NASA/CR–2014-218246

Assessing V&V Processes for Automation with Respect to Vulnerabilities to Loss of Airplane State Awareness

Stephen Whitlow, Chris Wilkinson, and Chris Hamblin
Honeywell International, Inc., Minneapolis, Minnesota

May 2014

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 NASA Aeronautics and Space Database and its public interface, the NASA Technical Report 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 having 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 organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

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 to help@sti.nasa.gov

- Fax your question to the NASA STI Information Desk at 443-757-5803

- Phone the NASA STI Information Desk at 443-757-5802

- Write to:
  STI Information Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320

NASA/CR–2014-218246

Assessing V&V Processes for Automation with Respect to Vulnerabilities to Loss of Airplane State Awareness

Stephen Whitlow, Chris Wilkinson, and Chris Hamblin
Honeywell International, Inc., Minneapolis, Minnesota

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

Prepared for Langley Research Center
under Contract NNL06AA05B

May 2014

Acknowledgments

We would like to thank the Honeywell analysis team and other Honeywell contributors, including Drs. William Roger and Emmanuel Letsu-Dake. We would also like to thank the NASA TPOC, Dr. Steven D. Young, for his technical input and review.

Available from:

NASA Center for AeroSpace Information
7115 Standard Drive
Hanover, MD 21076-1320
443-757-5802

Contents

Acronym List

A&A	Authority & Autonomy
AC	Advisory Circular
ADS-B	Automatic Dependent Surveillance - Broadcast
AOA	Angle of Attack
ARP	Aviation Recommended Practice
ASA	Airplane State Awareness
ATC	Air Traffic Control
CPDLC	Controller-Pilot Data Link Communication
CFR	Code of Federal Regulations
DOF	Degrees of Freedom
EASA	European Aviation Safety Administration
EFB	Electronic Flight Bag
FAA	Federal Aviation Administration
FARs	Federation Aviation Regulations
FMS	Flight Management System
F/O	First Officer
HF	Human Factors
HFE	Human Factors Engineering
HITL	Human in the Loop
HRA	Human Reliability Assessment
HUD	Head Up Display
MBD	Model Based Design
NASA	National Aeronautics and Space Administration
PF	Pilot Flying
PRA	Probabilistic Risk Assessment
SAE	Society of Automotive Engineers
SA	Situational Awareness
SVS	Synthetic Vision System
TIS-B	Traffic Information Services - Broadcast
V&V	Verification and Validation

Introduction

Automation has contributed substantially to the sustained improvement of aviation safety by minimizing the physical workload of the pilot and increasing operational efficiency. Nevertheless, in complex and highly automated aircraft, automation also has unintended consequences. As systems become more complex and the authority and autonomy (A&A) of the automation increases, human operators become relegated to the role of a system supervisor or administrator, a passive role not conducive to maintaining engagement and airplane state awareness (ASA). The consequence is that flight crews can often come to over rely on the automation, become less engaged in the human-machine interaction, and lose awareness of the automation mode under which the aircraft is operating. Likewise, the complexity of the system and automation modes may lead to poor understanding of the interaction between a mode of automation and a particular system configuration or phase of flight. These and other examples of mode confusion often lead to mismanaging the aircraft's energy state or the aircraft deviating from the intended flight path.

Authority, in the context of aircraft operations, refers to having the right, power, or requirement to execute a process associated with a function or action. Autonomy refers to the capability of an agent (human or mechanical) to perform functions/actions independent of other agents. This effort focuses on pilot awareness of those subsystems that are afforded authority and autonomy (A&A) to change aircraft states such as trajectory, modes, power settings, configuration, and status. Examples of subsystems include autopilot, autothrottle, flight guidance, flight management, fuel management, autotrim, and thrust reverser.

ASA is a subset of more global situation awareness (SA) of flight operations that would include environmental factors such as traffic, terrain, and weather. It is an emergent cognitive construct that pilots develop and maintain over time by observing various instruments and displays within the flight deck and integrating this with their mental representation of expected states based on training and flight experience. Airplane State Awareness is a complex emergent cognitive construct within the pilot's mind that:

- Involves attention, mental models, knowledge base, display annunciations, AC state, and evolving situations [1];
- Is impacted by workload, fatigue and stress—as well as situation complexity, system complexity[2]; and
- Includes a mental representation that evolves (and devolves) over time—with both negative and positive feedback loops between system displays and pilot SA.

