a review focused on the development of non-technical skills for surgeons (NOTSS) and their findings related to this approach.

We identified three forms of SA assessment in this literature:

- observer ratings
- participant probes
- information checklist

Several SA measures had observers rate specific types of practitioner behavior, such as talking. Measures for medical settings were modeled on CRM check cards in aviation (e.g., non-technical skills [NOTECHS] from Flin et al., 2003). The health care provider, typically a doctor, was rated on multiple aspects of the construct being measured. non-technical skills for surgeons (NTSS) was an early, influential measure of this type, based on aviation but adapted to surgery (Yule et al., 2008) and anesthesia (called ANTS for Anesthesia NTS). Observers rated performance in simulated or real surgery for successful execution of CRM-relevant actions, including SA. For SA, the rating scale consisted of three four-point scales to assess noticing, understanding, and thinking ahead. Later rating methods collected separate scores on three topics, awareness of patient, of procedure, and of team.

Observer ratings have the advantage that they can be done unobtrusively as the medical procedure is being carried out or at a later time from videotaping, which allows coding to be more easily controlled and coders to be blind to conditions. Reliability has been a concern with observer-rating methods (for a review, see Wood et al., 2017), and while some studies have found good levels of reliability, others have not (Rosqvist, Lauritsalo, & Paloneva, 2018). Training of observers seems to be an important factor.

More recently, studies have relied on assessing a participant's situation model with a probe, such as SAGAT (Endsley, 1995a). The increasing use of simulation in medical training has supported a move away from observer ratings to the use of in-the-moment participant reports, since posing questions to the participant does not impact a patient. Cooper, Porter, & Peach (2013) identified four studies using SAGAT-based measures in emergency health care, but these studies did not have a training intervention. We found only two studies (Hänsel et al., 2012; Lee Chang et al., 2017) with a training intervention that used a real-time probe, in both cases a variant of SAGAT. The method stops the simulation, makes the situation unavailable to the participant, and asks for a report. Use of response in context and measuring time to respond, as in Situation Present Assessment Method (SPAM) (Durso & Dattel, 2004), have been recommended because they are seen as less disruptive and more informative (Fioratou, Flin, Glavin, & Patey, 2010) than measures modeled on SAGAT, but we are unaware of medical studies using this. We believe that studies using an in-the-moment response to relevant questions rather than observer ratings may provide data more relevant to the flight deck, but useful insights have been gained from both.

An important third assessment method spans between measuring SA and measuring technical performance. This method is the degree of compliance with surgery-relevant checklists. Checklists such as the World Health Organization's (WHO) Surgical Safety Checklist identify information that should be verified prior to anesthesia, prior to incision, and prior to the patient leaving the operating room. Thus, the degree of compliance with this checklist provides a natural indicator of awareness of selective, treatment-relevant information. Typically, this checklist (or an extended version) is specified as part of the operating procedure, but compliance may be variable.

Several reviews concerning measures of SA, or NTS more generally, have been compiled, typically focused on a specific medical domain (Cooper et al., 2013; Robertson, Dias, Yule, & Smink, 2017; Sharma, Mishra, Aggarwal, & Grantcharov, 2011). For example, one analyzes the concepts and measurements of monitoring and distributed SA in anesthesia (Fioratou, Flin, Glavin, & Patey, 2010).

In the next section, we report on studies that include a performance measure that used either probes of the participant-in-action or observer ratings. We do not analyze training studies where (typically early in the development of a training approach) assessment only consisted of course ratings provided by participants and had no performance measures (for interesting examples, see Müller et al., 2007; or see one-day classroom from Flin et al., 2007). Many studies assess NTS, and some address aspects not directly related to SA, such as decision making or stress management. Many of the studies that include SA assessment do so as part of a package of NTS and we try to cover these. However, our coverage is erratic for studies where SA might have been included as part of NTS but was not mentioned in title or abstract or identified in relevant reviews.

7.2.2. Methods of Training and Results

Research in the medical domain speaks to three questions related to improving SA:

- Do particular training interventions lead to improved SA?
- Does better SA—either alone or as part of NTS—improve doctors' technical performance or patient outcomes? Or does better SA at least correlate with higher performance or better outcomes? (If there were no demonstrated operational benefit of SA, the relevance of training for improved awareness would be diminished.)
- What are the properties of training interventions that have greater impact?

Two research programs demonstrate substantial impact of training SA (or NTS broadly) on improving technical skills that directly impact the patient. Several additional studies show correlations between SA and better technical performance.

Brady et al. (2013) used a multi-aspect training intervention to improve SA and then assessed impact at the organizational level. This research did not describe how to train or measure SA at the individual level, but it enhanced organizational awareness through a multifaceted approach. This program shows the potential for major impact on SA when new, relevant factors are recognized and exemplifies resilient adaptation at an institutional level. Brady et al. began by targeting and analyzing risk factors for a prominent category of unsafe events—specific types of pediatric transfers from wards back to the intensive care unit (ICU). They analyzed 80 previous cases of this event type, which is somewhat analogous to accident analysis in aviation though with different types of data. They identified five different types of information in the patient records that were unrecognized predictors of these unsafe outcomes. To increase awareness of these risk factors, they introduced the significance of these factors across the care teams. They provided a decision aid (similar to a checklist) to the providers most responsible for early detection of these risk factors. They changed communication practices, adding frequent, team-based assessments linked to rapid mitigation of identified problems signaled by the predictors. And they included ongoing hospital-level assessment of all occurrences of the unsafe events. As a result of this program, the targeted, unsafe events were reduced by almost half at the hospital level a year after implementation. This study is important and informative for two reasons. First, the researchers developed a deep, well-reasoned analysis identifying important risk factors and provided practitioners with a method for maintaining a rich, relevant, shared mental model of patient state. Second, it produced a large operational impact at the institutional level—namely, a reduction in readmissions to ICUs. This type of change from an intervention (training plus structured support) can be seen as impact at Kirkpatrick's Level 4, or operational contribution.

A second line of research, presented in several reports, describes another training intervention that had an impact near the operational level. Data came from observations of surgical teams. Three reports (McCulloch et al., 2009; Mishra, Catchpole, Dale, & McCulloch, 2008; Mishra, Catchpole, & McCulloch, 2009) assessed correlations among measures. These studies found that NTS measured with NOTECHS correlated with measured technical performance (both operating and non-operative). In particular, the SA subscore for the surgical team was particularly strongly associated with lower errors (i.e., higher technical skill). This study (Mishra et al., 2008) showed that improved SA was associated with reduction in technical errors.

McCulloch et al. (2009) reported an intervention to train NTS, and they showed that this intervention improved both NTS and technical skills, measured as a reduction in operating and non-operative errors. The intervention was carried out as part of ongoing personnel training in a hospital context, and performance was evaluated in actual practice. McCulloch et al. (2009) targeted all 85 hospital personnel involved in either of the two selected operations, and 54 participated. Surgeons, anesthesiologists, and nurses were included. The intervention:

- integrated multiple aspects of NTS in three three-hour classroom sessions, one of which addressed SA
- distributed materials summarizing the main points from the course (e.g., in pocket cards, posters)
- provided a three-month "bedding in" period where briefings, operations, and debriefings of surgeries were observed and feedback provided

The study compared 48 operations over six months prior to the intervention and 62 operations in the six months after the "bedding in" period, assessing NOTECHS scores and technical performance measures. Reliability of NOTECHS scoring was high overall (.98) and high on all four subscores (> .90). Overall, there were significant improvements on NTS (measured by NOTECHS) and attitudes (Safety Attitudes Questionnaire) and on their two measures of technical skills. However, the NOTECHS subscale for SA did not improve. There were no effects on the clinical outcome indicators of surgery duration or length of hospital stay. Improvements were greatest for the nursing team and greatest for the teamwork and for the problem-solving subscales. The SA score of the surgeon(s) on the team was correlated with team technical error rates.

The McCulloch study raises interesting questions about training SA in this domain. Is it in fact more difficult to improve SA than other cognitive NTS such as problem-solving? Is SA more difficult to measure than other NTS?

Several studies looked for correlations between technical skill and either: a) SA specifically or b) a measure of NTS including SA. Black, Nestel, Kneebone, & Wolfe (2010) reported that overall NOTECHS scores (no subscores reported) were highly correlated with rated technical skills in simulated vascular surgery. Briggs et al. (2015) studied 20 trauma resuscitation teams and found a reliable correlation between a) the two NOTSS "cognitive scales," namely, SA and decision making (but not communication or leadership scales) and b) technical skills measured as critical tasks completed. In anesthesia, Zausig et al. (2009) found a strong correlation between NTS measured by ANTS and medical management score. Moorthy, Munz, Adams, Pandey, & Darzi (2005) used a self-designed measure of NTS based on LOSA elements from aviation; they found no association with overall technical skills, and no difference between more and less experienced groups. Of note in this study, the score for monitoring/vigilance was overall the area of worst performance.

Reviews on topics related to SA in medicine, specifically surgery (Hull et al., 2012; Ounounou et al., 2019; Wood et al., 2017), have identified only a small number of research projects that have assessed the relation between nontechnical and technical skills. Nicksa, Anderson, Fidler, & Stewart (2015) provided a three-month team training of NTS and produced improvement on some aspects but not on SA. However, SA was correlated with performance and Nicksa et al. suggested that lack of improvement with training may be due to participants' already-high scores. Associations between measurements of the two types of skills reassuringly suggest that each is being measured with some reliability and validity. A correlation, of course, does not imply cause, and both skills increase with experience.

7.2.3. Relevant Lessons from the Domain

A number of studies in the medical domain have assessed the effectiveness of training interventions on SA. The Brady et al. (2013) and Mishra et al. (2009) studies are important examples, though Brady et al. did not measure SA at the individual level. SA training is often delivered as part of broader training that targets other NTS, as well. A large number of studies trained NTS and, if SA were explicitly mentioned (e.g., Graafland, Schraagen, Boermeester, Bemelman, & Schijven, 2015; Wood et al., 2017), we considered the study potentially relevant. However, we did not include training studies that did not specifically name SA. A systematic review of NTS for surgery residents does not report any randomized clinical trials (RCT) studies of interventions involving SA (Dedy et al., 2013).

Training interventions reported in the medical literature have evolved in both approach and evaluation. Training delivery has shifted from classroom to simulator, and evaluation has shifted from participant rating to some measure of performance. We believe we can take away the following lessons from this literature:

1. *The training interventions reported in this literature focus on building understanding and sensemaking skills.* Many of these understanding-oriented training interventions train and measure a suite of NTS that include SA as a component; some focus very specifically on SA as assessed with SAGAT-based measures. We did not see evidence of training goals concerned with better perception of, or vigilance to, particular indicators or instruments; specifically, we did not see anything analogous to trying to train a "scan pattern." While our review was not so comprehensive to ensure such research does not exist, it was not conspicuous in our search or in recent systematic reviews. We suggest that in both the medical and aviation domains, the training that is both needed and feasible concerns noticing and integrating information that is guided by what is important for the work at hand.
2. *Identification of information that is highly relevant to observe, understand, or anticipate, but is not recognized or is underused, is critical to a successful training intervention.* The dramatic impact of the Brady et al. (2013) study is founded on discovery of risk factors that were under-appreciated and underused across the institution, not just an individual trainee. More typically, a successful training intervention teaches relevant information that is not recognized or is underused by the individual trainees, though the value of the information is understood and used by the trainers. Pointing out this information to trainees and giving them opportunities to practice using this information can contribute to successful training.
3. *Effective training must identify information that trainees lack and that is required for the work.* Good training outcomes suggest good information identification. When an intervention program does not have the intended effect on SA or related NTS, a

candidate explanation is that the information being taught was not impacting performance. This might either be because it was already appropriately used or because the trainee was not yet at the point where the information was accessible for the trainee (as suggested, e.g., by Zausig et al., 2009). Training benefits for one subgroup but not another may also be a symptom of mismatching content to trainee needs. For example, Nicksa et al.'s training benefited more but not less advanced residents, while Rosqvist et al.'s (2018) training benefitted participants who had never participated in a simulation. When training includes multi-disciplinary teams (e.g., surgeons, anesthesiologists, and nurses), training may have differential impact by discipline.

Although training to be aware of particular information may benefit people with some levels of expertise but not others, it is very useful to identify the information any person should be aware of to do the task well. Efforts in medicine have been made to characterize the aspects of the situation that are most important; this information may then be incorporated in checklists to guide individual or team awareness at critical points. WHO has developed awareness of information that is pervasively important yet also often missed (e.g., the Surgical Safety Checklist mentioned above). Particular medical disciplines or hospital groups have derived requirements for particular types of surgery, and performance on these checklists can also be used to measure SA and the impact of a training intervention. If a checklist item is called out, the practitioner was aware of it; if not called out, the practitioner likely was unaware. Checklists can provide an extremely valuable type of measure: one that is both a natural part of the operational activity (e.g., the surgery) and a direct measure of SA.

4. *For the medical domain, interventions for training SA did not address fixating, looking at, or finding information.* While we did not frame our searches to ensure discovery of any such studies, we think the invisibility of this focus is noteworthy. The actual process of finding or looking at information is not taken to be a limiting factor in SA in any of the training studies we encountered.

5. *Successful studies in the medical domain showed great variability in the scale of the intervention and in the magnitude and scope of impact.* The Brady et al., (2013) and Mishra, Catchpole, & McCulloch (2009) studies discussed above had large-scale interventions spanning months, targeting multiple groups or a whole hospital, and included substantial procedural support. These interventions successfully shifted both technical skills and NTS over a multi-month assessment period. On the other hand, small-scale interventions also had an effect; e.g., a one-hour session with a brief, a simulation, and debrief produced better use of medical checklists on retest 20 months later (Nguyen, Elliott, Watson, & Dominguez, 2015). And, watching a one-hour DVD produced better checklist use and better NOTECHS scores (Gillespie et al., 2017). Thus, even very modest interventions can have measurable impact, not only on NTS ratings but on performance of a clinically relevant task, such as carrying out the medical actions specified in a checklist.

6. *Simulators are frequently used delivery methods for training SA and NTS more broadly and, when assessed, have contributed to learning.*
    a. In a review of studies examined for "tools" for surgeons' NTS, 44 of 84 studies used simulators (Wood et al., 2017). Simulator scenarios typically consist of:
        – a briefing (about what the case will be)
        – the actual scenario in which trainees manage the case, typically as part of a team
        – a debriefing for an individual or team that provides some form of evaluation and coaching
    b. Several studies have used a pre/post design to compare performance of a group before and after participating in a simulator-based experience; participating in the briefing/core simulation/debriefing may itself be the training (Catchpole, Dale, Hirst, Smith, & Giddings, 2010) or the simulation may be preceded by other activities, such as a short lecture and role assignment (Rosqvist et al., 2018). Trainees often show improvement after participation when assessed immediately or after months. Studies often target surgery residents as the trainees, but some use established practitioners as well as students. Impact for established practitioners may interact with the local culture, with effectiveness varying among departments or sites (Catchpole et al., 2010).
    c. In addition to widespread use of simulators across medical studies, two well-designed studies compared simulation-based training to alternatives. Hänsel et al. (2012) compared simulator training without NTS training to NTS training without simulation training and to a no-training control. They found SAGAT scores increased after the simulation-based training but did not improve in either of the other conditions. Lee Chang et al. (2017) compared simulator-based training (running five hours with eight scenarios) to a lecture form of training situation awareness (two hours) and found better SAGAT scores following the simulation-based training. Thus, there is direct evidence that simulation-based training of SA produces better learning than an appropriate lecture- and classroom-based alternative (though time and expense may be greater).
7. *Debriefing is a key aspect of much simulator training and can also be used following an actual medical procedure.* Many studies have more time allocated to debriefing than to the simulation scenario itself (Jankouskas, Haidet, Hupcey, Kolanowski, & Murray, 2011; Nguyen et al., 2015; Nicksa et al., 2015; Rosqvist et al., 2018). This allocation suggests that benefits of participating in a simulation may be heavily mediated by what happened in debriefing. Studies comparing the benefits of self-debriefing versus debriefing by a coach found no difference in one case (Boet et al., 2011) but greater improvement with coach-mediated debriefing in another (Yule et al., 2015). Indeed, Yule et al. found no improvement without coaching. When self-briefing was equally successful, trainees were told to review their videotaped performance by comparison to the NOTECHS scoring rubric. Boet et al. emphasized the importance of self-monitoring and self-criticism as an important skill for ongoing independent learning. The aspects of content, structure, and participants in debriefing that contribute to learning is an important open question.

8. *We do not know whether NTS should be trained with technical skills.* Studies have tried to layer teaching NTS with technical training by trying to increase the amount that trainers' debriefs addressed NTS (Alken, Luursema, Weenk, Yauw, Fluit, & Van Goor, 2018) or by directly integrating NTS training topics with technical training (Zausig et al., 2009). These two studies did not find improved NTS in the condition that added in NTS training. However, a simulation-based study comparing training of basic medical skills alone versus with addition of CRM training showed more improvement in non-technical skills in the condition with CRM training (Jankouskas et al., 2011). Thus, factors concerning whether and when integrated training is beneficial still need to be identified.
9. *There are two different views concerning the data availability and monitoring.* One view considers that there is a stable set of key variables that are important across situations, and good situation awareness means awareness of these variables; this can be measured by randomly sampling across this set. The second view considers that relevant information changes substantially with the situation. This focuses on awareness of a set of variables known to be important at specific points in a procedure. This is usually structured around following different information procedures or checklists at different points in the procedure (e.g., before, during, or after surgery). Good awareness means ability to report variables relevant in a particular situation or at a particular event.
10. *As with most training programs, institutional support is important*. For example, the impact of a training intervention interacted with site, in accord with differences in how the training was viewed by the site (Catchpole et al., 2010). In addition, site support for using the SA concepts in actual operations, via "bedding in" briefings, seems to contribute to successful training, though we do not know of a direct test of its value. Similarly, hospitals invested in the importance of NTS provide recurrent training.
11. *Gamification for understanding equipment may be helpful.* Surgery and other medical domains increasingly rely on complex technology and, thus, monitoring the equipment as well as the patient and the medical team is increasingly important. An exploratory study of gamification—using game-like activities—to train equipment monitoring showed improvement that transferred to monitoring in a later, realistic training exercise (Graafland, Bemelman, & Schijven, 2017).