Many incidents and accidents are related to pilots losing awareness of the modes of subsystems with A&A, also known as mode confusion or automation surprises. Air traffic controllers also form a representation of airplane state awareness for all instrument flights operating in their area of responsibility; however, their airplane state awareness is generally limited to speed, altitude, and separation from other aircraft, consistent with the scope of their responsibility, their limited workload bandwidth, and the capabilities of their surveillance systems. It is worth noting that none of the incidents/accidents reviewed involved improper air traffic control (ATC) monitoring or guidance. Further, ATC is not currently required to maintain state awareness of aircraft systems with A&A. Accordingly, the current approach is tailored to pilot awareness. For current aircraft operations, pilots are responsible for operating the aircraft and ATC is responsible for maintaining separation of the aircraft. Pilots often use automation to improve performance and

efficiency of the aircraft but this can sometimes result in inadvertent conflicts between intended performance and actual performance of the aircraft due to poor situation awareness of the automation's mode (mode awareness) or incomplete understanding of the automation's authority and autonomy. In these cases, the pilot is responsible for monitoring the performance of the automation to assure it performs as intended and to reclaim authority should it not perform as expected, or otherwise "de-couple" per its design logic. This paradigm has worked well and has been demonstrated to be safe for many situations – due largely to well-established validation and verification processes (V&V) for systems and procedures, as well as pilot training per 14CFR [3] and its associated guidance material, advisory circulars, etc. Representative examples of such accidents/incidents are given in Table 2.

It is well known that any process controller (e.g. a pilot) has to maintain an accurate and up to date model of the process being controlled (the aircraft). In control theory, this is termed the process model and in human factors, it is generally referred to as a mental model. There is thus two-way feedback between the pilot and aircraft to keep the pilot's mental model synchronized with the airplane's system model. A common theme among incidents and accidents is that the pilot's mental model of the aircraft state became de-synchronized, i.e. the pilot's mental model of the aircraft's state did not accurately reflect the actual state. The pilot's mental models were inaccurate or incomplete due to a lack of information or poor understanding of the information they were provided. In 2008, the Loss of Control Joint Safety Analysis Team, chartered by the Commercial Aviation Safety Team (CAST), also identified 50 aviation incidents occurring over a period of the previous five years involving energy state management and automation mode awareness [4]. In almost all cases, the flight crews lost awareness of what the automation was doing or was not able to manipulate the automation to resolve the incident. In every case, crews were unable to return the aircraft to the desired flight path in a timely manner.

This situation will likely be exacerbated by the increased levels of automation, system complexity, and operational requirements for NextGen operations [5]. For example, the accuracy, precision, and data interchange requirements of NextGen operations such as Required Navigation Performance (RNP) and Interval Management (IM) will impose substantially higher information processing demands on the flight crew. They will increase cooperative engagement with controllers and require greater precision and responsiveness from the control automation than current-day operations. The increased flight crew responsibilities NextGen will require advances in automation that can support more precise operations, can adapt dynamically to changing situations, and can exercise more authority and control on the flight deck. Flight deck user interfaces must support the dynamic transition of both the authority to exercise aircraft controls and the autonomy to act independently between pilots and flight deck automation. As research indicates, automated systems increase the complexity of human-automation interaction. The potential for accidents due to unanticipated automation behavior and resultant loss of aircraft state awareness will similarly increase. These transitions and interactions have significant safety implications, as surreptitious or frequent transitions may compromise flight crew awareness of aircraft state.

While preliminary work in computational models of pilot awareness have been developed for a circumscribed problem space—taxiway errors—the state of the practice will not support scaling up to the broad category of aircraft state awareness within dynamic, full-mission operations. Aircraft state awareness, including authority/autonomy (A&A) awareness, cannot be reliably modeled or estimated for complex, dynamic situations. Currently, state awareness must be

explored through human in the loop (HITL) evaluations. However, exhaustive HITL evaluations to identify A&A awareness issues would be impractically costly and time-consuming, so any new methodology must include a means to identify test cases to support selective HITL evaluations.