7.3. Air Traffic Control and Management Domain

7.3.1. Training Studies and Results

Training for ATC/ATM seems particularly relevant to understanding how to train monitoring since the controller's job is monitoring aircraft for potential separation violations or a need for re-routing. Technology development has produced new tools such as those for conflict detection, conflict probes that show potential conflicts, and data link communication. These new tools and revised tasks have motivated investigations about how to train controllers to best use the new technologies and how to control when old and new methods are both in use.

In one study relevant to monitoring, Knecht, Muehlethaler, & Elfering (2016) developed a detailed prototype training program for NTS for ATM. Development was based partially on interviews with controllers (ATCOs), both for training design and review-based evaluation. This training program was structured around cycles of theory, briefing, simulation, and debriefing, and the 21 simulation scenarios were based on critical incidents collected from ATCOs. The simulations were run on standard computers. After development, ATCOs participated in the program and evaluated with ratings. They rated the program and individual scenarios as useful and realistic, which suggests they can provide a valuable, deeply motivated training program. It would be valuable if performance-based, as well as ratings-based (Kirkpatrick level 1), evaluation of the implemented program could be carried out. This training is striking in both providing detailed assessment of the contents of training for NTS and using a concept-brief-simulation-debrief training cycle. The concept-brief-simulation-debrief cycle integrates multiple useful training methods likely to be of wide applicability.

Several studies (Billinghurst et al., 2011; Vu, Kiken, Chiappe, & Strybel, 2013; Vu et al., 2012) have compared the impact of alternative training schedules on performance in a loss-of-separation task and also on SA and workload. These studies manipulated several variables, including:

- the proportion of aircraft that were NextGen-equipped
- whether manual or NextGen control methods were taught first or were taught together
- whether the test session was given immediately after training the first method or after both control methods were trained

Performance variables were loss-of-separation and the time for the aircraft to pass through the sector.

The benefits were found for Part-Task training over Whole-Task training as follows:

- Part-Task training was better when, in the Part-Task training, participants first learned to control manually and then learned to use the NextGen tools and control methods (Vu et al., 2013).
- Part-Task training was found to provide greater benefit for the slower learners (with equivalent numbers of slower learners in each condition).

Additional findings were that:

- Performance (maintaining separation) was better on the first half of a scenario (vs the second half).
- Performance was marginally better on the control mode more recently taught; i.e., manual and NextGen were better at a midterm test if they were taught first, and at a final test if they were taught second.
- The impact of the proportion of NextGen-equipped aircraft interacted with several variables across studies.
- Effects sometimes occurred with SAGAT-style measures of SA and sometimes in ATC control actions (aircraft requests).

This group of studies is valuable because it explored factors affecting learning of air traffic management (ATM) skills. Further, some findings were related to generalizable principles—such as a part-whole approach vs a whole-throughout approach—and assessing the impact of recency of relevant training to the assessment (first or second part of the session). However, these studies should be considered exploratory because of factors such as complex designs, effects from complex interactions that are often less stable and harder to measure, and challenges in statistical identification of reliable effects.

Researchers have also worked to identify domain content for the design of overall training and of interaction functionality, such as control strategy cues and means for collecting data and making decisions (Malakis, Kontogiannis, & Psaros, 2014). Inclusion of additional measures may also be valuable, such as measuring the time both a radar and a planner controller look at the same, future-relevant information (Hauland, 2008).

A third type of contribution comes from studies of meta-cognitive training in learning complex microworlds. Microworlds are computer-based simulations of some natural or engineered system that includes operator interaction. Meta-cognitive interventions have been widely explored in educational contexts but rarely in applied training settings. Kim (2018) investigated whether providing meta-cognitive training would produce performance improvements. The meta-cognitive training included giving feedback between the trainee's self-judgments and their actual performance. Kim investigated learning aircraft identification and provision of warning in a simulation-based, two-day monitoring task that had similarities with military aircraft monitoring (an Anti-Air Warfare Coordinator analog) and included 30 (probably student) participants. Providing feedback and comparison about self-assessment increased learning on the identification task. While the relevance of the simulation game to ATM is indirect, it is valuable to have research with a substantial number of participants, more than is typically feasible even with ATM controller students or retirees.

7.3.2. Relevant Lessons from the Domain

In summary, we identified few studies on training ATC skills. We believe the most-relevant contributions may be the emphasis on the cycle of teaching concepts/theory, followed by a brief-simulate-debrief training cycle and the potential effectiveness of lower-cost simulators with relevant fidelity. Both aspects have direct applicability to training pilots.

7.4. Driving Domain

7.4.1. Training Studies Results

Note that, in driving research, several terms are variously used to refer to the ability to notice, understand, or respond to hazards appropriately; we use "hazard anticipation" to refer to this spectrum of skills. In driving, the strong link between hazard-anticipation skill and accident risk has made measurement and training of this skill a priority research topic. Poor hazard anticipation is a stronger contributor to accident risk than is vehicle handling (discussed in Horswill, Falconer, Pachana, Wetton, & Hill, 2015; Pradhan, Pollatsek, Knodler, & Fisher, 2009). High-risk groups—both new drivers and elderly drivers—are also associated with poor hazard anticipation. Accident risk of newly licensed, young drivers is initially very high but decreases rapidly over the first six to 12 months. While hazard anticipation improves with experience, it does not become perfect in low-risk driver groups. One early study looked at the impact of a broad "advanced driver training" program on noticing and managing hazards and found improvement both in SA and behavior (Walker, Stanton, Kazi, Salmon, & Jenkins, 2009). The change in hazard awareness with experience, as well as improvements from a broad hazard-management training program, suggests this skill could be trained, and that training methods might accelerate and refine whatever improvements naturally occur from time on the road.

Exciting, recent results show a large reduction in accidents from a short hazard-anticipation training intervention, called Risk Awareness and Perception Training (RAPT; reported in Thomas, Rilea, Blomberg, Peck, & Korbelak, 2016). This large study involved over 5,000 young drivers and was administered through the California Department of Motor Vehicles at the time participants were issued their first license. They found that crash rates of males in the first year after licensing were 23% lower in the group who had done the RAPT training. Interestingly, taking RAPT training did not significantly affect accident rates for female drivers. This study was well-designed, and results were analyzed appropriately. This version of RAPT training provided trainees with a series of nine video-taped driving clips that contained hazards. Participants clicked on the area of the hazard and got explanatory feedback on two screens, first identifying the hazard and then explaining what should be done. Similar to previous findings, PC-based hazard anticipation generalized to on-road driving behavior and to scenarios quite different from those used in training (Pradhan et al., 2009). It is impressive that a 17-minute intervention produced an operationally significant impact over a year (Thomas et al., 2016). While no single study is definitive, these findings suggest it may be possible to see dramatic improvement from training.

We believe that a key factor in improving hazard anticipation in driving is inducing participants to engage with examples of hazards in a task-relevant way, thus enabling them to build a richer, generic mental model of hazards and a richer situation model of a particular driving environment that includes the nature and location of possible hazards. Hazard anticipation is measured multiple ways; e.g., by where participants look, by verbal identification of the hazard type, by circling areas of potential hazard, or by driving behavior, such as speed. The hazard situations and the training activities are both important inputs to training.

New and experienced drivers do not differ in detecting certain hazards; for example, all skill levels do comparably well at recognizing that a child visible between two parked cars is a potential hazard. Therefore, these hazards are not strong training targets. However, situations where a potential hazard could be obscured or hidden but still offers predictive cues seem particularly helpful to train. All the cases used by RAPT were of this type. Indeed, a topic of ongoing research is the character of the situations where performance is most affected by expertise (Crundall et al., 2012).

Several training activities produced benefits, and often multiple activities produced greater benefit. Horswill provides reviews (Horswill, 2016a; 2016b).

- Marking hazard areas or where the driver should be looking are useful tasks. RAPT requires users to identify areas showing primary and secondary hazards in a scene and provides feedback. In the Thomas et al. version, participants viewed a map perspective and a driver's-eye perspective and marked responses on the driver-perspective view.
- Listening to accompanying expert commentary while viewing scenarios can aid learning, and self-generated commentary taught in a training period can also produce improvement in subsequent performance (Horswill, 2016a; Isler, Starkey, & Sheppard, 2011), but may also slow hazard detection (Young, Chapman, & Crundall, 2014; Horswill, 2016a; 2016b).
- Stopping scenarios to have participants project what will happen next, and then continuing the scenario is also an effective method. [Note that the comparison of the prediction with actual events can serve as feedback.]
- Various forms of feedback can be helpful, and identification of the relevant hazard is important; e.g., feedback may take the form of extended commentary by an expert (Petzolt, Weiss, Franke, Krems & Bannert, 2013, cited in Horswill, 2016a.). Interestingly, just providing feedback on overall correctness did not produce improvement, and participants tended to reject the validity of the feedback (Dogan et al., 2012, cited in Horswill, 2016a).

Some research approaches provided multiple types of information or actions, such as viewing annotated clips, explaining hazard locations, plus "what happens next" (Chapman et al., 2002, cited in Horswill 2016a); or listening to instruction and classifying video hazards (Meir, Borowsky, & Oron-Gilad, 2014; Wetton, Hill & Horswill, 2013, cited in Horswill, 2016a). Providing information in multiple representation formats may also be helpful, such as the “schematic map view plus driver view” used in RAPT. One very intensive intervention used multiple methods to teach multiple skills in a summer-camp setting (Isler et al., 2011). While effective, the amount of training time makes this approach unlikely to scale well.

It is also true that some training methods were ineffective; specifically, lecture instruction alone and viewing videos of hazards without additional activity failed to produce learning (Horswill 2016a).

7.4.2. Relevant Lessons from the Domain

There are interesting commonalities and differences between training for awareness in driving and in aviation. Both fields require sustained attention to the path of the vehicle. Also, in both fields, lack of awareness has been identified as a critical contributor to accidents. The key element limiting SA for driving is hazard detection and projection. In driving, hazards are familiar objects poised to do something dangerous, such as a truck covering the edge of a cross walk. Learning is a matter of construing these familiar things in less familiar, task-relevant ways. In aviation, FPM hazards are primarily undesired states of the airplane relative to its energy or trajectory (and, rarely, objects out of place in the air or on the runway). That is, information about FPM hazards generally are not objects but abstract states, represented primarily on flight deck displays but sometimes only in the pilot’s head. However, for experienced airline pilots, this abstract representation may be quite familiar and be recognized as a hazard. In both domains, normal operations are unlikely to provide varied or comprehensive exposure to hazards. Further, a hazard may go unnoticed, particularly when it does not result in an incident or accident. Concentrated exposure to hazards in simulators is a valuable method to accelerate learning in a safe environment.

Several other similarities in domain structure or relevant training interventions may also ring true:

- For driving, and almost certainly for aviation, it is critical to identify which hazards are difficult to recognize, especially for less-experienced operators. It is very likely that in both domains it is difficult to anticipate and project emerging hazards, though projecting into the future is an important skill.
- Feedback about the overall level of performance is unlikely to help learning in aviation and was shown to be unhelpful for driving.
- Information about the nature of the hazard to be identified is important, so explanatory or evaluative descriptions accompanying events (not just success feedback) is useful. In aviation, this often occurs as debriefing after a longer session or scenario. Shorter descriptions for individual cases are effective for driver hazard training. The best use of explanations linked to example cases is unclear:
  - whether the explanation needs to be contingent on the individual’s behavior or can a more general response.
  - whether it can be generated by the learner in self-debriefing and self-explanation or whether it is best produced by an expert.
- In both domains, stating observations out loud as events unfold is a method used to improve or measure awareness; when proceduralized in aviation, these are called callouts. In both domains, there are questions and trade-offs concerning when such vocalizations help awareness of self or others, and when reporting observations

becomes a burdensome dual task and creates interference with work: how much talking is too much talking?

- Across domains, the skills limiting performance will change with training and experience, and identifying different skill levels so that the most appropriate training can be provided is likely to be valuable. However, both domains are concerned with providing large sets of people with an adequate, standardized skill level, so differentiation by large groups (individuals with or without a type rating in aviation, drivers with training permits) may be feasible, where individualized training may not.
- In RAPT, dual forms of representation are used. Both driver-view and schematic plan-view diagrams are used to explain the nature and means of addressing the hazard. We expect this would also be helpful in training SA, for example, diagrams showing how being above path might implicate ability to meet future targets, but we do not know of training that integrates pilot-view (simulator-view) and schematic presentation.

Differences are important as well. One difference between domains is the shorter response window in driving than in aviation. In driving, recognizing and responding to a hazard may need to be done in seconds, while minutes are often available in aviation. This is fortunate because much hazard processing in aviation likely requires slower, deliberative thinking, such as reasoning about how modes of the autoflight system control flight, or determining the feasibility of meeting a waypoint restriction. A very conspicuous difference between domains is the overall complexity and amount of training required for safe performance. Probably in large part due to the much greater simplicity of driving, the training of SA, or hazard anticipation, seems to be much farther advanced in this simpler domain. Successes here may provide suggestions for aviation.

7.5. Sports Performance Domain

7.5.1. Methods of Training and Results

Attention, awareness, dynamic understanding, and decision-making have been active research topics in sports, addressing both coaching and player roles (e.g., Debanne, Fontayne, & Bourbousson, 2014; Macquet & Stanton, 2014; for a review of SA from a sports perspective, see Meireles, Alves, & Cruz, 2018). For example, an observational study on individual sports performance (rowing and hammer throwing) found a complex pattern of content similarities and differences between coach and athlete based on their talk-aloud commentary (Macquet & Stanton, 2014). They did not suggest or evaluate a training intervention, however. While training is certainly important and monitoring an important component in many sports, for example, in life guarding, we found no investigation of outcomes (Yu & Chen, 2014).

Only a small body of the research in this domain addresses how required SA capabilities are learned or how they can be better trained. There are two strands of inquiry within this broad topic:

- awareness of other players
- how eye-fixations may be used to regulate attention and control

Note that in team sports such as soccer, the dynamic aspects of the situation are driven by players: the immediate past actions (trajectory of the ball), current configuration (who is open), and future capabilities (will that team member be able to complete this pass?), so awareness of players and their perspective constitutes much of the needed dynamic awareness and assessment.

Kermarrec and colleagues (2015, 2017; Kermarrec, Yohann Cardin, & Bossard, 2014) have explored uses of several training strategies to improve awareness and understanding, including what other players are seeing and considering. These methods include a variety of replay and commentary techniques, including interrupting the play for a quick side-line review of a configuration of particular interest. These methods have shown benefit for elite youth players. Somewhat analogous to the position of SA within CRM in aviation, particular methods for developing awareness may be nested within a broader program such as the 4P strategy: Positioning, Practicing, Picturing, and Post-analyzing (Kerrmarrec, 2017).

Gaze control has been shown to be correlated with successful performance and has become a useful training intervention in sports. Vickers's analysis of "quiet eye" fixations advocates sustained fixations prior to critical actions, though the optimal fixation location differs with the type of motor control, whether targeting, interceptive, or tactile (Piras & Vickers, 2011; Rienhoff, Tirp, Strauß, Baker, & Schorer, 2016; Vickers, 1996; Wood & Wilson, 2011). The association of a sustained "quiet eye" with better performance has been identified in many sports and can be trained effectively. For example, elite youth soccer players were trained to fixate the target (e.g., upper left or right in the net) prior to the run up for a penalty kick, with sustained fixation on the ball thereafter; that is, a long "quiet eye" period. This training resulted in improved performance and better performance when compared to a "normal practice" control (Wood & Wilson, 2011).

7.5.2. Relevant Lessons from the Domain

Gaze training in sports performance has been shown to be effective, raising the question about whether there is a relevant application in aviation. In sports performance, gaze control is:

- part of a sensory-motor control loop
- operating over milliseconds
- simplified by the extremely short and direct coupling

Possibly, there are analogies to manual flight control. However, the monitoring requirements of an airline pilot flying with a modern flight deck are very different. Rather than a steady, brief fixation on one target to regulate motor control, the pilot must sample across a considerable number of fixation points to meet changing information needs. Thus, the situations in which gaze training has been shown to be successful, either in gaze regulation or in impacting performance, differ considerably from FPM.

7.6. Cross-Domain Themes for Training Pilot Monitoring

7.6.1. The Situation Model is the Target of Training

Getting the content right is critical to training, and this can be defined at multiple levels. As depicted previously in Figure 9, the target of training should be: a) important for the work; b) undeveloped in the trainees; and c) trainable.

What information should be monitored? Identifying and teaching the relevant information to monitor is important. Thus, identifying unrecognized relevant information can lead to a dramatic jump in awareness and outcomes, even at an institutional level as in the study by Brady et al. in the medical domain. More typically, the relevant information is known to experienced personnel but not to those being trained.

What behaviors or skills can and should be trained? Eye movements and visual attention are probably not useful as direct training targets. First, this aspect of behavior may not be very trainable.

We found few attempts to train a scan pattern, and when it was tried, it failed. Changing where people look by changing what they are looking for may be a more effective strategy. We also found a small number of studies that specifically tried to train workload management or attention regulation, and these did not succeed in changing performance. "Low level" behavior may be constrained by intrinsic attention limits and require extensive, routine practice on stereotyped, repeating patterns to change.