Current validation and verification (V&V) processes include neither HITL evaluations nor V&V of requirements related to aircraft state awareness (i.e., situational awareness by the crew). Such awareness includes the current internal state of the aircraft and systems during nominal and off-nominal scenarios. System-level requirements are typically validated through analysis, which is driven by processes documented by SAE ARP-4754 [6] and its later revision SAE-4754A [7], SAE ARP-4761 [8], to satisfy AC 25.1309 [3] none of which include HITL evaluations. Further, introducing the nearly infinite variations of human cognitive states under stressful conditions produces a combinatory explosion of possible scenarios to be validated and verified. Current V&V practices include selecting only a small number of nominal cases that are tested until failure or success. For future V&V processes, the challenge is to define a subset of off-nominal conditions and scenarios that are manageable from a schedule and cost perspective while insuring sufficient V&V coverage to satisfy the airworthiness regulations. This process will require a revamped approach to V&V that systematically generates requirements related to pilot aircraft state awareness, including HITL evaluations, and which leverages advanced simulation capabilities to verify acceptable levels of human factors (HF) constructs, such as pilot aircraft state awareness across boundary cases for A&A transitions.

The need to address this issue was confirmed by the recent PARC/CAST Flight Deck Automation Working Group (FDAWG) report which identified pilot awareness of system states and improved V&V processes as outstanding needs, as illustrated by the following recommendations [9]:

- Recommendation 2-- Autoflight mode awareness-- confirms that the pilot awareness continues to be compromised by overly-complex autoflight modes.
- Recommendation 5-- V&V for equipment design-- explicitly calls out need to improve processes and method of V&V to address pilots need to respond to non-normal situations on highly-integrated avionics systems, confirming our working hypothesis.
- Recommendation 6- Flight Deck System Design-- identifies an ongoing need to enable pilot awareness of system behavior after failure of another system.

Statement of Work Tasks

This report is organized into four main sections corresponding to the tasks described in the statement of work (SOW).

Task 1: Reviews current processes used by the industry to conduct V&V activities to assure that airplane state awareness is maintained by flight crew in cases where automated systems have been delegated the authority to change airplane state.

Task 2: Discusses a more rigorous and comprehensive methodology for the verification and validation of A&A-management constructs that minimizes the potential for loss of airplane state awareness by flight crew and air traffic control (ATC). Within this method, the role of linked ground-air simulation capabilities is also evaluated and discussed.

Task 3: Describes a generic cost/benefit analysis methodology that would provide decision support to help engineers consider what combination of testing elements, simulated and otherwise, would adequately and efficiently investigate ASA issues for A&A-management constructs.

Task 4: Describes an approach to determine simulation requirements to support HITL evaluations of pilot awareness of complex subsystem states involving A&A management. The approach includes: 1) identifying hazardous scenarios, 2) identifying fragile human-system interaction points that could compromise awareness and 3) considering the span of modeling and simulation environments that may be used to examine scenarios prior to operational service.

Current Processes Used by the Industry

Current V&V practices related to aircraft state awareness were surveyed. Researchers reviewed FAA Advisory Circulars (AC), Federal aviation regulations (FARs), and NextGen operational concepts. Review of these documents quickly revealed that the existing regulations thoroughly address the V&V of automated systems, components, and their integration; however, the same guidance does not exist when evaluating the system's influence on the operator's performance, or the operator's influence on the system's performance. For this reason, regulatory documents from other domains were reviewed in an effort to understand if and how other industries address V&V of human-operated systems. The findings from four domains including FAA, DOD, NRC, and NASA, revealed that although all the domains invoke human factors requirements, human-system interaction is generally not part of the V&V process.

Systems Certification Guidance and Standards

The certification framework for systems (hardware and software) is built around CFR Title 14 (hereafter 14CFR) requirements. Figure 1 shows the general process flow and the applicable de-facto standards of current certification practice and are described in Table 1 along with how they are invoked in certification projects. Note that many of these are invoked simply by being required by the FAA divisions listed. These standards relate to 1) system development, 2) safety assessment and 3) design assurance of system hardware and software. These documents provide guidance on acceptable means of compliance to 14CFR but other methods may be acceptable to the certification authorities if proposed by an applicant. Details of all activities and deliverables to be fully compliant are not shown in interest of focusing on the key steps; these can be found within the documents referenced. The scope here is to give an overview and not a full descriptive narrative. In each case there is a direct equivalence between US and European editions of these documents. These are denoted by SAE/RTCA document numbers and corresponding EuroCAE documents numbers. We refer only to the US editions here for brevity.

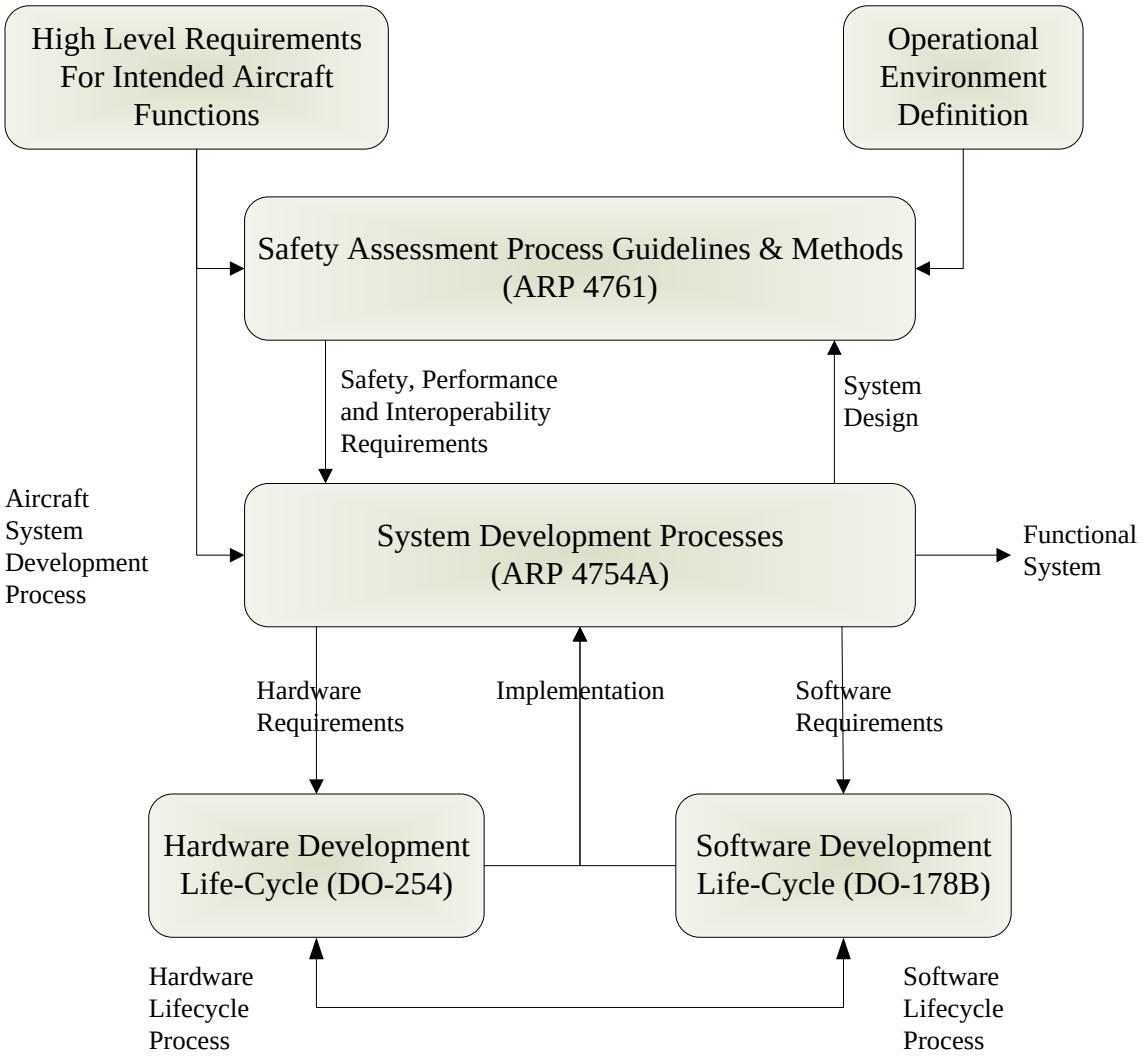

Figure 1 – Certification Process Flow and Applicable Standards

Some revisions of these documents are underway or have been recently completed but are not yet formally invoked by 14CFR through Advisory Circulars. SAE ARP-4761 is under revision by the SAE S-18 Committee and is scheduled for completion in 2014. RTCA released DO-178C in late 2011 [10]. Although not shown in Figure 1, RTCA DO-200A also applies in cases where software (or hardware) utilizes data coming from off-board sources (e.g. navigation databases).

Table 1 – Systems Development Standards

Reference	Description	Applicability	Invocation
SAE ARP-4754A	Guidelines for Development of Civil Aircraft and Systems	Highly-Integrated or Complex Aircraft Systems	No generic invocation at this time. Selectively invoked for some certifications through either customer request (contractual) or IP/CRI process.
SAE ARP-4761	Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment	Aircraft, Systems and hardware components	SAE ARP-4754 (if invoked), commonly accepted as means to support AC25.1309-1A compliance
AC 25.1309-1A or AC 23.1309-()	Describes various acceptable means for showing compliance with the requirements of 14 CFR section 25.1309(b), (c), and (d).	Applies to any system on which compliance with any of those requirements is based. Section 25.1309(b) and (d) specifies required safety levels in qualitative terms, and requires that a safety assessment be made.	FAA ANM-110