Does training to build the connection between general knowledge and specific situations result in better SA/monitoring? Recall that we described that a study designed to train a scan pattern also provided (inexperienced) pilots with simple but accurate explanations about control and produced much better performance than those given only scan training. We refer to this as mental model training, where a mental model is a person's organized, general knowledge, typically about how something works, such as how autoflight works or the impact of weather on aircraft. Mental models provide schemas or frames for interpreting situations. Pilots typically have incomplete mental models.

We introduced the concept of a situation model in Section 3. While a mental model represents information at a fairly general level, a situation model is specific to a particular point in time. A situation model integrates the general model with the specifics of the current time (see Figure 13); for example, projecting a descent path based on knowledge of aircraft capabilities combined with the current airspeed, distance, and predicted wind. We think the distinction between a mental model and situation model is useful for understanding how people think in dynamic situations: a situation model is constantly updated. A pilot's situation model allows him/her to reason, explain, and make projections about the future. Accuracy of a situation model depends both on the accuracy of the general mental model(s) on which it is based and on the accuracy of information about the current state. The *situation model* relates to and extends ideas about *situation awareness*. We suggest that the process of retrieving a mental model and filling values from the current situation may be particularly difficult and may benefit particularly from training.

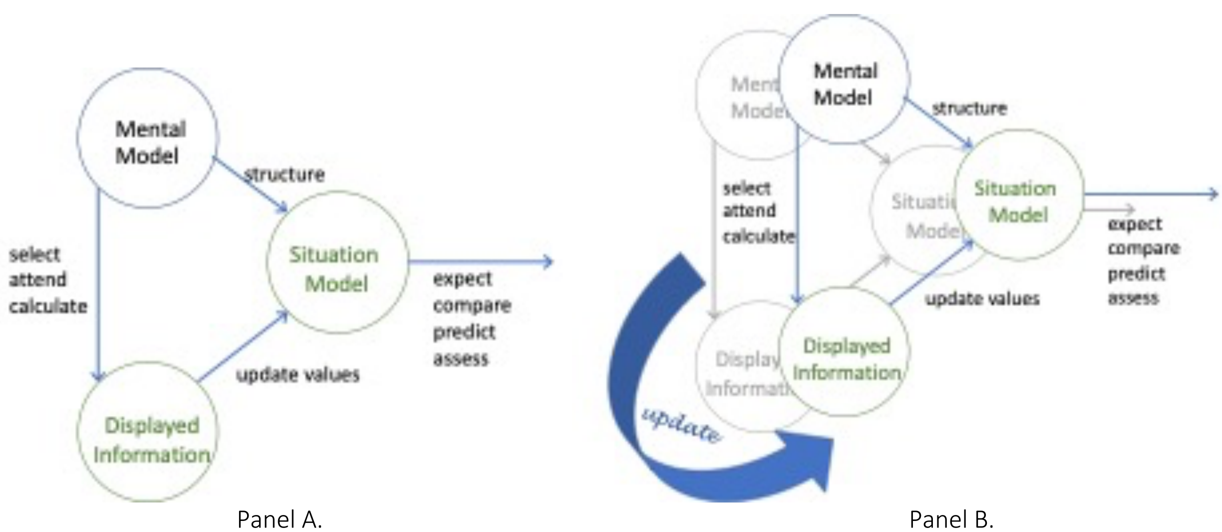

Figure 13. Panel A: Integrating displayed information with a mental model from long-term memory to build a situation model.

Panel B: Information is updated. Not drawn: information from the situation model can feed back to revise the mental model, in learning through experience.

We know of no research that has tried to train building a situation model, but two studies are related. One suggests that applying a general principle to a specific situation is difficult: the ACRM program reduced the need for pilots to build the link between general CRM principles and specific actions, by writing the relevant CRM actions into the procedures, and this improved performance on items related to monitoring and awareness (Holt et al., 2001). The second study found that more experienced pilots outperformed less experienced pilots in making predictions specifically on items that required integrating a general mental model with the specifics of the situation (Doane et al., 2004).

We suggest that training pilots' understanding of how things work (to build a long-term mental model) will often benefit monitoring, particularly where gaps in pilot understanding can be identified. Further, we suggest that training on how to integrate information about the current state with the model may be both difficult to do and helpful if accomplished.

7.6.2. Use an Integrated Structure to Organize Training

Structuring training around activity cycles of teaching concepts and theory, providing a briefing, conducting a simulation, and doing a debriefing (CBSD) is likely to be a powerful approach for organizing training (see Knecht, Muehlethaler, & Elfering, 2016). Improving the ability to build up a situation model and use it to understand the ongoing situation is a complicated learning goal. We think CBSD training cycles may be particularly effective for complex learning tasks such as this. Building a useful situation model involves multiple cognitive activities and might help pilots learn to:

- build up and retrieve an accurate mental model
- identify what information is important to sample from the current situation
- make the relevant observations
- integrate these with expectations from the mental model
- reason and predict using the situation model to assess whether events are unfolding as they should

While this training structure may be particularly valuable for complex training goals, we suggest it is a generally useful framework.

The importance of simulation is widely recognized, but framing training design this way provides several distinctive benefits. First, this emphasizes the relations between the four types of activity: all should be coordinated around a specific training objective. Second, it may be valuable to iterate through this cycle in relatively short "loops" so that familiarity and understanding can be built up in manageable units, without overwhelming the learner; of course, the cycle size can be adjusted to the complexity of the training goals. Part-task training across several topics can then be integrated into longer and more complex simulation cycles. Third, a large number of training plans can be developed within this framework, selecting or weighing different components of the cycle appropriately. For example, for a very simple training goal, it might be sufficient to rely heavily on presentation and discussion without simulation time for that topic. We discuss the applicability of CBSD further in Section 8.

7.6.3. Identify Efficient and Relevant Training Methods

Simulators play a role in learning. The cycle of reflective and in-action experience is a great benefit of the CBSD framework. Dynamic simulation is particularly helpful in dynamic domains. The person must respond with the pace relevant to the work. Higher-fidelity simulators also provide great value in providing high realism (and therefore easier transfer). They are also engaging and motivating. Simulators, however, are expensive. The studies reviewed here included a variety of simulation types and fidelity, often finding training impact from simple simulation, and occasionally finding as much benefit to SA from that simpler training tool. An important theme in training is identifying the least sophisticated simulation that is relevant to the training goals. This is an active research area not addressed here. Of note, microworlds and games may also be effective, motivating training environments where the dynamic characteristic of simulators is preserved but a narrower part of the task is trained. It is important not to think of simulators for training monitoring and SA as "procedure trainers" because the competencies to be learned will not be just procedures.

The learner also plays a role in learning. The learner role in a training activity can take different forms. While it is generally understood that an "active" learner is good (and the CBSD cycle encourages an active role), it can take many forms, and there are some interesting complications. Watching flight videos produced as much improvement in SA as did flying with yoke and rudder controls. Self-debriefing provided comparable benefit to debriefing by the expert; possibly the greater activity of self-organization and reviewing traded off against greater knowledge to produce comparable outcomes. Training focused on meta-cognitive skills is unusual as part of monitoring training, but like self-debriefing, it puts the learner in a self-evaluative position. In short, there is a complex and sometimes surprising relation between learner role and outcomes of training for SA.

7.6.4. Details of Method Choices Impact Outcomes

Training of topics related to monitoring may have interesting facilitating or interfering effects. On the one hand, it is important not to overwhelm the learner in a training session with too much material (Sweller, 1988). On the other hand, "desirable difficulty" may decrease performance during training but result in better retention and training (Soderstrom & Bjork, 2015). These trade-offs are explored in a variety of learning situations, but impact on pilot training is unknown. Similarly, the timing of feedback is an important variable. Immediate feedback is useful in many cases, in part because delay may produce forgetting. However, playback in debriefing may allow good remembering of the events, possibly leading to better situation models and assessment.

One important open question is whether or how training on one aspect of monitoring may benefit (or possibly interfere) with another. Consider a pilot monitoring who has a good situation model and organized structure for a set of questions, checks, and predictions; will such pilots be less vulnerable to lapses of attention or disengagement? The ESSAI study found that one effect that favored the ESSAI pilots was not from increased ESSAI performance on a second session, but that the ESSAI-trained group did not decline, as the control group did. Possibly, the high-level training impacted low-level likelihood of boredom, disengaging, or reducing effort.

8. Findings Regarding Effective Training of Monitoring for FPM

A discussion of effective training should begin with a description of how pilot performance has fallen short and the types of knowledge and skills that underlie skilled performance. In Section 2, we described monitoring for FPM and the ways in which it can fail. Specifically, we described the following failure types, stressing operational outcomes:

- Imminent Upset, in which the flightcrew loses awareness of airplane state, and the airplane is near or outside the bounds of a stable flight regime, perhaps entering an upset.
- Incorrect Airplane State or Configuration, in which the flightcrew fails to confirm that the airplane is appropriately configured (e.g., flight control surfaces, automation state) and has the correct flight path targets.

- Failing to meet flight path targets, in which the airplane is not currently achieving the cleared flight path targets or is in danger of not achieving future flight path targets.
- Failing to identify important changes in the flight environment, in which the flightcrew is unaware of important changes that will affect the flight path.
- Failing to maintain awareness of crew resources, in which flightcrew members have lost awareness of how monitoring for FPM is being accomplished.

We also identified these two failure types that were given less weight in our analysis:

- Failing to call out a deviation, in which the PM notices a deviation away from an FPM target but fails to call the PF's attention to it.
- Failing to intervene when a deviation is not being managed, in which the PF fails to intervene in order to ensure that appropriate corrective actions are initiated.

We also, in Section 3, laid out our understanding of the full range of knowledge and skills that underlie monitoring performance.

For task management, they are:

- *Strategic task management*: to manage operational tasks (and, therefore, workload) through planning to ensure there are ample resources for monitoring.
- *Attention switching*: to remain responsive to important changes in the operational environment and also ensure that periodic monitoring for FPM does not lapse.

For Understanding, they are:

- *Efficient information extraction*: to efficiently comprehend relevant data and information in the operational environment.
- *Operational knowledge* (including appropriate mental models).
- *Situation model*: to understand the airplane's current state relative to the operational environment.
- *Interface configuration*: to set up the flightdeck interface to facilitate monitoring.
- *Threat identification*: to identify and understand the FPM threats that need to be managed.

And these skills:

- *Problem identification*: to identify deviations from the expected state that need to be managed.
- *Intervention*: to take action to ensure that the deviation is being corrected.
- *Communication*: to develop a shared understanding of FPM objectives and threats; communication supports monitoring activities in a number of ways.

In the next section we identify the subset of these knowledge and skills that should be the focus of training.

8.1. Training Focus

The two major areas that training should focus on are task management and understanding.

8.1.1. Training for Task Management

Training should focus on ***strategic task management***, which is managing operational tasks (and, therefore, workload) through planning to ensure there are ample attentional resources for monitoring. Due to the many barriers to sustained attention (see Section 2.3), and because workload from other tasks can reduce a flight crew's ability to monitor FPM well, there is value in thinking ahead about when workload will increase, when FPM monitoring will become more intensive (due to a more dynamic flight phase), and how those two can combine to degrade FPM-related monitoring. Generally, the training objective is to aid pilots in planning and shifting tasks to free up attentional resources for FPM monitoring. The 2014 Flight Safety Foundation report on monitoring offered some useful concepts for helping pilots understand these demands.

Specific operational skills that may be relevant here are:

- planning for task management
- communication/briefings
- prompts and external reminders

Planning for task management should consider how workload is likely to change over the course of the flight and identify windows in which monitoring for FPM is likely to be threatened. Flight crew communication is an important element for executing a task management plan. Flight crew briefings are one technique to ensure there is a shared understanding of how tasks will be allocated. They can also be used to better focus on specific monitoring concerns, such as ensuring that the airplane is descending fast enough to make a waypoint. Ideally, the PF shares FPM objectives with the PM to make it easier for the PM to identify deviations from the intended goals.

Another potential aid to ensuring that the plan is executed are monitoring prompts—explicit, spoken prompts to check FPM. Because interruptions and distractions can be powerful forces on capturing attention, there needs to be an equally strong and reliable prompt for returning to FPM monitoring. Prompts are naturally generated by transition points between phases of flight and ATC communications on flight path (e.g., cleared for the approach). Outside of these prompts, the flight crew is responsible to generate others. Ideally, additional prompts are tied to other external events, which is a more reliable technique than hoping that a pilot will recall the need to check FPM. And these prompts should be associated with a set of FPM-related items.

Thus, skill here is not just the planning, but also identifying strategic means to generate external pointers back to the plan. Another aid available in some airplanes is reminding technology that generates a prompt/alert at a certain point in time, and these can also be shaped as reminders.

Training should *not* focus on ***attention switching***. As we have said above, attention switching, a meta-cognitive skill, does not lend itself to training. However, there may be some techniques for preventing switching failures. Specific operational skills that may be relevant are:

- communication regarding monitoring
- monitoring heuristics
- prompts

A flight crew allows each pilot to look out for the other, and by communicating your intentions to the other pilot, you create a second check on switching failures. Ideally, the other pilot can tell you to return to basic checks when you get too involved in a single task. More generally, this communication aids each pilot in understanding how the other pilot is allocating attention and for how long. Some pilots have developed heuristics for checking their own performance. We have talked to pilots who have developed rules for themselves, such as "if I cannot get the FMS entry correct after three tries, I move my attention back to other tasks." While this type of rule is also vulnerable to attentional failures, it may be effective for managing attention on less stressful or urgent tasks.

And, as with strategic task management, the creation of and use of prompts for monitoring can reduce the number of switching failures. The goal is to prevent a pilot from becoming overly focused on a single task when other items need attention as well.

8.1.2. Training to Enhance Understanding

This report on monitoring is different from previous reports (see Appendix A) due to our emphasis on the role of understanding or sensemaking in monitoring. In particular, we think training should focus on developing and using a situation model, the relevant operational knowledge, and threat identification.

The situation model is a rich mental representation of the current situation. A situation model integrates background knowledge with awareness of current indications to represent the pilot's understanding of the current situation. This allows a skilled pilot to generate expectations about the airplane's energy and trajectory in relation to the flight path targets and also allows projection into the future, thus supporting prediction and planning. Those expectations guide what information is monitored in the environment and support judgments about how well FPM is progressing.

Operational knowledge provides the foundation for a skilled pilot to develop a situation model, and it influences how the situation model is updated. Operational knowledge refers to a range of topics: airplane systems and indications, autoflight modes and behavior, energy management, expectations about airplane performance, typical ATC clearances, approach geometries, weather and its effects on performance and routing, airspace and airline procedures, and others. This knowledge is derived from training and experience, and it is largely organized into mental models and can include heuristics for calculations or assessing risk. Operational knowledge also allows the pilot to fill in the situation model with information that has not been observed. Operational knowledge identifies relationships between the various indications; e.g., transitioning into VNAV PTH on climb indicates that we must be crossing a waypoint restriction, which I can see on the LEGS page. Thus, operational knowledge leads to techniques for building and assessing an understanding of what is happening in FPM and for determining information relevance.

Threat identification refers to the recognition that some aspect of the current situation poses a threat to FPM and leads to an understanding of actions that are available to manage that threat. This is a form of operational knowledge that specifically focuses on problem anticipation and corrective actions. As a simple example, when ATC gives a "direct to" clearance that removes many track miles, there is a threat to meeting altitude and airspeed targets. You are now high on the path because of the shorter distance. A skilled pilot not only draws these inferences from the situation model but also updates the situation model to include the actions required to maintain FPM targets and how to monitor that progress. Skilled pilots will also modify the **interface configuration** to make sure the most appropriate displays are "on the surface" for monitoring the current situation.

Effective **communication** within the flight crew is a critical activity to support these monitoring skills. Briefings provide the opportunity to identify specific threats to FPM and articulate FPM objectives, which in turn points to objectives for monitoring: what indications are relevant and where are thresholds for further action. For example, as the PF determines a strategy for meeting altitude and airspeed targets, that strategy should be communicated to the PM so that the PM can watch for significant deviations from the plan.

Communication is not just about content; it also conveys attitudes about the role of the PM in FPM. The Captain, by being more explicit about FPM objectives and expectations for meeting those objectives, makes it easier for the PM to monitor the appropriate indications and know when to speak up about deviations that are not being managed. Ideally, the flight crew finds complementary roles in monitoring to ensure that the full range of indications is covered.

8.1.3. Poor Candidates for Training

Occasionally, in our discussions with pilots and airline trainers about monitoring, and in our reading, we encountered the idea that skilled pilots develop a **scanning pattern** (i.e., a sequence of fixations) that allows them to monitor effectively, and, if you could capture this scanning pattern, you could use it to teach less-experienced pilots to monitor effectively. However, we have found no evidence that such a scan pattern exists. A few studies (see the eye-tracking literature review in Appendix B) have been able to demonstrate that if you consider only core flight instruments—such as attitude, airspeed, altitude—you can identify a small number of scan patterns that pilots use. However, given the amount of information that is spread across the displays in the modern jet transport (PFD, NAV display, the flight plan, and the mode control panel), this narrowly defined scan pattern cannot account for the full range of FPM activities.

Our analysis indicates that pilots look at places where they think they can get information that is relevant to their current understanding of FPM. Because FPM needs change over the course of the flight and need to be balanced with other operational tasks, scanning the full set of displays is not likely to look like a recurring fixation pattern. We think the search for a scanning pattern should end; it should not be a focus of training.

Another theme that we encountered was that monitoring could be improved if pilots “tried harder” or focused more on monitoring. In this view, pilots do not lack monitoring skill but lack the discipline or motivation to apply it. In particular, this idea has been tied to LOC accidents in which the flightcrew seems to lose awareness of basic flight parameters, such as airspeed. We believe, however, that this type of failure is tied to the issues described in Section 2.3: high workload, significant distractions or interruptions, stress and fatigue, scattered indications, and the tendency to see what was expected. Training should not be dedicated to **vigilance or sustained effort**.

8.2. Training Approaches for Monitoring for FPM

Section 7.6 laid out a set of conclusions from the research literature about the ways in which monitoring, awareness, and understanding can be trained effectively. Although there is variability across domains and across individual studies, we believe that our conclusions about what techniques work can serve as steppingstones for building effective airline training for monitoring and awareness tied to FPM.

8.2.1. CBSD: An Integrated Approach to Monitoring Practice

The research results described in Section 7 demonstrate that a valuable pattern for structuring a training program for monitoring and awareness consists of cycles of these four activities:

- theory and concepts
- briefing
- practice (simulation-based)
- debriefing

While the value of flight simulation is widely recognized, the other activities are also important. These four activities need to be designed together to target specific concepts and related simulation scenarios to address suspected gaps in monitoring skills and knowledge.

We use CBSD to refer to this sequence of concept, briefing, simulation, and debriefing. Knecht et al. (2016) articulates how this cycle of activities can be an effective structure for training. To illustrate how and why this training structure can be effective in the design of training for monitoring, we use the example of trying to monitor and manage flight path when high above the descent path. Within this cycle, the specific material can vary with the level of the trainee. For example, when concepts are first introduced, it may be helpful for the briefing to alert the pilot to the indications that will be most important. Later training may require the pilot to prioritize what indications will require closest monitoring or may present the pilot with surprising scenarios.

Presenting theory and concepts through brief lecture units, reading, or discussion provides information on core concepts, knowledge, and vocabulary, ideally enriching relevant mental models. There are two broad types of theory and concepts that are valuable to train:

- operational knowledge
- understanding the process of building and using a situation model

Concerning operational knowledge, training provides (or reviews) the target concepts, theories, heuristics, and examples relevant to the scenarios. For managing the situation where the airplane is getting high above the descent path, relevant knowledge might include the following:

- Knowing what ATC or environmental prompts should trigger a concern.
- The 3-to-1 rule for quickly assessing your ability to descend to an altitude over a certain set of track miles.
- How VNAV will behave and which indications can help you track your position relative to the path.
- Other interface tools that can be used to judge distance if you are off path laterally as well.
- Reasonable expectations about airplane performance concerning descending and slowing down.

With a broad exposure to relevant knowledge, mental models will support understanding a wider range of situations. Training that focuses exclusively on procedures or a small set of nominal cases can result in “brittle” learning that is unlikely to generalize well to other cases. This more general training also helps the pilot build an organization for setting FPM objectives, generating expectations, anticipating potential threats, interpreting outcomes, and remembering experiences. An important part of building a useful mental model is learning what indications are relevant for understanding the behavior of a system or a situation. In our example, information tied to energy management will be central. More general mental models support the pilot in building situation models relevant to a broader range of situations.

Training can also focus directly on the basic theory and concepts of monitoring. Having an explicit theory can help link together and generalize various experiences and "rules of thumb" that trainees may already have, as well as providing a framework for managing new situations. Because it is impossible to directly train all situations, providing a general understanding of monitoring processes can help pilots when encountering the unfamiliar and also provide a shared vocabulary for pilot communication about monitoring. For example, the process of identifying where there are gaps or inconsistencies in the situation model (refer back to Figure 1) and hence what information is needed can be broken out and illustrated. Reviewing for inconsistent or unexpected values can be helpful here, as can reference to a monitoring plan set up in briefing.

Second in CBSD is a briefing, which serves to orient the pilot to the simulation scenario. The briefing can also direct the pilot's attention to the learning goal. Early in training, it may be helpful to point out the aspects of the upcoming situation that are particularly relevant to managing it well. The briefing can also provide a scaffold for learning and reduce the learner's cognitive load, allowing the pilot to focus on a relevant subset of the information encountered and the actions taken in the upcoming scenario. In our example, the briefing could preview items such as determining how long you can be held high without jeopardizing your ability to meet an altitude constraint, how to assess your ability to slow down to meet an airspeed target, and how to assess whether the VNAV mode will return to VNAV PTH. Later in training, a trainer-provided briefing that describes the flight without guiding the pilot's attention may be valuable, but with reminders about the monitoring cycle and about the responsibilities of each pilot for communicating. The pilots may be responsible for part of the briefing, and discussion of their briefing included in the debriefing.

Third, a simulated scenario—performance of an operationally relevant task—is the heart of the training activity. Per our example, the pilot or flightcrew would fly a descent and manage it as the airplane gets high above the path, perhaps due to being slowed down by ATC. Each scenario provides an opportunity to construct a situation model, to identify what information is relevant, and to update the situation model with these relevant values and use it to make decisions about the ongoing course of action. Some evidence suggests that integrating the specific situation with a more generic mental model to build an accurate situation model is particularly hard (Doane et al., 2004), and, therefore, structured practice through the simulation is valuable. Because of the obvious similarity between the work environment and realistic simulated scenarios, the scenario work allows the pilot to transition from "schoolhouse to wheelhouse," i.e., transfer into the operational work.

Finally, the simulation or practical case is followed by debriefing, which allows the pilot to think back over events without the time pressure of the need for action. In debriefing, the pilot can identify or recall problems or questions encountered. Debriefing also provides an opportunity for feedback by the pilot, from the other crew, and from an instructor; in particular, feedback about process is more valuable than feedback about outcomes. For many types of cognitive skills, and particularly monitoring, feedback that only informs the user that an error occurred does not make learning efficient or easy. Process feedback also allows identifying whether successes were 'fortunate accidents' or were based on a generalizable understanding. Process feedback can be provided at multiple levels. Pilots can be asked to recall parts of the scenario where they stepped through the monitoring cycle: what gap was identified, what relevant information found, and what risk assessed. Feedback is useful that identifies the relevant indications important at different points in the scenario, as well as comparison of the actual and the expected values. Discussion or review during debriefing can also "debug" gaps or errors in the pilot's mental model, such as elaborating pilot understanding of autoflight mode behavior. If the debriefing is structured so that the pilot(s) can efficiently provide much of the feedback themselves, this active learning can enhance retention. Typically, there will be places where instructor feedback is also crucial.

The effects of feedback timing are well-understood: typically, quicker feedback leads to faster immediate improvement in performance. However, the benefits of immediate feedback may be outweighed by other qualities of a more-delayed debriefing, particularly when CBSD sequences are relatively short. In debriefing, if feedback is provided over a meaningful unit of performance, it allows the learner to actively retrieve and reflect on the scenario, and it provides the information at a point where the cognitive load is low. Thus, although immediate feedback (during the scenario) may lead to a quick increase in performance, feedback deferred until debriefing should improve transfer and generalization of the target skills and knowledge.

Debriefings are usually carried out by an expert who reviews the learner's performance. Interestingly, learners also benefit from structured self-debriefing. Further, self-debriefing may be a particularly valuable "learning to learn" skill that a pilot can apply to capitalize ongoing experiences. Learners also benefit from listening to the debriefings of others' performance.

Multiple cycles of CBSD increase its benefits. In particular, a second cycle of briefing, simulated flight(s) analogous to the first, and debriefing can provide a particularly valuable opportunity to consolidate and extend learning. Short elements of the first concepts-and-theory part of the cycle can be reviewed in the briefings. The duration and content of an individual scenario is important, as is the coverage of the scenario set. Ideally, similar monitoring issues can be posed in multiple scenarios. Working with multiple examples of the same principles aids the learner in identifying the principle and recognizing the examples where the principle is relevant. Designing training as a set of shorter CBSD cycles with shorter scenarios allows more frequent exposure to each part of the CBSD cycle. Structuring training in CBSD cycles also can provide a form of part-whole training, where early scenarios expose the learner to a small part of monitoring for the domain. As learning progresses, longer sequences may be helpful, particularly for managing monitoring tasks in the context of other activities.

8.2.2. Training Good Practices in FPM Communication

In our interviews with pilots, it became clear that pilots did not always provide good communication about their understanding of how the flight path was being managed. Pilots may not express their concerns about potential FPM threats, their assessments of making altitude and airspeed restrictions, or their developing plans for meeting a clearance. Room for improvement was illustrated when we conducted the cognitive task analysis-type interviews described in Section 4.B. While these pilots were able to describe how they think about FPM, they also made it clear that this thinking often is not expressed in communications within the flight deck.

Therefore, a useful training approach may be to provide or impose a structure for flight crew communication that covers:

- current FPM objective(s)
- current potential hazards or threats to FPM
- relevant indications to monitor
- expected behavior—e.g., for autoflight or airplane performance—and signs that things are not going as planned
- trigger points, or the point at which the PM should ensure that the PF is aware of a deviation

An example might include this content, as hypothetically spoken by the PF:

- "Our next objective is to cross (waypoint) at or below 17,000 and 260 kts."
- "We started down a bit past the FMS top of descent point, and we have a tailwind we didn't expect."
- "Please keep an eye on our position above path on the vertical deviation indicator, and monitor how quickly we are losing airspeed."
- "I think we need to get down before we can bleed off airspeed, but I hope we are close to our target airspeed by 10 nm before the waypoint. Also, we are in VNAV SPD now, but I expect to revert back to VNAV PTH soon."
- "If we are not back to VNAV PTH within 15 nm before the waypoint, let me know."

Clearly, this type of exchange goes well beyond what one might hear in a descent briefing. Using this type of exchange during training scenarios models the kinds of information that flight crew communications should cover. There are several potential benefits of this type of exchange:

- It reveals the current understanding and thinking of the PF, which then allows the PM to point out any inaccuracies or differences of opinion.
- It creates a shared situation model, especially concerning what aspects of FPM may be threatened and PF's intent of how to manage it.
- It explicitly calls out the flight deck information that is relevant.
- It establishes an agreed-upon threshold for when the PM needs to call out a deviation to the PF.

To implement this type of training, an airline will need to fully specify the desired content and structure of this "FPM check-in." Training should be scenario-oriented (but may not need a simulator) and allow pilots to practice using the structure, identifying the important FPM content, and how to keep it short but informative. The PM, after hearing the PF's download of information, should repeat back his/her own version to ensure that there is a shared understanding. The PM should also identify a way to supplement the check-in or ask questions to clarify. An open question is the desirable degree of flexibility encouraged for the crew vs degree of structure stipulated by the airline. Of note, the CBSD format described in B.1 may provide an appropriate structure for training here.

A potential benefit to this approach is that it replaces FMA call-outs with a more meaningful check. A concern, expressed by quite a few pilots and confirmed by several studies (e.g., Mumaw et al., 2000), is that FMA call-outs fail to convey the significance of the current automation state. Ideally, pilot communications around monitoring can enhance understanding for both PF and PM. However, a concern is that many pilots will not have sufficient understanding to generate a concise "check in" initially, and an objective of this training is to force trainees to think this way and improve their own understanding.

It is worth pointing out that three recent (but unpublished) training intervention studies have taken an approach similar to that described here; specifically, improving knowledge to generate accurate expectations about relevant FPM information. A study by Mauro and Barshi (personal communication, 2019) trained airline pilots to generate more accurate expectations for autoflight modes, hoping that pilots could anticipate and confirm that the appropriate mode was engaged. A study by Billman and others at NASA trained airline pilots to better use flight deck indications to manage FPM tasks in which monitoring can be difficult. Information relevance is a major element of that training. Similarly, a recently completed study by Peyle and Stewart (personal communication, 2019) combined a structured method to prompt monitoring with training to improve understanding and expectations about autoflight state. In all of these studies, monitoring is seen as understanding or sensemaking that relies on informed expectations about how the system will behave.

8.3. Final Thoughts on the Role of Training for Monitoring

Our focus in this report was on improving monitoring and awareness through training interventions. However, we feel the need to point out that training is just one leg of the human factors three-legged stool. The primary influences on performance of the human operator in a system setting are:

- *Training*: A formal method for identifying and improving task-relevant knowledge and skills.
- *Operational procedures*: The set of formal operational documents that guide some operational activities. Aviation has developed a number of normal and non-normal checklists, as well as other well-defined procedures (e.g., "flows"), that impose a structure on how pilots gather data or perform cross-checks. These are geared largely toward confirming appropriate airplane state or configuration and ensuring flight path targets are entered correctly. Some airlines even shape pilot communication with techniques like VVM (verbalize, verify, monitor) to encourage them to share intentions and follow through on monitoring progress toward FPM objectives.
- *Interface design*: The interface, ideally, can influence which information is most salient or most easily noticed, and also how well it conveys the broader system state. Older system interfaces—especially for nuclear power plants—used a single-sensor, single-indicator design in which each indication had a dedicated location on a large display space. Operators would need to know where each indication was located and when to look at it. As technology has dramatically altered system interfaces, there is now potential for changes in:
  - Presence and location; e.g., indications might be grayed out or suppressed when they are not available or relevant, or pop-up when they are relevant; indications may be moved more centrally or pushed on to an overview display. The goal here is to highlight the subset of indications or information that is easiest for the operator to find or notice.
  - Context; e.g., an indication/system variable can be placed in an appropriate operational context, such as appropriate boundary or threshold values, to enhance the operator's ability to interpret its meaning or significance.

– Salience; e.g., various visual and auditory schemes can be applied to the interface to enhance the salience of certain information when that information changes or becomes more important to notice.

The point is that there are various approaches for enhancing pilot/operator awareness and all three of these mechanisms should be used to improve pilot performance. Further, designing these three in a coordinated way would improve their overall effectiveness.

References

Alken, A., Luursema, J.-M., Weenk, M., Yauw, S., Fluit, C., & Van Goor, H. (2018). Integrating technical and non-technical skills coaching in an acute trauma surgery team training: Is it too much? *The American Journal of Surgery*, *216*(2), 369–374. https://doi.org/10.1016/j.amjsurg.2017.08.011

Anders G. (2001). Pilot's attention allocation during approach and landing—Eye and head-tracking research in an A330 full flight simulator. Presentation at the 11th International Symposium on Aviation Psychology. Columbus (OH): Ohio State University.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*, Issue 6, pp. 775-779.

Banbury, S., Dudfield, H., Hoermann, H.-J., & Soll, H. (2007). FASA: Development and validation of a novel measure to assess the effectiveness of commercial airline pilot situation awareness training. *The International Journal of Aviation Psychology*, *17*(2), 131–152. https://doi.org/10.1080/10508410701328557

Bektas, K., Knecht, C.P., Ferrari, F., Spillmann, K.P.C., & Muehlethaler, C.M. (2018). Improved pilot training using head and eye tracking system. *ETH Zurich*. https://doi.org/10.3929/ethz-b-000222482

Bellenkes, A.H. (1999). *The use of expert pilot performance models to facilitate cockpit visual scan training*. University of Illinois at Urbana-Champaign.

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.

Billinghurst, S., Morgan, C., Rorie, R.C., Kiken, A., Bacon, L., Vu, K.-P., L., … Battiste, V. (2011). Should students learn general Air Traffic Management skills before NextGen tools. Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting.

Billman, D., Mumaw, R.J., & Feary, M. (2020). Methods for Evaluating the Effectiveness of Programs to Train Pilot Monitoring. NASA/TM-20210000045.

Björklund, C.M., Alfredson, J., & Dekker, S.W.A. (2006). Mode monitoring and call-outs: An eye-tracking study of two-crew automated flight deck operations. *The International Journal of Aviation Psychology, 16*, 263-275.

Black, S.A., Nestel, D.F., Kneebone, R.L., & Wolfe, J.H.N. (2010). Assessment of surgical competence at carotid endarterectomy under local anaesthesia in a simulated operating theatre. *British Journal of Surgery*, *97*(4), 511–516. https://doi.org/10.1002/bjs.6938

Boet, S., Bould, M. D., Bruppacher, H. R., Desjardins, F., Chandra, D. B., & Naik, V. N. (2011). Looking in the mirror: Self-debriefing versus instructor debriefing for simulated crises, *Critical Care Medicine*, *39*(6), 1377–1381. https://doi.org/10.1097/CCM.0b013e31820eb8be

Bolstad, C.A., Endsley, M.R., Howell, C.D., & Costello, A.M. (2002). General aviation pilot training for situation awareness: An evaluation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46(1), 21–25. https://doi.org/10.1177/154193120204600105

Bolstad, C., Ensdley, M.R., Costello, A.,M., & Howell, C.D. (2010). Evaluation of computer-based situation awareness training for general aviation pilots. *The International Journal of Aviation Psychology*, *20*(2). Retrieved from https://doi.org/10.1080/10508414.2010.487013

Brady, P.W., Muething, S., Kotagal, U., Ashby, M., Gallagher, R., Hall, D., Wheeler, D.S. (2013). Improving situation awareness to reduce unrecognized clinical deterioration and serious safety events. *PEDIATRICS*, *131*(1), e298–e308. https://doi.org/10.1542/peds.2012-1364.

Briggs, A., Raja, A.S., Joyce, M.F., Yule, S.J., Jiang, W., Lipsitz, S.R., & Havens, J.M. (2015). The role of nontechnical skills in simulated trauma resuscitation. *Journal of Surgical Education*, *72*(4), 732–739. https://doi.org/10.1016/j.jsurg.2015.01.020.

Carbonell, J.R. (1966). A queueing model for many-instrument visual sampling. *IEEE Transactions on Human Factors in Electronics, 7*, 157-164.

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.

CAST (2014). Airplane state awareness joint safety analysis team. Located at https://www.skybrary.aero/bookshelf/books/2999.pdf

Catchpole, K.R., Dale, T.J., Hirst, D.G., Smith, J.P., & Giddings, T.A.E.B. (2010). A multicenter trial of aviation-style training for surgical teams. *Journal of Patient Safety*, *6*(3), 180–186. https://doi.org/10.1097/PTS.0b013e3181f100ea.

Chapman, P., Underwood, G., & Roberts, K. (2002). Visual search patterns in trained and untrained drivers. Transportation Research Part F, 5, 157–167. doi:10.1016/S1369-8478(02)00014-1

Cooper, S., Porter, J., & Peach, L. (2013). Measuring situation awareness in emergency settings: A systematic review of tools and outcomes. *Open Access Emergency Medicine*, 1. https://doi.org/10.2147/OAEM.S53679

Crandall, B., Klein, G., & Hoffman, R.R. (2006). Working minds: A practitioner's guide to cognitive task analysis. Cambridge, MA: MIT Press.