RTCA DO-178B/C	Software Considerations in Airborne Systems and Equipment Certification	Provide guidelines for the production of software for airborne systems and equipment that performs its intended function with a level of confidence in safety that complies with airworthiness requirements	TSO, AC 20-115B
Order 8110.49	Software Approval Guidelines	This order guides Aircraft Certification Service (AIR) field offices and Designated Engineering Representatives (DER) on how to apply RTCA/DO-178B, “Software Considerations in Airborne Systems and Equipment Certification,” for approving software used in airborne computers.	FAA AIR-1

AC 20-115B	Radio Technical Commission for Aeronautic, Inc. Document RTCA/DO-178B	Calls attention to RTCA/DO- 178B, "Software Considerations in Airborne Systems and Equipment Certification," issued December 1992. It discusses how the document may be applied with FAA technical standard order (TSO), authorizations, type certification (TC), or supplemental type certification authorization (STC).	FAA AIR-130, Aviation Safety - Aircraft Certification Service, Aircraft Engineering Division
RTCA DO-254	Design Assurance Guidance for Airborne Electronic Hardware	complex custom micro-coded components or programmable logic devices (PLD), such as Field Programmable Gate Arrays (FPGA) and Application Specific Integrated Circuits (ASIC) Note: that DO-254 was written to address all hardware items, but the FAA through AC 20-152 has limited applicability to PLDs.	AC 20-152, TSO's
AC 20-152	RTCA, Inc., Document RTCA/DO-254, Design Assurance Guidance For Airborne Electronic Hardware,	Applies to manufacturers and installers of products or appliances incorporating complex custom micro-coded components with hardware design assurance levels of A, B, and C.	FAA AIR-100, Aircraft Engineering Division, Aircraft Certification Service
Order 8110.105	Simple And Complex Electronic Hardware Approval Guidance	This order explains how FAA can use and apply RTCA/DO-254, Design Assurance Guidance for Airborne Electronic Hardware, when working on certification projects.	FAA AIR-100, Aircraft Engineering Division, Aircraft Certification Service

The close collaboration between the FAA and industry working groups creates consistency within the V&V processes that OEMs employ; however, the role of human factors engineering is not specifically considered in the formalized systems certification process and these topics are given little or no attention in these standards. As such, human factors issues such as ensuring the system promotes situation awareness, minimizes the potential for human error, reduces workload, etc., are not addressed by the industry in a consistent manner. Likewise, the employment of human factors engineers and incorporation of human factors best practices is highly variable across OEMs.