Crundall, D., Chapman, P., Trawley, S., Collins, L., van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident Analysis & Prevention*, *45*, 600–609. https://doi.org/10.1016/j.aap.2011.09.049

Debanne, T., Fontayne, P., & Bourbousson, J. (2014). Professional handball coaches management of players' situated understanding during official games. *Psychology of Sport and Exercise*, *15*(6), 596–604. https://doi.org/10.1016/j.psychsport.2014.06.008

Dedy, N. J., Bonrath, E. M., Zevin, B., & Grantcharov, T. P. (2013). Teaching nontechnical skills in surgical residency: A systematic review of current approaches and outcomes. *Surgery*, *154*(5), 1000–1008. https://doi.org/10.1016/j.surg.2013.04.034

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. DOI: 10.1145/2702123.2702521

Dehais, F., Behrend, J., Peysakhovich, V., Causse, M., & Wickens, C.D. (2017). Pilot flying and pilot monitoring's aircraft state awareness during go-around execution in aviation: A behavioral and eye-tracking study, *The International Journal of Aerospace Psychology, 27*:1-2, 15-28, DOI: 10.1080/10508414.2017.1366269

Dill, E.T, & Young, S.D. (2015). Analysis of eye-tracking data with regards to the complexity of flight deck information automation and management - Inattentional blindness, system state awareness, and EFB usage, 15th AIAA Aviation Technology, Integration, and Operations Conference, AIAA AVIATION Forum, (AIAA 2015-2901) https://doi.org/10.2514/6.2015-2901

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making, 1*, 271-285.

Doane, S. M., Sohn, Y. W., & Jodlowski, M. T. (2004). Pilot ability to anticipate the consequences of flight actions as a function of expertise. *Human Factors, 46*(1), 92–103. https://doi.org/10.1518/hfes.46.1.92.30386

Dogan, E., Steg, L., Delhomme, P., & Rothengatter, T. (2012). The effects of non-evaluative feedback on drivers' self-evaluation and performance. *Accident Analysis and Prevention, 45*, 522–528. doi:10.1016/j.aap.2011.09.004

Durso, F. T., & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 137–154). Aldershot, UK: Ashgate.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 65–84. https://doi.org/10.1518/001872095779049499

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Ephrath, A.R., Tole, J.R., Stephens, A.T., & Young, L.R. (1980). Instrument scan – Is it an indicator of the pilot's workload? Proceedings of the Human Factors Society 24th Annual Meeting (pp. 257–258). Santa Monica, CA: Human Factors and Ergonomics Society.

FAA (1996). Human factors team report: The interfaces between flightcrews and modern flight deck systems. Washington, DC: FAA.

FAA (2013). Operational use of flight path managements systems; Final report of the performance-based operations Aviation Rulemaking Committee / Commercial Aviation Safety Team Flight Deck Automation Working Group. Washington, DC: FAA.

Fioratou, E., Flin, R., Glavin, R., & Patey, R. (2010). Beyond monitoring: Distributed situation awareness in anaesthesia. *British Journal of Anaesthesia*, *105*(1), 83–90. https://doi.org/10.1093/bja/aeq137

Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review, 9*, 24-29.

Flight Safety Foundation (2014). A practical guide for improving flight path monitoring. Alexandria, VA: FSF.

Flin, R. (2004). Identifying and training non-technical skills for teams in acute medicine. *Quality and Safety in Health Care*, *13*(suppl_1), i80–i84. https://doi.org/10.1136/qshc.2004.009993

Flin, R., Yule, S., Paterson-Brown, S., Maran, N., Rowley, D., & Youngson, G. (2007). Teaching surgeons about non-technical skills. *The Surgeon*, *5*(2), 86–89. https://doi.org/10.1016/S1479-666X(07)80059-X

Flin, R., Martin, L., Goeters, K.-M., Hormann, H.-J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety*, *3*(2), 95–117.

Froger, G., Blättler, C., Dubois, E., Camachon, C., & Bonnardel, N. (2018). Time-interval emphasis in an aeronautical dual-task context: A countermeasure to task absorption. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *60*(7), 936–946. https://doi.org/10.1177/0018720818783946

Gaba, D. M., Howard, S. K., Flanagan, B., Smith, B. E., Fish, K. J., & Botney, R. (1998). Assessment of clinical performance during simulated crises using both technical and behavioral ratings: *Anesthesiology*, *89*(1), 8–18. https://doi.org/10.1097/00000542-199807000-00005

Gillespie, B. M., Harbeck, E., Kang, E., Steel, C., Fairweather, N., Panuwatwanich, K., & Chaboyer, W. (2017). Effects of a brief team training program on Surgical Teams' Nontechnical Skills: An interrupted time-series study. *Journal of Patient Safety*, 1. https://doi.org/10.1097/PTS.0000000000000361

Graafland, M., Schraagen, J. M. C., Boermeester, M. A., Bemelman, W. A., & Schijven, M. P. (2015). Training situational awareness to reduce surgical errors in the operating room. *British Journal of Surgery*, *102*(1), 16–23. https://doi.org/10.1002/bjs.9643

Graafland, M., Bemelman, W. A., & Schijven, M. P. (2017). Game-based training improves the surgeon's situational awareness in the operation room: A randomized controlled trial. *Surgical Endoscopy*, *31*(10), 4093–4101. https://doi.org/10.1007/s00464-017-5456-6

Haber, J.A., Ellaway, R.H., Chun, R., & Lockyer, J.M. (2017). Exploring anesthesiologists' understanding of situational awareness: A qualitative study. *Canadian Journal of Anesthesia/Journal Canadien d'anesthésie*, *64*(8), 810–819. https://doi.org/10.1007/s12630-017-0904-2

Hah, S., Willems, B., Mueller, G., Johnson, D.R., Schulz, K., DiRico, J., Hallman, K., Maliko-Abraham, H., O'Brien, T.G., Alvidrez, S., Bastholm, R., Dworsky, M., Fincannon, T., Bassett, P., & Hailston, T. (2017). Human-in-the-loop simulation experiment of integrated arrival/departure control services for NextGen operational improvement. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society.

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

Hänsel, M., Winkelmann, A. M., Hardt, F., Gijselaers, W., Hacker, W., Stiehl, M., … Müller, M. P. (2012). Impact of simulator training and crew resource management training on final-year medical students' performance in sepsis resuscitation: A randomized trial. *Minerva Anestesiologica*, *78*(8), 901–909.

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.

Hauland, G. (2008). Measuring individual and team situation awareness during planning tasks in training of en route Air Traffic Control. *The International Journal of Aviation Psychology*, *18*(3), 290–304. https://doi.org/10.1080/10508410802168333

Helmreich, R.L., Merritt, A.C., & Wilhelm, J.A. (1999). The evolution of Crew Resource Management training in commercial aviation. *International Journal of Aviation Psychology*, *9*(1), 19-32.

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

Hoermann, H.-J., Banbury, S., Blokzijl, C., Dudfield, H., Lamers, J., Lehmann, O., … Soll, H. (2003). Experimental Validation ESSAI/DLR&Q_Q/WPR/WP5/2.0—Status: National Aerospace Laboratory (NLR), Netherlands: ESSAI -Enhanced Safety through Situation Awareness Integration in training.

Hoermann, H-J., & Soll, H. (2004). Training of situation awareness and threat management techniques. In *Practice and Research*. *Aviation Psychology*. Retrieved from https://www.researchgate.net/publication/224794538_Training_of_situation_awareness_and_threat_management_techniques

Hoermann, H-J., Soll, H., Banbury, S., & Dudfield, H. (2004). Effects of situation awareness training on flight crew performance. In *Measures and Application*. *A Cognitive Approach to Situation Awareness: Theory*. Aldershot/UK: Ashgate.

Hoermann, H.-J., Soll, H., Banbury, S., Dudfield, H., Blokzijl, C., Bieder, C., … Polo, L. (2003). A validated recurrent training programme for Situation Awareness and Threat Management ESSAI/NLR/WPR/WP6/2.0. National Aerospace Laboratory (NLR), Netherlands.

Holder, B., Jacobsen, A., Nikolic, M., & Whitlow, S, (2016). Flight deck task management. FAA White Paper; Volpe Task Order: T0026; Volpe Project: FB22C1 PY277

Holt, R. W., Boehm-Davis, D. A., & Hansberger, J. T. (2001). *Evaluation of Proceduralized CRM at a Regional and Major Carrier*. Retrieved from https://www.researchgate.net/publication/237580032_Evaluation_of_Proceduralized_CRM_at_a_Regional_and_Major_Carrier

Horswill, M. S. (2016a). Hazard perception in driving. *Current Directions in Psychological Science*, *25*(6), 425–430. https://doi.org/10.1177/0963721416663186

Horswill, M. S. (2016b). Improving fitness to drive: The case for hazard perception training: Hazard perception training. *Australian Psychologist*, *51*(3), 173–181. https://doi.org/10.1111/ap.12132

Horswill, M. S., Falconer, E. K., Pachana, N. A., Wetton, M., & Hill, A. (2015). The longer-term effects of a brief hazard perception training intervention in older drivers. *Psychology and Aging*, *30*(1), 62–67. https://doi.org/10.1037/a0038671

Huettig, G., Anders, G., & Tautz, A. (1999). Mode awareness in a modern glass cockpit: Attention allocation to mode information. In R. Jensen (Ed.), Proceedings of the 1999 Ohio State University Aviation Psychology Conference. Dayton, OH: Ohio State University.

Hull, L., Arora, S., Aggarwal, R., Darzi, A., Vincent, C., & Sevdalis, N. (2012). The impact of nontechnical skills on technical performance in surgery: A systematic review. *Journal of the American College of Surgeons*, *214*(2), 214–230. https://doi.org/10.1016/j.jamcollsurg.2011.10.016

Hutchins, E.J., Holland, J.D., & Norman, D.A. (1985). Direct manipulation interfaces. *Human-Computer Interaction, 1*, pp. 311-338

International Civil Aviation Organization. (1994). Safety analysis: Human factors and organizational issues in controlled flight into terrain (CFIT) accidents, 1984–1994. Montreal, Quebec, Canada: ICAO.

International Civil Aviation Organization. (2016). Guidance material for improving flightcrew monitoring. Montreal, Quebec, Canada: ICAO.

Isler, R. B., Starkey, N. J., & Sheppard, P. (2011). Effects of higher-order driving skill training on young, inexperienced drivers' on-road driving performance. *Accident Analysis & Prevention*, *43*(5), 1818–1827. https://doi.org/10.1016/j.aap.2011.04.017

Jankouskas, T. S., Haidet, K. K., Hupcey, J. E., Kolanowski, A., & Murray, W. B. (2011). Targeted crisis resource management training improves performance among randomized nursing and medical students: *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, *6*(6), 316–326. https://doi.org/10.1097/SIH.0b013e31822bc676

Jarvis, S.R. (2017). Concurrent pilot instrument monitoring in the automated multi-crew airline cockpit. *Aerospace Medicine and Human Performance,* Vol 88, No. 12 pp 1100-1106.

Jung, J. J., Borkhoff, C. M., Jüni, P., & Grantcharov, T. P. (2018). Non-Technical Skills for Surgeons (NOTSS): Critical appraisal of its measurement properties. *The American Journal of Surgery*, *216*(5), 990–997. https://doi.org/10.1016/j.amjsurg.2018.02.021

Kalar, D.J., Liston, D., Mulligan, J.B, Beutter, B., & Feary, M. (2016) Considerations for the use of remote gaze tracking to assess behavior in flight simulators. NASA Technical Memorandum NASA/TM-2016-219424.

Kearns, S. (2010). *An Instructional Strategy for Asynchronous e-Learning of Pilot Safety Skills*.

Kearns, S. (2011). Online single-pilot resource management: Assessing the feasibility of computer-based safety training. *The International Journal of Aviation Psychology*, *21*(2), 175–190. https://doi.org/10.1080/10508414.2011.556499

Kermarrec, G. (2015). Enhancing tactical skills in soccer: Advances from the naturalistic decision-making approach. *Procedia Manufacturing*, *3*, 1148–1156. https://doi.org/10.1016/j.promfg.2015.07.192

Kermarrec, G. (2017). *Tactical skills training in team sports: Technological supports for the 4P strategy*. Springer.

Kermarrec, G., Yohann Cardin, & Bossard, C. (2014). Shared understanding and coordination in team sports contribution of viewpoints changes and shared information displays for team situation awareness training. *Unpublished*. https://doi.org/10.13140/2.1.2191.3605

Khoo, Y., & Mosier, K. (2005). Searching for cues: An analysis of factors affecting the decision making process of regional airline pilots. In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (pp. 578–581). Santa Monica, CA: Human Factors and Ergonomics Society

Kim, J. H. (2018). The effect of metacognitive monitoring feedback on performance in a computer-based training simulation. *Applied Ergonomics*, *67*, 193–202. https://doi.org/10.1016/j.apergo.2017.10.006

Kirkpatrick, D.L. (1959). Techniques for evaluation training programs. *Journal of the American Society of Training Directors, 13*, 21-26.

Knecht, C. P., Muehlethaler, C. M., & Elfering, A. (2016). Nontechnical skills training in Air Traffic Management including computer-based simulation methods: From scientific analyses to prototype training. *Aviation Psychology and Applied Human Factors*, *6*(2), 91–100. https://doi.org/10.1027/2192-0923/a000103

Lee Chang, A., Dym, A. A., Venegas-Borsellino, C., Bangar, M., Kazzi, M., Lisenenkov, D., … Eisen, L. A. (2017). Comparison between simulation-based training and lecture-based education in teaching situation awareness: A randomized controlled study. *Annals of the*

American Thoracic Society, *14*(4), 529–535. https://doi.org/10.1513/AnnalsATS.201612-950OC

Lefrancois, O., Matton, N., Gourinat, Y., Peysakhovich, V., & Causse, M. (2016). The role of pilots' monitoring strategies in flight performance. Paper presented at the European Association for Aviation Psychology Conference, Cascais, Portugal.

Lefrancois, O., Matton, N., Peysakhovich, V., & Causse, M. (2018). Pilot flying vs pilot monitoring: Study of gaze allocation during dynamic and critical flight phases. Proceedings of the 33rd European Association for Aviation Psychology conference, Dubrovnik, Croatia.

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

Loukopoulos, L.D., Dismukes, R.K., & Barshi, I. (2008). The multi-tasking myth: Handling complexity in real-world operations. Ashgate.

Mack, A., & Rock, I. (1998). *MIT Press/Bradford Books series in cognitive psychology. Inattentional blindness.* Cambridge, MA, US: The MIT Press.

Macquet, A.-C., & Stanton, N. A. (2014). Do the coach and athlete have the same «picture» of the situation? Distributed Situation Awareness in an elite sport context. *Applied Ergonomics*, *45*(3), 724–733. https://doi.org/10.1016/j.apergo.2013.09.014

Malakis, S., Kontogiannis, T., & Psaros, P. (2014). Monitoring and evaluating failure-sensitive strategies in air traffic control simulator training. Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '14, 1–6. https://doi.org/10.1145/2674396.2674462

McCulloch, P., Mishra, A., Handa, A., Dale, T., Hirst, G., & Catchpole, K. (2009). The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Quality and Safety in Health Care*, *18*(2), 109–115. https://doi.org/10.1136/qshc.2008.032045

Meir, A., Borowsky, A., & Oron-Gilad, T. (2014). Formation and evaluation of Act and Anticipate Hazard Perception Training (AAHPT) intervention for young novice drivers. *Traffic Injury Prevention*, *15*(2), 172–180. https://doi.org/10.1080/15389588.2013.802775

Meireles, L., Alves, L., & Cruz, J. (2018). Conceptualization and measurement of individual situation awareness (SA) in expert populations across operational domains: A systematic review of the literature with a practical purpose on our minds. Proceedings of the Human Factors and Ergonomics Society 2018 Annual Meeting.

Mishra, A., Catchpole, K., Dale, T., & McCulloch, P. (2008). The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surgical Endoscopy*, *22*(1), 68–73. https://doi.org/10.1007/s00464-007-9346-1

Mishra, A., Catchpole, K., & McCulloch, P. (2009). The Oxford NOTECHS system: Reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Quality and Safety in Health Care*, *18*(2), 104–108. https://doi.org/10.1136/qshc.2007.024760

Moorthy, K., Munz, Y., Adams, S., Pandey, V., & Darzi, A. (2005). A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre: *Annals of Surgery*, *242*(5), 631–639. https://doi.org/10.1097/01.sla.0000186298.79308.a8

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.

Muehlethaler, C. M., & Knecht, C. P. (2016). Situation awareness training for general aviation pilots using eye tracking. *IFAC-PapersOnLine*, *49*(19), 66–71. https://doi.org/10.1016/j.ifacol.2016.10.463

Müller, M. P., Hänsel, M., Stehr, S. N., Fichtner, A., Weber, S., Hardt, F., … Koch, T. (2007). Six steps from head to hand: A simulator based transfer oriented psychological training to improve patient safety. *Resuscitation*, *73*(1), 137–143. https://doi.org/10.1016/j.resuscitation.2006.08.011

Mumaw, R.J., Haworth, L., & Feary, M. (2019). The role of alerting system failures in loss of control accidents. CAST SE-210 Output 2. NASA/TM-2019-220176.

Mumaw, R.J., Sarter, N., Wickens, C., Kimball, S., Nikolic, M., Marsh, R., Xu, W., & Xu, X. (2000). Analysis of pilots' monitoring and performance on highly automated flight decks. (Final Project Report: NASA Ames Contract NAS2-99074). Seattle, WA: Boeing Commercial Aviation.

National Transportation Safety Board (1994). Safety study: A review of flightcrew-involved major accidents of U.S. air carriers, 1978 through 1990. NTSB/SS-94/01, PB94-917001. Washington, DC: NTSB.

Nguyen, N., Elliott, J. O., Watson, W. D., & Dominguez, E. (2015). Simulation improves nontechnical skills performance of residents during the perioperative and intraoperative phases of surgery. *Journal of Surgical Education*, *72*(5), 957–963. https://doi.org/10.1016/j.jsurg.2015.03.005

Nicksa, G. A., Anderson, C., Fidler, R., & Stewart, L. (2015). Innovative approach using interprofessional simulation to educate surgical residents in technical and nontechnical skills in high-risk clinical scenarios. *JAMA Surgery*, *150*(3), 201. https://doi.org/10.1001/jamasurg.2014.2235

Ounounou, E., Aydin, A., Brunckhorst, O., Khan, M. S., Dasgupta, P., & Ahmed, K. (2019). Nontechnical skills in surgery: A systematic review of current training modalities. *Journal of Surgical Education*, *76*(1), 14–24. https://doi.org/10.1016/j.jsurg.2018.05.017

Peißl, S., Wickens, C.D., & Baruah, R. (2018). Eye-tracking measures in aviation: A selective literature review, *The International Journal of Aerospace Psychology,* DOI: 10.1080/24721840.2018.1514978

Petzoldt, T., Weiß, T., Franke, T., Krems, J.F., & Bannert, M. (2013). Can driver education be improved by computer-based training of cognitive skills? *Accident Analysis & Prevention*, *50*, 1185–1192. https://doi.org/10.1016/j.aap.2012.09.016

Piras, A., & Vickers, J. N. (2011). The effect of fixation transitions on quiet eye duration and performance in the soccer penalty kick: Instep versus inside kicks. *Cognitive Processing*, *12*(3), 245–255. https://doi.org/10.1007/s10339-011-0406-z

Potter, B. (2011). *Effects of online training on aircrew monitoring behaviors: A field study*. Embry-Riddle Aeronautical University, Daytona, Beach, FL.

Pradhan, A. K., Pollatsek, A., Knodler, M., & Fisher, D. L. (2009). Can younger drivers be trained to scan for information that will reduce their risk in roadway traffic scenarios that are hard to identify as hazardous? *Ergonomics*, *52*(6), 657–673. https://doi.org/10.1080/00140130802550232

Reynal, M., Colineaux, Y., Vernay, A., & Dehais, F. (2016). Pilot flying vs pilot monitoring during the approach phase: An eye-tracking study. In Proceedings of the International Conference on Human-Computer Interaction in Aerospace. ACM.

Reynal, M., Rister, F., Scannella, S., Wickens, C., & Dehais, F. (2017). Investigating pilot's decision making when facing an unstabilized approach: An eye-tracking study. In Proceedings of the 6th Congress of the International Symposium on Aviation Psychology, 18- 23 June, Dayton, Ohio.

Rienhoff, R., Tirp, J., Strauß, B., Baker, J., & Schorer, J. (2016). The 'Quiet Eye' and motor performance: A systematic review based on Newell's constraints-led model. *Sports Medicine*, *46*(4), 589–603. https://doi.org/10.1007/s40279-015-0442-4

Robertson, J. M., Dias, R. D., Yule, S., & Smink, D. S. (2017). Operating room team training with simulation: A systematic review. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, *27*(5), 475–480. https://doi.org/10.1089/lap.2017.0043

Robinson, D. (2000). The Development of Flight Crew Situation Awareness in Commercial Transport Aircraft. British Airways.

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

Rosqvist, E., Lauritsalo, S., & Paloneva, J. (2018). Short 2-H in situ trauma team simulation training effectively improves non-technical skills of hospital trauma teams. *Scandinavian Journal of Surgery*, 145749691878900. https://doi.org/10.1177/1457496918789006

Rothkopf, C.A., Ballard, D.H., Hayhoe, M.M., & Regan, O. (2007). Task and context determine where you look. *Journal of Vision, 7*, 1–20.

Ryffel, C. P., Muehlethaler, C. M., Huber, S. M., & Elfering, A. (2019). Eye tracking as a debriefing tool in upset prevention and recovery training (UPRT) for general aviation pilots. *Ergonomics*, *62*(2), 319–329. https://doi.org/10.1080/00140139.2018.1501093

Salas, E., Burke, C. S., Bowers, C. A., & Wilson, K. A. (2001). Team training in the skies: Does Crew Resource Management (CRM) training work? *Human Factors*, *43*(4), 641–674.

Schaudt, W.A., Caufield, K.J., & Dyre, B.P. (2002). Effects of a virtual airspeed error indicator on guidance accuracy and eye movement control during simulated flight. Proceedings of the Human Factors andErgonomics Society Annual Meeting.

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors, 50*(6), 864–879. DOI:10.1518/001872008X374974

Schulz, C. M., Endsley, M. R., Kochs, E. F., Gelb, A. W., & Wagner, K. J. (2013). Situation awareness in anesthesia: Concept and research. *Anesthesiology*, *118*(3), 729–742. https://doi.org/10.1097/ALN.0b013e318280a40f

Senders, J.W. (1983). Visual scanning processes. Netherlands: University of Tilburg Press.

Sharma, B., Mishra, A., Aggarwal, R., & Grantcharov, T. P. (2011). Non-technical skills assessment in surgery. *Surgical Oncology*, *20*(3), 169–177. https://doi.org/10.1016/j.suronc.2010.10.001

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

Soderstrom, N.C. & Bjork, R.A. (2015). Learning versus performance: An integrative review. *Psychological Science*, *10*(2), 176-199.

Soll, H., Hoermann, H.-J., Blokziji, C., Bieder, C., Paries, J., Lehmann, O., … Polo, L. (2003). *A validated recurrent training programme for situation awareness and threat management* (p. 49).

Spady, A. A. (1978). Airline pilot scan patterns during simulated ILS approaches. NASA Technical Paper 1250. NASA: Langley Research Center.

Sumwalt, R. L. (1995). Altitude awareness programs can reduce altitude deviations. *Flight Safety Digest*, (December).

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Taatgen, N.A., van Rijn, H., & Anderson, J. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention and learning. *Psychological Review, 114*(3), 577–598.

Thomas, F., D., Rilea, S., L., Blomberg, R., D., Peck, R., C., & Korbelak, K., T. (2016). *Risk awareness and perception training program for novice teen drivers*. National Highway Traffic Safety Administration.

United Kingdom CAA (2013). Monitoring matters: Guidance on the development of pilot monitoring skills. CAA Paper 2013/02. West Sussex, UK: CAA.

van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye movements as an indicator of situation awareness in a flight simulator experiment. *The International Journal of Aviation Psychology,* 22:1, 78-95, DOI: 10.1080/10508414.2012.635129

Vicente, K.J., Roth, E.M., & Mumaw, R.J. (2001). How do operators monitor a complex, dynamic work domain? The impact of control room technology. *International Journal of Human-Computer Systems, 54*, pp 831-856.

Vickers, J. N. (1996). Visual control when aiming at a far target. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 342–354.

Vu, K.-P., L., Kiken, A., Chiappe, D., & Strybel, T., Z. (2013). Application of Part-Whole training methods to evaluate when to introduce NextGen Air Traffic Management tools to students. *American Journal of Psychology*, *126*(4), 433–447.

Vu, K.-P., L., Silva, H., Ziccardi, J., Morgan, C., Morales, G., Grigoleit, T., … Battiste, V. (2012). How does reliance on automated tools during learning influence students' Air Traffic Management skills when the tools fail? Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting.

Walker, G. H., Stanton, N. A., Kazi, T. A., Salmon, P. M., & Jenkins, D. P. (2009). Does advanced driver training improve situational awareness? *Applied Ergonomics*, *40*(4), 678–687. https://doi.org/10.1016/j.apergo.2008.06.002

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

Wetton, M.A., Hill, A., & Horswill, M.S. (2013). Are what happens next exercises and self-generated commentaries useful additions to hazard perception training for novice drivers? *Accident Analysis & Prevention*, *54*, 57–66. https://doi.org/10.1016/j.aap.2013.02.013

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.

Wickens, C.D., Goh, J., Helleberg, J., Horrey, W., & Talleur, D.A. (2003). Attentional models of multi-task pilot performance using advanced display technology. *Human Factors, 45*(3), 360-380.

Wood, G., & Wilson, M. R. (2011). Quiet-eye training for soccer penalty kicks. *Cognitive Processing*, *12*(3), 257–266. https://doi.org/10.1007/s10339-011-0393-0

Wood, T. C., Raison, N., Haldar, S., Brunckhorst, O., McIlhenny, C., Dasgupta, P., & Ahmed, K. (2017). Training tools for nontechnical skills for surgeons—A systematic review. *Journal of Surgical Education*, *74*(4), 548–578. https://doi.org/10.1016/j.jsurg.2016.11.017

Young, A. H., Chapman, P., & Crundall, D. (2014). Producing a commentary slows concurrent hazard perception responses. *Journal of Experimental Psychology: Applied*, *20*(3), 285–294. https://doi.org/10.1037/xap0000016

Yu, L., & Chen, J. (2014). Comparative study on the security personnel training for high-risk sports in America and China. International Conference on Education, Langauge, Art and Intercultural Communication. Atlantis Press.

Yule, S., Flin, R., Maran, N., Rowley, D., Youngson, G., & Paterson-Brown, S. (2008). Surgeons' non-technical skills in the operating room: Reliability testing of the NOTSS behavior rating system. *World Journal of Surgery*, *32*(4), 548–556. https://doi.org/10.1007/s00268-007-9320-z

Yule, S., Parker, S. H., Wilkinson, J., McKinley, A., MacDonald, J., Neill, A., & McAdam, T. (2015). Coaching non-technical skills improves surgical residents' performance in a simulated operating room. *Journal of Surgical Education*, *72*(6), 1124–1130. https://doi.org/10.1016/j.jsurg.2015.06.012

Zausig, Y.A., Grube, C., Boeker-Blum, T., Busch, C.J., Bayer, Y., Sinner, B., … Graf, B. M. (2009). Inefficacy of simulator-based training on anaesthesiologists' non-technical skills. *Acta Anaesthesiologica Scandinavica*, *53*(5), 611–619. https://doi.org/10.1111/j.1399-6576.2009.01946.x

Appendix A. A Short History of Institutional Findings and Training Recommendations on Pilot Monitoring

Concerns about failures in pilot monitoring and awareness are not new. Over the last several decades, a number of aviation regulatory, operations, and safety organizations have identified concerns with pilot monitoring and awareness and made recommendations to mitigate the problems. This Appendix briefly describes the notable efforts that are relevant to the current project and provides a compilation of the training recommendations that these various organizations have made through their reports.

In 1994, the National Transportation Safety Board (NTSB), released a Safety Study on 37 flightcrew-involved accidents between 1978 and 1990 (NTSB, 1994). They identified 302 pilot errors across these events and determined that 70 of them (23%) were monitoring/challenging errors. In fact, this class of error occurred in 31 of the 37 accidents, and 76% of the 70 monitoring/challenging errors were failures to catch an error that the NTSB had labeled as causal to the accident. They also noticed that the First Officer (FO) was the PM (and the Captain was the PF) in 80% of the accidents flights and pointed to difficulties for the FO to intervene when the Captain is making poor decisions or fails to take an action.

In 1996, the FAA published a report called, "The interfaces between flightcrews and modern flight deck systems," which focused primarily on human factors issues concerning the autoflight system. One of the sets of recommendations in this report was motivated by the new complexity in the flightdeck interface and its effects on flight crew situation awareness. Two of the Situation Awareness recommendations are relevant to monitoring:

- Recommendation SA-5: "The FAA should encourage the exploration, development, and testing of new ideas and approaches for providing effective feedback to the flightcrew to support error detection and improved situation awareness."
- Recommendation SA-8: "The FAA should ensure that flightcrews are educated about hazardous states of awareness and the need for countermeasures to maintain vigilance. The FAA should encourage operators to: a) Develop operational procedures and strategies to foster attention management skills with the objective of avoiding hazardous states of awareness; and b) Develop techniques to apply during training to identify and minimize hazardous states of awareness.

In 2010, the CAST, prompted by a set of LOC accidents in the previous 10 years, formed a team to analyze incidents and accidents in which the pilot or flightcrew experienced a loss of awareness for energy state or aircraft attitude that led to a loss of control. These events (listed in Table A.1) were linked to a loss of airplane state of awareness (see the CAST, 2014 report for more details: https://www.skybrary.aero/bookshelf/books/2999.pdf). Failures in pilot monitoring or the flightcrew's ability to maintain attention on FPM were seen as significant contributors to this set of events. CAST generated a set of Safety Enhancements (SEs) to address issues identified by the accident analysis. Some of the SEs that are most relevant include:

- SE 195: Flight Crew Training Verification and Validation
- SE 199: Enhanced Crew Resource Management Training
- SE 208: Airplane Systems Awareness
- SE 211: Training for Attention Management

More information can be found on these specific SEs at the Skybrary site (www.skybrary.aero).

Table A.1. Safety Events from the CAST ASA Report

Low-energy or Low-airspeed Events	*Attitude Awareness Events*
Icelandair; B757; Oct 19, 2002	Formosa Airlines B-12255; Saab 340B; Mar 18, 1998
Midwest Express; B717; May 12, 2005	Korean Air 8509; 747-200F; Dec 22, 1999
Provincial Airlines; DHC-8; May 27, 2005	Gulf Air 072; A320; Aug 23, 2000
West Caribbean 708; MD-82; Aug 16, 2005	Icelandair 315; 757-200; Jan 22, 2002
Thomsonfly; 737-800; Sept 23, 2007	Flash Airlines 604; 737-300; Jan 3, 2004
XL Airways 888T; A320; Nov 27, 2008	Armavia 967; A320; May 3, 2006
Colgan Air 3407; DHC-8-Q400; Feb 12, 2009	Adam Air 574; 737-400; Jan 1, 2007
Turkish Airways 1951; 737-800; Feb 25, 2009	Kenya Airways 507; 737-800; May 5, 2007
Empire Air 8284; ATR-42; Jan 27, 2009	Aeroflot Nord 821; 737-500; Sept 14, 2008

In 2013, the FAA (specifically, the Performance–based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team (PARC/CAST) Flight Deck Automation Working Group (FltDAWG)) did a follow-up report to its 1996 "interfaces" report (mentioned above), called, "Operational use of flight path managements systems." While the focus of this report was not on monitoring, the topic showed up in numerous discussions. For example, these snippets:

- The accident investigation boards identified that pilots were out of the control loop in over 50% of the accidents reviewed by the Working Group.
- Factors that contributed to insufficient awareness included...insufficient methods for monitoring mode changes.
- Operators have increased emphasis on crew communication and cross verification in many airlines. They recognize the value of formalized confirmation and cross-verification of selected modes, such as verbalize, verify, monitor; or confirm, analyze, monitor, and intervene).
- Although there is general industry consensus that monitoring, cross verification, and error management are important, these topics are not always explicitly trained.
- Long-term use of FMS-derived flight path trajectory without the need to critically assess or intervene may atrophy the skills needed to anticipate, monitor and react.
- Recommendation 8 from this report specifically mentions the need for improved methods for monitoring to improve airplane state and flight path awareness.

Since 2013, there have been three aviation-related reports specifically addressing challenges to monitoring and proposals for improving training for monitoring. In 2013, the Civil Aviation Authority of the United Kingdom (UK CAA) delivered a report titled, "Monitoring Matters: Guidance on the development of pilot monitoring skills." The aim of this report was to:

- promote a good understanding among pilots about why active monitoring is important; to highlight human frailties; and to highlight strategies that can improve monitoring

- place emphasis on training of monitoring, through procedures, assessment scenarios, and behavioral markers
- look to airlines to put in place monitoring procedures, focus on monitoring in FOQA/FDM, and promote a monitoring culture

This report:

- provides a number of good examples of accidents and incidents in which monitoring failures played a role
- provides a set of "tips" to make monitoring more effective
- identifies a set of behavioral markers (behaviors that can be observed) for monitoring, which can be used in evaluating a pilot's use of monitoring in a training setting
- discusses methods for measuring monitoring skills
- discusses the potential role for standard operating procedures (SOPs) and briefings in monitoring

In 2014, Flight Safety Foundation published a report from the Active Pilot Monitoring Working Group, called, "A practical guide for improving flight path monitoring." This report identified the barriers to effective monitoring and then made a set of recommendations regarding pilot monitoring practices, procedures and policies, and, finally, training and evaluating monitoring skills.

In 2016, the International Civil Aviation Organization (ICAO), through their International Air Transport Association (IATA) Pilot Training Task Force, compiled their 1st edition of "Guidance material for improving flightcrew monitoring." This report starts with a discussion of monitoring and its limitations and also covers topics from the other two reports—e.g., tips on how to make monitoring more effective, defining operational policies to improve monitoring, potential behavioral markers, and training recommendations. A unique element of this report is an Appendix that pulls together example training materials from a number of operators around the world.

Finally, smaller, more focused recommendations have come recently from the FAA's Flight Path Management Working Group, which is an element of the Air Carrier Training Aviation Rulemaking Committee.

Below is a compilation of guidance or recommendations on training pilot monitoring from some of the reports just described plus a few other sources. These are all captured at a high level and there is sometimes more detail in each report. As an even more concise summary, we have identified the following training topics from those various lists, and ordered them roughly by how often they appear in the various reports:

1. The responsibility to challenge the PF when deviations occur, plus effective intervention techniques, including judgment and decision-making about when to intervene and how to take over control when needed.
2. Monitoring of automated systems and their behaviors, including mode progressions through a normal flight.
3. Pilot vulnerabilities to monitoring errors and lapses.
4. Areas of vulnerability, which are points in the flight when workload is expected to be high.