Traditionally, if human factors engineers are used, they are primarily used in the hardware and software development cycles. However, we suggest that earlier involvement of human factors engineers in the overall process has distinct advantages. Involving human factors engineers in the early design processes, including defining operational environments and high level functional requirements, ensures good human factors design practices are incorporated throughout the process.

The process described above defines the operational environment from an engineering perspective when a user-centered perspective may be a necessary complement, or more appropriate. The reductionist nature of the engineering process often loses sight of the end user, focusing more on the functional requirements of the technology rather than the goals and objectives of the user. By defining the operational environment from a user-centered perspective, the goals and objectives of the user can be clearly defined and the functionality needed to safely achieve those goals is consistently addressed during product design and development. Engaging human factors experts in the requirements development phase is another opportunity to ensure good human factors practices are incorporated throughout the program. Traditionally, human factors personnel have had limited opportunity to influence high level requirements, but this practice may be changing as more regulators begin to include human factors requirements in their regulations. The creation of high level human factors requirements will naturally lead to human factors involvement in formal validation efforts as well. The inclusion of human factors personnel in defining the operational environment, developing high level requirements, a participation in verification and validation efforts early in the program is an important step to ensuring human factors constructs such as situation awareness, workload, usability, etc. are addressed throughout the design.

Human Factors Engineers can also contribute to the safety assessment process by ensuring that user-centered perspective is used throughout the various analyses (e.g., proper assumptions regarding the pilots and their tasks are used during the analysis). In addition, human factors experts can conduct human reliability analyses to evaluate the system's design and identify factors that influence human performance. Human Reliability Analysis (HRA) is closely related to Probabilistic Risk Analysis (PRA) methods developed to identify and quantify potential failure modes of complex systems. HRAs apply engineering reliability analysis to the human operator to identify potential opportunities for human error and quantify the probability of their occurrence.

Ideally, hardware and software design processes would involve a collaborative team of designers, flight test pilots, and human factors engineers working together to ensure the design enables sufficient flight crew awareness. Flight test pilots provide operational expertise while human factors engineers provide detailed knowledge of the psychological foundations of situation awareness and the associated design attributes that enable it. However, the level of

human factors involvement varies amongst OEMs. Some OEMs may not employ human factors experts, while others may have entire departments staffed with dedicated human factors engineers. The result is inconsistencies in the application of the human factors knowledge and best practices during design and development as well as the level of scrutiny applied to the type of evaluations conducted during V&V. Some OEMs may conduct comprehensive human factors evaluations, while others may simply rely on the expert judgment of flight test pilots.

The requirements recently published in AC 25.1302 [11] are the first step in ensuring human factors constructs like situation awareness are addressed during the design process. However, additional efforts should be made to provide industry guidance for the inclusion of human factors engineering practices within the existing system certification guidance and standards shown in Figure 1.

V&V of Airplane State Awareness

The development of verifiable human-systems requirements related to situation awareness, usability, and workload can be challenging as it requires quantification of a set of measurable criteria that represent these constructs. While the empirical literature concerning these constructs is well known, the challenge of converting them into verifiable engineering requirements is a daunting one.

If the intention of V&V is to assure situation awareness, then valid and efficient testing depends on clear notions of what situational awareness is and how it can best be tested. Although numerous definitions of situational awareness have been proposed, Endsley's definition [1], "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future," is firmly established and widely accepted.

Aircraft state awareness is an emergent cognitive property that pilots build and maintain over time by observing various instruments and displays within the cockpit. For example, autopilot mode awareness is maintained by observing annunciations displayed on primary flight displays, flight director panels and the FMS MCDU. The best practices for the design and display of autopilot mode annunciations is outlined in AC 25.1329 [12] which includes recommendations for annunciating automation modes, mode changes, and mode transitions. Additional guidance regarding the optimal location of the displays and indications can be found in AC 25.1321 [13].