5. Workload management, which is about shifting workload to a more appropriate time in the flight when monitoring demands are lower.
6. Managing distractions that can pop up and interfere with monitoring.
7. How monitoring relates to threat and error management (TEM)and CRM.
8. Operator policies and procedures related to monitoring the flight path.
9. Monitoring as a task vs monitoring as a role.
10. Methods to resume effective monitoring if it becomes degraded.
11. Monitoring for altitude changes.
12. Access, anticipate and analyze FMS data.
13. Access and analyze information tools, such as Aircraft Communications Addressing and Reporting System (ACARS).

The following are the individual reports and their training recommendations.

From an NTSB Accident Report, Recommendation A-07-13 to FAA:
Require pilot training programs be modified to contain modules that teach and emphasize monitoring skills and workload management and include opportunities to practice and demonstrate proficiency in these areas.

From "A Practical Guide for Improving Flight Path Monitoring: Final report of the Active Pilot Monitoring Working Group" Flight Safety Foundation, 2014. This report makes a number of recommendations for training, including:

- Train pilots about why they are vulnerable to errors and monitoring lapses.
- Reinforce the responsibility of monitoring pilots to challenge deviations.
- Develop and publish clearly defined monitoring tasks, training objectives and proficiency standards. Ensure instructors and evaluators are proficient at training and evaluating these standards.
- Implement a comprehensive approach to training and evaluating autoflight and flight path monitoring.
- Incorporate monitoring training into simulator or other device training.
- Place greater emphasis on monitoring in operator flight standards programs.

The report also makes a number of other recommendations related to training content or policy, including the following:

- Practice interventions to maintain effective monitoring or to resume effective monitoring if degraded.
- Practice interventions to resume effective monitoring after distractions and interruptions.
- Promote policies, procedures and practices to improve monitoring of altitude changes.
- Develop and refine training to improve the monitoring of automated systems as incorporated in the flight path management policy.

From "Guidance material for improving flightcrew monitoring," ICAO, 2016. Training monitoring knowledge, skills and attitudes for pilots should include:

1. Defining what monitoring is and training monitoring as a task:
    - Train the "monitoring process."
    - Train how monitoring is integrated in all pilot skills and competencies.
    - Discuss how monitoring relates to TEM.
    - Discuss how monitoring relates to CRM.
    - Discuss the differences between functional and assigned roles.
    - Discuss the differences between monitoring as a "task" and monitoring as a "role."
2. Ensuring pilots are trained to possess the knowledge, skills, and attitudes to function as active monitors. Pilots need to be trained for what they should be looking for, verifying that something is working properly or that there is a problem. For example, pilots should have the ability to proficiently:
    - Predict flight mode progression
    - Plan and verbalize intentions regarding the use of automation to another crew member
    - Access, anticipate, and analyze all FMS data
    - Access and analyze information tools (e.g., ACARS, EFB, CPDLC, etc.)
3. Defining what monitoring is and training specific monitoring tasks for PF and PM positions:
    - Develop monitoring qualification standards for training programs
    - Integrate FPM training into type rating and recurrent training
    - Incorporate Areas of Vulnerability (AoV) into flight training, defined by phase of flight
    - Define appropriate monitoring "sample rate" for each AoV and reinforce with scenario-based training
4. Ensuring desired monitoring skills are blended into standard operating procedures (SOPs) (e.g., verbalizing FMA changes) and identifying any SOPs that possibly hinder effective flight path monitoring (e.g., eliminating unnecessary talking during critical phases of flight).
5. Developing and defining PM tasks and creating specific training/learning objectives and proficiency standards to be incorporated into the training syllabus. This will be necessary for instructors to train and evaluate monitoring performance.
6. Operators to consider intervention training and the appropriate skill development for taking over and swapping PF and PM roles.
7. Ensuring flight path monitoring skills are emphasized and incorporated from the initial stages of training through type-rating, and then continually emphasized during recurrent training and line operations.
    - Identify different means for training flight path monitoring at different phases of training (e.g., computer-based training, desktop training, fixed-based simulator, full flight simulator). For example, learning about AoV during ground training and practicing/applying during scenario-based training in the flight training phase.
    - Create training scenarios that introduce system failures to condition a level of skepticism and encourage maintaining a higher level of awareness.
    - Validate/evaluate monitoring KSAs (knowledge, skills, and abilities) and observable behaviors during all phases of training.

8. Improving FSTD (Simulator) Instructor and Line Check Pilot (Flight Standards) training to enhance their ability to train/evaluate flight path monitoring skills.
    - Increase the level of training instructors/evaluators receive in training pilot monitoring skills. This training should be incorporated into instructor/evaluator annual training.
    - Importance of training/evaluating EFPM must be emphasized to instructors.
    - Develop monitoring qualification standards for AQP, ATQP training programs to give instructors definitive guidelines for expected pilot performance. Failure to make one call out should not rise to the level of severity that actually violating an assigned altitude may.
9. Emphasizing monitoring tasks and procedures throughout a pilot's entire training and line operations. If a flight crewmember's monitoring tasks and procedures are not continuously emphasized, the natural assumption of a flight crew member would be that they are not considered as important as other tasks/skills/SOPs. Instructors/evaluators must emphasize to pilots that monitoring skills/compliance is just as important as any other task they are trained in.
10. Including flight path monitoring elements in all briefings and debriefings (e.g., considering energy management, expected configuration, specific stabilized approach criteria to monitor and expected action of the PM, and what the expected FMA changes will be when briefing a non-ILS approach).
11. Developing special purpose operational training (SPOT) or event sets to be administered during recurrent training that emphasize effective monitoring as a task, flight path monitoring, PM monitoring tasks and SOPs, etc.
    - Identify specific monitoring skills and observable behaviors to focus on, instead of training to a broader scope
    - Specific training scenarios should be designed to emphasize certain monitoring knowledge and skills
12. Developing interactive distance learning modules dedicated to teaching enhanced monitoring skills using scenario-based training and making them available for flight crew members to review.
13. Developing Verbal Communication Skill Training as part of augmenting flight crew member CRM skills.
14. Lack of manual flying in training and line operations also means lack of exposure to monitoring of hand flown approaches. Additional training to improve manual flying skills also creates an opportunity to train predictive monitoring of the PM.
15. Operators to consider providing "Takeover training" to help prevent the situation where both pilots end up acting as PF while monitoring is overlooked. This can easily happen when one pilot takes over control of the aircraft from the other during critical moments and the PF fails to switch to the PM role.

From ACT ARC Recommendation 15-10: Guidance Material Addressing Intervention Strategies. There are recommendations for training flight path intervention; human-to-human and human-to-machine; including:

- PM communication (content) for intervening with the PF, including communication style and assertiveness

- judgment and decision-making around intervention
- basic hand-flying skills (when you intervene from autoflight)
- autoflight system behavior
- autoflight system management
- cognitive skills for manual flight operations

From ACT ARC Recommendation 16-4: Training Elements for Training the Pilot Monitoring. This document provides a list of topics that should be covered in training for the PM position, including training on:

- The operator's policies and procedures related to monitoring the flight path (e.g., callouts, double-pointing, etc.). This training should also include any of the carrier's recommended practices.
- Applicable common errors in monitoring the flight path.
- The concept that there are predictable situations during each flight when the risk of a flight path deviation is increased, heightening the importance of proper task/workload management.
- Managing distractions that interfere with monitoring the flight path.
- The responsibilities of the PM to monitor the flight path.
- Recognizing when the PF is not adequately controlling the flight path or when the PM is not adequately monitoring the flight path.
- Intervention methods that PM can use to help the PF regain proper control of the flight path.
- Operationally relevant combinations/levels of flight guidance and flight control automation.
- Anticipating, recognizing, and recovering from known flight guidance (includes FMS) and flight control (includes autopilot and, autothrottles) system-behavioral challenges (e.g., subtle mode reversions), and environmental/ circumstantial traps that are known to lead to flight path-related errors (e.g., vectors off, then back on, a Standard Terminal Arrival Route (STAR) during a "descend via" clearance).

From "Standard operating procedures and pilot monitoring duties for flight deck crewmembers" FAA Advisory Circular 120-71B, 2017.
An operator should train its pilots on all policies and procedures related to monitoring the flightpath (e.g., callouts, double-pointing, etc.). This training should also include any of the operator's recommended practices:

1. Pilots should be trained on the responsibilities of the PM to monitor the flightpath. In particular, pilots should be trained to recognize when the PF is not adequately controlling the flightpath or when the PM is not adequately monitoring the flightpath. This training should include pilot task loading and signs of diminished performance. Some examples include lack of communication, channelized attention, and failure to make required callouts.
2. Pilots should be trained on applicable common errors in monitoring the flightpath. This includes training on appropriate methods of recognizing precursors to, and signs of, degraded monitoring and on resolving monitoring errors and/or lapses.
3. Pilots should be trained on the concept that there are predictable situations during each flight when the risk of a flightpath deviation is increased, heightening the importance of

proper task/workload management. If the PM is trained to recognize the flight phases or situations when they are most vulnerable to flightpath deviations (including when little time exists to correct deviations), he or she could strategically plan tasks and workload to maximize monitoring during those phases.

4. Pilots should be trained on CRM/TEM principles and human performance vulnerabilities related to monitoring, the importance of monitoring, and the operator approved practices that achieve effective monitoring of the flightpath.
5. Pilots should be trained on system failures that may distract from effective monitoring and proper flightpath management.
6. Pilots should be trained to manage distractions that interfere with monitoring the flightpath. Provide guidance on managing task priorities and train them to effectively switch between other tasks and monitoring of the flightpath so that flightpath vigilance is always maintained. Include information and task management strategies that enable pilots to use charts, EFB, ACARS, etc. while also effectively monitoring the flightpath and airplane energy state.
7. Pilots should be trained on intervention methods that the PM can use to help the PF regain proper control of the flightpath and provide opportunities for the PM to practice these methods (e.g., calling out deviations, levels of assertiveness).
8. Pilots should be trained on and be able to demonstrate understanding of operationally relevant combinations/levels of flight guidance and flight control automation (e.g., given a certain set of circumstances, what will happen next?).
9. Ensure pilots can transition seamlessly between combinations/levels of flight guidance/flight control automation (including manual flight) by training them to anticipate, recognize, and recover from known flight guidance (includes FMS) and flight control (includes autopilot [AP] and autothrottles) system-behavioral challenges (e.g., subtle mode reversions), and environmental/ circumstantial traps that are known to lead to flightpath-related errors (e.g., vectors off, then back on, a Standard Terminal Arrival Route (STAR) during a "descend via" clearance).
10. Flight guidance and flight control systems training should include an assessment of a pilot's understanding of those systems and what will happen 'next' given a certain set of flight circumstances, and the reasons why. The training should incorporate FMS degradations and failures and operational consequences requiring flightcrew action, known flight guidance and flight control system-behavioral challenges (e.g., subtle mode reversions), and environmental/circumstantial traps (e.g., vectors off, then back on, a STAR during a "descend via" clearance) that are known to lead to flightpath-related errors.

Appendix B. Review of Eye-Tracking Studies of Commercial Transport Pilots

Given that we have not identified an existing prescription—related to eye fixations or scanning—for pilot monitoring (as described in Section 5), we wanted to know what lessons we could draw from ET studies about how pilots monitor. Generally, ET studies have used a wide range of pilots, but our focus in this review was on ET studies as they relate to commercial transport pilots. Note that two reviews of the broader set of pilot-related ET studies were recently published (Ziv, 2017; Peißl et al., 2018). In addition to covering some of the studies described in this Appendix, those reviews also cover studies of visual attention with low-time pilots using desk-top mock-ups of aviation displays.

B.1. Assumptions in Eye-tracking Studies

A key assumption in measuring ET is that the eye-fixation point is where visual attention is directed. This assumption can be incorrect for, at least, three reasons. First, visual attention can be on stimuli in the periphery of vision. Foveal vision captures much less than 10º of the visual field. Peripheral vision typically extends beyond 90º to the left and right of the focus point. Thus, although the pilot's eyes may be fixated on the airspeed indication, his/her attention could be on a sudden change on the engine display. Indeed, salient stimuli in the periphery can attract visual attention, leading to a shift in fixation to that area (e.g., Schaudt, Caufield, & Dyre, 2002).

Second, our attention can be disconnected from the visual field; that is, we may not be processing what is currently at or near the center of the visual field. One known phenomenon, referred to as "inattentional blindness" (Mack & Rock, 1998), reveals that a viewer can fail to have awareness of objects or events near the fixation point, demonstrating that visual attention can be narrow and selective. Also, attention can be directed inward and fail to notice what is at the fixation point.

Third, it is important to note the inherent difficulties with data collection and analysis of physiological measurement, including eye-tracking, in realistic environments, such as flight training or operations. There are several challenges that make it difficult to make confident conclusions about monitoring behavior solely on the basis of eye-tracking data. (Kalar et al., 2016). These challenges include:

- *Measurement techniques*. There are different methods for determining eye-fixations. Some of these methods are more robust to high-noise environments (e.g., flight simulators used for pilot training), and some provide more precise and accurate data in more-controlled environments.
- *Individual differences*. Eye color, eyelid closure, and other anatomical differences can result in differing quality of eye-tracking data across individuals.
- *Analysis techniques*. Different analysis algorithms and the varying treatment of data can result in different conclusions from the same dataset. For example, some algorithms may be very aggressive about what is considered an eye-fixation, whereas others may be more conservative.
- *Physical layout*. The physical layout of instrumentation and the placement of eye-tracking sensors may result in reduced precision and accuracy for some areas of the flight deck interface.
- *Behavioral differences* .The nature of the eye-tracking equipment and how it is placed on the participant (e.g., pilot) may alter the behavior of the participant being measured.

- *Noise levels*. Environmental noise sources include vibration and glare.
- *Calibration and setup of equipment*. Eye-tracking precision and accuracy are significantly impacted by reliable setup and calibration of the eye-tracking sensors.

However, typically, in an operational setting, pilots do attend to where their eyes are fixating, and fixations can inform us about visual attention / monitoring. There are several measures of monitoring that can be captured with ET, including:

- The fixation points on the interface, which is where the eye is directed.
- Dwell time, which is the length of a fixation.
- % (total) dwell time, which is the percentage of total time that fixations are in a specific region.
- Scan pattern, which captures the sequencing of fixations over a period of time.

The following is a summary of what the research studies tell us about how pilots are allocating visual attention.

B.2. Review of Eye-tracking Findings

As we said above, because our interest is in commercial transport pilots—pilots who use a complex interface, such as the one in Figure 9—this review focuses on ET studies of these pilots in a simulator that is a realistic representation of the airplane they fly.

B.2.1 Areas of Interest

ET studies establish context by defining AoIs, which are areas in the environment that have meaning for the tasks being performed (and can be reliably distinguished from each other). With ET studies of pilots, AoIs are elements of the flight deck interface and the out-the-window view. An AoI can designate a large element of the interface, such as a Navigation Display, or it can designate a smaller element of a display, such as the airspeed indication or the Flight Mode Annunciation (FMA) area on the PFD. The following are commonly used AoIs from these studies:

- Primary flight display, which is sometimes further partitioned into
  - Airspeed (AS)
  - Attitude (ADI)
  - Altitude, which is sometimes combined with vertical speed (ALT)
  - Heading (HDG)
  - Flight mode annunciation (FMA), and sometimes the individual modes
- Navigation display (ND)
- Out the window (OTW)
- Engine display / Central alerting (Eng/CAS)
- Mode control panel / Flight control unit (MCP)
- Control display unit (CDU), which is the interface to flight planning and performance
- Head up display (HUD)
- Other, which is a catch-all for fixations that were not in any of the other AoIs

As mentioned above, the ability to reliably determine which AoI is being fixated is based on the precision of the ET equipment, size and location of the AoI, and the quality of the calibration for each pilot. Not all experiments captured the smaller elements of the PFD.

B.2.2. AoI Dwell Time Findings

One meaningful measure of where pilots are allocating their visual attention is a summary of how much time was spent in each AoI. This measure is typically presented as percent dwell time (%DT) 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. We identified a number of studies that put current, qualified airline pilots in a high-fidelity simulator, had them perform a maneuver, and measured fixations relative to a set of AoIs.

The Tables in this section present data on %DT. We start with those cases where we could find multiple studies of the same maneuver. We separate data for different maneuvers because different maneuvers can have different monitoring requirements. The Tables also separate fixations of the PF from the PM. These two roles also may have different monitoring requirements.

B.2.2.1. Take-off (%DT)

We found two studies of the Take-off maneuver. For both, there were %DT measures for the elements of the PFD as well as measures for the larger displays. Tables B.1 and B.2 are from the same studies but focus on different AoI groups. Also, in one of the studies, there were measures for both the PF and the PM. One striking finding in these tables is the difference between the A320 PF and PM regarding which AoIs were primary. The PF is expected to be looking out the window as the airplane accelerates down the runway; the PM is expected to look at the Engine display to monitor that the engines are working properly, and airspeed to determine when the airplane is going too fast to perform a rejected take-off (RTO). The first table here shows that the PM spent more time on these displays and less time looking out the window.

Table B.1. Percent Dwell Time for the Take-off Maneuver (Large AoIs)

	PFD	*ND*	*Eng/CAS*	*OTW*	*MCP*	*CDU*
B747-400 PF[1]	--*	2	3	70		
A320 PF[2]	--*	1	9	64		
A320 PM[2]	--*	2	27	28		

Table B.2. Percent Dwell Time for the Take-off Maneuver (PFD AoIs)

	ADI	*AS*	*ALT*	*HDG*	*FMA*
B747-400 PF[1]	4	4	0	0	2
A320 PF[2]	9	12	2	2	4
A320 PM[2]	8	19	3	4	3

** %DT for the PFD is partitioned into sub-elements, presented in the 2nd Table*

A blank cell means that this study did not use this AoI.

[1] Mumaw et al., 2000 (data from 17 PFs).

[2] Lefrancois et al., 2018 (data from 10 PFs and 10 PMs).

B.2.2.2. Approach: Manual (%DT)

We found three studies of a manual approach maneuver (see Table B.3), one of which measured both the PF and PM. In this case, the focus of the AoIs was the elements of the PFD. The ADI, generally, gets the most attention for this maneuver. Also, note that there is quite a bit of variability, especially for AS, ALT, and HDG.

Table B.3. Percent Dwell Time for the Manual Approach Maneuver (PFD AoIs)

	ADI	AS	ALT	HDG	FMA
B777/A330 PF[3]	35	4	7		
A320 PF[4]	28	5	8	6	1
A320 PF[5]	26	12	20	25	
A340 PF[5]	32	16	18	21	
B777/A330 PM[3]	21	8	13		

A blank cell means that this study did not use this AoI.
[3] Dehais et al., 2017 (data from 12 PFs & 12 PMs).
[4] Lefrancois et al., 2016 (data from 4 "best performing" PFs).
[5] Haslbeck & Zhang, 2017 (data from 51 PFs).

B.2.2.3. Approach: Autoflight (%DT)

We found six studies of an approach maneuver with autoflight engaged (see Table B.4), two of them measured both the PF and PM. Again, ADI gets the most attention. Altitude and heading get fewer fixations when compared to the manually flown maneuver. The FMA, understandably, gets more attention since that is an important indicator of the autoflight system's behavior.

Table B.4. Percent Dwell Time for the Autoflight Approach Maneuver (PFD AoIs)

	ADI	AS	ALT	HDG	FMA
B747-400 PF[1]	14	9	7	1	2
A320 PF[4]	15	2	3	1	2
B777/A330 PF[6]	31	8	10	4	5
737NG PF[7]	30	8	7		
A330 PF[8]	11	8	12	2	2
A330 PF[9]	11	8	8	2	3
A320 PM[4]	26	18	11	2	2
B777/A330 PM[6]	16	7	10	4	4

A blank cell means that this study did not use this AoI.
[1] Mumaw et al., 2000 (data from 17 PFs).
[4] Lefrancois et al., 2018 (data from 10 PFs and 10 PMs).
[6] Reynal et al., 2016 (data from 8 PFs and 8 PMs).
[7] Reynal et al., 2017 (data from 10 PFs).
[8] Huettig at al., 1999 (data from 1 PF).
[9] Anders, 2001 (data from 8 PFs and 8 PMs).

B.2.2.4. Go Around (%DT)

There were two studies of a go-around maneuver, both measured the PF and PM (see Tables B.5 and B.6). Again, there were measurements for two sets of AoIs; the elements of the PFD are in Table B.6. The PM looks more at the MCP, which makes sense since it is the responsibility of the PM to make inputs there. It is surprising that the A320 PM has the highest %DT out the window since the expectation is for the PF to focus there.

Table B.5. Percent Dwell Time for the Go-around Approach Maneuver (Large AoIs)

	PFD	ND	Eng/CAS	OTW	MCP	CDU
B777/A330 PF[3]	--*	*4*		*3*	*4*	
A320 PF[4]	--*	18	2	5		
B777/A330 PM[3]	--*	12		1	13	
A320 PM[4]	--*	13	7	10		

Table B.6. Percent Dwell Time for the Go-around Maneuver (PFD AoIs)

	ADI	*AS*	*ALT*	*HDG*	*FMA*
B777/A330 PF[3]	50	13	11		
A320 PF[4]	31	13	2	6	16
B777/A330 PM[3]	16	15	7		
A320 PM[4]	12	14	7	3	16

** %DT for the PFD is partitioned into sub-elements, presented in the 2nd Table*

A blank cell means that this study did not use this AoI.

[3] Dehais et al., 2017 (data from 12 PFs and 12 PMs).

[4] Lefrancois et al., 2018 (data from 10 PFs and 10 PMs).

B.2.2.5. Findings Not Tied to Specific Maneuvers

There were also three studies that captured %DT for the larger AoIs—PFD, ND, and OTW—across a phase of flight (see Table B.7). These give a general sense of %DT for these two primary displays in realistic simulated flight for commercial transport pilots. Not surprisingly, %DT for the out-the-window AoI can vary dramatically, depending on proximity to the ground. (Note that the Dill & Young [2015] study covered a wide range of approach conditions.)

Table B.7. Percent Dwell Time by Phase of Flight (Large AoIs)

	Flight Phase	*PFD*	*ND*	*OTW*
747-400 PF[1]	Take-off roll	14	2	70
747-400 PF[1]	Climb	38	26	4
747-400 PF[1]	Cruise	22	22	1
747-400 PF[1]	VNAV descent	32	33	1
747-400 PF[1]	Approach/landing	40	23	12
A320 PF[2]	Cruise	37	43	
787 PF[10]	Approach	20	16	54
787 PM[10]	Approach	28	25	34

A blank cell means that this study did not use this AoI.
[1] Mumaw et al., 2000 (data from 17 PFs).
[2] van de Merwe et al., 2012 (data from 6 PFs).
[10] Dill & Young, 2015 (data from 10 PFs and 10 PMs).

B.3. Event-Triggered Fixations

The measure %DT is a rough measure; it has the advantage of capturing central tendencies for a group of pilots over a period of time. However, it loses information. Specifically, %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 an operational event. We found two studies that reported fixations tied to operational events; more specifically, tied to autoflight mode behavior.

Mumaw et al. (2000) evaluated pilot monitoring, awareness, and understanding of the autoflight system using 20 airline pilots flying in a 747-400 simulator from San Francisco to Los Angeles. These pilots were given a number of potentially confusing autoflight situations. In one situation, the pilots were transitioned to the VNAV ALT mode in cruise. In this mode, the airplane will not start descending automatically at the top of descent point; the mode needs to be VNAV PTH to start down as expected. Only four of the 20 pilots made sure they were in VNAV PTH prior to the top of descent. Of the other 16 pilots, the fixation data showed that nine pilots fixated the VNAV ALT mode during cruise but took no action to change it. For the other seven pilots, either they did not fixate VNAV ALT or fixation data were lost during that period.

This study also used direct manipulations of the autoflight modes. On three occasions (for each pilot), the mode was artificially changed (and displayed for several minutes) to a mode that was inappropriate for the phase of flight. On the first change, the fixation data showed that 11 of the pilots fixated the mode. For each of the second and third changes, 10 of the pilots fixated the mode. In these 31 opportunities, there was only one case when a mode was called out as erroneous. Thus, either fixation was not a good indicator of pilot attention and the value was not noticed, or failure to call out the mode may have more to do with a poor understanding of mode behavior.

For autoflight mode monitoring, pilots are often trained to notice and call out each mode change. The objective is to ensure that the flightcrew is aware of the current autoflight modes. Mumaw et al. tested whether an autoflight mode was fixated after it changed by looking for a fixation within 10 seconds of the change (when a green box is placed around it) and within 20 seconds of the change. Over all pilots, failure-to-fixate rates within 20 seconds were:

- 34% for modes that were changed manually by the pilot
- 31% for modes that changed without pilot input but were expected (due to airplane behavior)
- 40% for modes that changed without pilot input and were less expected

Bjorklund et al. (2006) also focused on autoflight mode callouts. They were checking to see if flightcrews complied with policy by noticing a mode change, visually verifying the mode change, and then calling out that change to ensure both pilots were aware of it. Six flightcrews flew, in a 737NG simulator, from Amsterdam to London; they flew it twice with each pilot serving as PF, generating 12 flights. Bjorklund et al. determined that there were 521 mode changes during the 12 flights (however, there were fixation data for only 418 of the 521 changes).

Across the 12 flights, pilots made only 146 mode callouts. Of these, only 32 were called out after a visual verification, as determined by the fixation data; the other 114 callouts were made prior to a visual verification. That is, the pilot calling out the mode change did not actually fixate the FMA prior to calling it out. Also, for 44 other mode changes, there was some communication (not a formal callout), but only seven of these 44 communications occurred after a visual verification. Focusing just on whether a mode change was fixated, Bjorklund at el. found that (similar to the Mumaw et al., 2000 data) 40% of the 418 mode changes were not fixated within 20 seconds of the change. Thus, these two studies indicate that pilots are not reliable in visually confirming mode changes in a realistic operational setting. It is likely that these pilots will not be aware of many mode changes.

More generally, these two studies demonstrate that for measuring attention, awareness, and understanding, it can be beneficial to capture event-triggered fixations on specific interface elements. Specifically, in addition to seeing where pilots are allocating their visual attention generally, we can also see if those interface elements were fixated in specific time periods, and whether fixations match other performance measures that reflect awareness and understanding.

B.4. Fixation Dwell Time

Another eye-tracking measure that has received attention is the duration of each fixation. Wickens and Dehais (2019), in a review of pilot expertise, identified a number of studies that 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 advantage of this efficiency is that fewer attentional resources are spent on information extraction and, therefore, are available for higher-level cognitive functions.

However, in examining the six studies cited by Wickens & Dehais, we found that often the novice pilots were pilots with very low time. Specifically, in four of the six studies cited, the novices were general aviation pilots who were working on their initial pilot's license or getting introduced to instrument flight rules (IFR). Typically, these pilots have 100 hours or less of flight time. One of the other two studies used novices with several hundred hours of flight time, and the sixth study did not report flight time details.

While these studies suggest that a shorter dwell time may be a marker of developing expertise, the pilots used in these studies were far less experienced than are commercial transport pilots (airline pilots), who are currently required to have 1500 hours of flight time prior to being hired in the United States. Extraction efficiency may already have developed for airline pilots, with little more to be gained.

Alternatively, it could be that each transition to a new airplane and a new interface requires a period of familiarization in which information extraction becomes more rapid. We believe that more data on fixation dwell time are needed relative to higher-time pilots to assess whether increased extraction efficiency would benefit airline pilots and would be a useful training goal.

B.5. Scanning (Fixation Sequences)

A topic of interest in a few ET studies is fixation sequence. The belief has been expressed that expertise is tied to a specific fixation sequence or scanning pattern. In attempting to answer this question, Haslbeck and Zhang (2017) applied a transition-matrix analysis to capture fixation sequences within the core flight instruments. They were trying to determine if A320 (short-haul) pilots and A340 (long-haul) pilots were applying a consistent scan pattern. They focused on pilots flying a manual approach from 3300 ft down to 270 ft, which was about five minutes of flying time. Almost 60% of the 48 airline pilots were categorized as applying a radial, or attitude-oriented, scan pattern. About 30% of the pilots were categorized as applying a pattern that was less attitude-oriented and more reflective of a circular or triangular pattern that increased emphasis on airspeed, altitude, and heading.

Their categorizations showed that the A340 pilots were more likely to use a radial pattern, whereas the A320 pilots were divided between the two patterns equally. Haslbeck and Zhang then compared flight performance for just the group of A320 pilots and found that pilots characterized by the triangular pattern had significantly smaller localizer deviations (they also had smaller glideslope deviations, but this difference was not significant). Note that this finding may have been influenced by a crosswind that needed to be managed on the approach. Perhaps the greater emphasis on heading helped manage the crosswind-induced deviations.

In a much earlier study of scanning, Spady (1978) also focused on transitions between core flight instruments for seven 737 pilots, who also were flying an approach. This was a much older flight deck interface that was still quite similar to a 6-pack presentation. Spady's data showed that the majority of scanning was of the radial type, with pilots returning to the ADI after looking at nearby instruments. Spady also found some interesting differences between a manually flown approach and a coupled approach. Specifically, for the manual approach, pilots spent more time on the ADI than they did on the coupled approach; and, for the coupled approach, there were more fixations on the other instruments.

One important caveat about these fixation sequence studies is that they focus on a small section of the modern jet transport interface: the PFD indications. While these instruments are essential for flight path management, especially during manual flight, over the course of the complete mission, many other displays/indications will require attention. Thus, the idea of an "ideal scan pattern" fails to address the full interface and the full duties of the flightcrew.

Other researchers have focused on variability in monitoring as reflected 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. In two studies, Ephrath et al. (1980) and Dinocera et al. (2007), there was a demonstrated relationship between increased workload and an increase in scanning entropy (variability). However, these studies focused on low-time pilots; we did not find any similar study with experienced commercial transport pilots.

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. In their study, they gave pilots an automation "surprise," which was an unexpected automation behavior that prevented meeting a flight path objective. This surprise, when it was noticed, shifted pilots significantly toward exploring behavior.

B.6. 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 (Moray (1983) used the term "mean first passage time"). 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, which has been at the heart of a number of aviation accidents.

We found only a single study with airline pilots that exploited this measure. Dehais et al. (2017) captured fixations for 12 B777 and A330 flightcrews during an approach and unexpected go-around. As one of their measures of performance, they captured AoI neglect latency for airspeed, attitude, altitude, and the NAV display. While they found pilots that had neglect latencies of more than 30 seconds for some indications, they did not see any overall negative effects on performance as a result.

B.7. AoI Relevance

A few studies have focused on the range of 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. For example, when performing an altitude-change task, the pilots were also looking at airspeed, which may need to be managed during a climb.

Similarly, Schriver et al. (2008) studied pilots performing a decision-making task and found that more experienced pilots allocated attention to the more diagnostic cues during failure trials than did the less experienced pilots. Unfortunately, each of these two studies compared skilled performance to that of very-low-time pilots in a general aviation setting. More data are required to determine whether this same finding would occur for commercial transport pilots.

In a more-relevant study, Khoo and Mosier (2005) studied regional airline pilots but used an interactive web-based display (not a realistic simulator). They also determined that the more experienced pilots used a wider range of cues in managing a non-normal condition.

While these studies are limited in relevance to commercial transport pilots (our target group of pilots), they offer a potentially important measure that could be applied to that group.

B.8. Pilot Flying vs Pilot Monitoring

The flightcrew on a modern commercial jet transport consists of two pilots, the PF and the PM. In the United States, the Captain and First Officer take turns filling these roles. While airlines clearly identify different duties for these two roles, they typically do not specify in detail how monitoring should be different for PF and PM. A few studies have compared monitoring between the two roles.

Jarvis (2017) had 17 737 flight crews fly approaches in instrument meteorological conditions (IMC) and measured fixations for each pilot. Half of the approaches were flown with autopilot engaged, and half were flown manually. Jarvis measured fixations to a standard set of AoIs, including the ADI and the ND (called here the Horizontal Situation Indicator [HSI]). He found that, when flying manually, the PF spent more than 50% of dwell time on the ADI. But, when the autopilot was engaged, the PF's dwell time on the ADI decreased significantly, and attention to the HSI more than doubled. Interestingly, the PM shifted in the same way when moving from manual flight to autopilot: less ADI and more HSI. Jarvis points out that the flightcrew did not seem to work in a coordinated fashion. That is, the PM could have taken more of the HSI monitoring burden during manual flight, since manual flight seems to require the PF to focus on the ADI.

Reynal et al. (2016) had airline pilots fly a set of approaches with the autopilot engaged. Each of the crewmembers flew two approaches (as PF) and served as PM for two approaches. The %DT data, which were quite similar to the Jarvis findings, showed that there was a significant difference tied to which role the pilot served. PFs spent significantly more time on the ADI than did PMs. And PMs did not seem to complement the PF with greater coverage of other AoIs. The PF and PM were quite similar across the other AoIs.

B.9. Summary and Conclusions

Our focus in this review of the ET research literature was on studies of qualified, current airline pilots performing flight tasks in a realistic simulator setting. The goal was to determine what is known about how these pilots monitor the flight deck interface and to gain insights about skilled performance. The following points offer a summary:

- A number of ET-related measures have been developed to characterize pilot performance and identify markers of skilled performance. Specifically, we identified the following:
  - % Dwell time
  - Event-triggered fixations
  - Fixation dwell time
  - Fixation sequences, or scanning, including measures of entropy and the explore/exploit ratio
- For common flight maneuvers, we have a general understanding of how visual attention is being allocated. That is, the %DT measures reveal generally where the average pilot is allocating attention. Specifically, the PFD and ND, which are the primary displays on the interface, receive much more attention than other display areas. And on the PFD, the ADI, which is at the center of the PFD, receives more

attention than the other PFD elements. Thus, the interface design offers a good configuration in terms of where visual attention is being allocated.

- Specific tasks or other demands can produce large shifts in visual attention; e.g., for flight near the ground, there is a major increase in fixations at the world out the window.
- Fixations on the core flight instruments have received more research attention than fixations across the full flight deck interface. Indeed, a number of studies measured fixations only for PFD-related AoIs. These studies have supported understanding of how visual attention is allocated across those instruments, and they have better characterized the fixation sequences (scanning patterns) that pilots use. However, these studies do not speak directly to the full range of monitoring failures that can impact flight path management.
- No study has established a broader “scan pattern” (outside of the PFD) that characterizes skilled performance. Indeed, there is no consensus developing regarding larger patterns of fixations to characterize skilled performance.
- When event-triggered fixations are studied to identify more fine-grained connections between fixations and performance, they can reveal failures in monitoring and awareness. Specifically, this approach has been used to reveal monitoring weaknesses tied to autoflight mode awareness.
- Studies that compare the PF and PM have identified strong differences in attention on the ADI but also find that, overall, the PM allocates visual attention in a manner similar to the PF.

This review has also raised a number of questions about how ET research can continue to improve our understanding of skilled performance:

- Are eye tracking data really as precise as researchers think they are? Is there a need to be more conservative with assigning fixations to small AoIs?
- The literature has largely focused on characterizing %DT for the average pilot and looking for markers of skill differences. Now that we have some understanding of general attention allocation strategies (broadly defined), it may be more useful to shift to the more fine-grained analysis of fixations and try to understand how understanding, monitoring, and performance are linked.
- More data are needed before we can make strong claims about fixation duration as a marker of skilled performance for commercial transport pilots.
- If we look outside of fixations on just the core flight instruments, do PF and PM monitoring start to look more different?