For human factors engineers, the purpose of verification is to determine that the design conforms to regulations and enables the crew to successfully perform the necessary tasks. Advisory Circular 25.1302 [11] describes five methods to show compliance with the requirements, though for situation awareness, the only applicable methods are evaluations and tests (the only distinction between the methodologies is that tests require a conforming product and system interface). Though not prescribed in the advisory circular, human factors engineers typically assess situation awareness using one or more of the methods described below:

- Subjective Ratings – participants rate their perceived situational awareness [14][15].
- Direct Query - situation awareness of the participants is assessed by questioning their knowledge of particular aspects of the situation [16][1].
- Performance-Based – participants' situation awareness is inferred based on their actions and responses to stimuli [17][18].

Each of these methods has inherent advantages and disadvantages and the rigor of evidence increases from subjective ratings to direct query and performance based methods. Selection of the appropriate method is dependent upon several factors including the novelty, complexity, level of integration, and intended function of the system. It is likely that a combination of methods will be needed to demonstrate a system promotes situation awareness. For example, early in the design process subjective measures may be used to select between various prototypes. While later in the design process, as the fidelity of the prototypes increase, direct query and performance based methods provide more experimental rigor and validity. For many human factors issues and requirements, such as workload and situation awareness, there is no good substitute for the rigor of results from human-in-the-loop studies designed to evaluate integrated, full-mission crew performance under normal and off-nominal situations.

Survey of Industry practices for V&V of Human Factors Constructs

V&V of human factors constructs is a common concern and problem across a broad range of industries that are characterized as safety critical, strictly regulated and incorporate strong interactions between automated systems and human users/operators. A survey of the common current practices and related standards and guidance across the aviation, space, defense and nuclear power sectors is described below.

Federal Aviation Administration

Best practices for V&V of aircraft systems and their components are alluded to in several FAA documents, including regulations (e.g. §25.1301, §25.1309, etc.), advisory circulars (e.g. AC 20-115, AC 20-152, etc.) and Policy Memos (e.g., PS-ANM111-2001-99-01). These documents address the V&V of hardware and software components and their integration but do not address the systems influence on human performance.

Certification of autopilot systems on transport category airplanes is outlined in §25.1329 and TSO-C9C. TSO-C9C invokes industry recommendations for the design of autopilot systems described in SAE-402B. Test and evaluation criteria for certification of autopilot systems is described in AC 25.1329-1B. These criteria are primarily concerned with the effects of autopilot failures on the airplane. The most recent revision to AC 25-7A, "Flight Test Guide for Certification of Transport Category Airplanes," also defines some evaluation criteria for determining whether the autopilot is performing as intended. Policy Memo PS-ANM111-2001-99-01 was issued in 2001 to address incidents and accidents involving pilot-autopilot interactions. The memo provides additional design guidelines to improve flight crew mode awareness, specifically for speed and attitude awareness during operations when the autopilot system is activated.

Unfortunately, none of these documents address authority- and autonomy-management issues from a human-centered perspective. Likewise, none address the issue of mode awareness, a common cause of accidents and incidents involving the autopilot [19]. The guidance cited above implies that that flight crew awareness is maintained by ensuring that the systems perform as designed and provide the necessary operational cues, annunciations, and alerts. Unfortunately, accident reports are rife with examples in which flight crew lost situational awareness even though they were complying with operational procedures and the aircraft was operating as designed and certified [4]. Awareness is an active, complex cognitive process that resides in the

mind of the flight crew; the successful presentation of "situation awareness" information on the flight deck is a necessary but not sufficient condition for pilot awareness.

The documents cited above support traditional system engineering processes and address the V&V of integrated systems as well as their hardware and software components. However, they do not address the systems influence on human performance. The commonly held assumption is that, ensuring the safe and reliable operation of system will result in the safe and reliable performance of the operator. Unfortunately, this system-centered approach does not accurately account for the operator's role in the overall safety of the system. While systems engineers often see the human operator as a source of error and uncertainty, and thus seek to minimize their interaction with the system, they fail to recognize that positive human intervention is often the key to recovering from a system failure. Thus, the systems-centric approach often leads to the overuse of automation and interfaces that fail to keep the operators engaged; which in turn, results in a loss of situational awareness. Current V&V practice dictates tests run until failure, but humans can fail to maintain airplane state awareness and then subsequently recover. The dynamic interaction between pilot performance, mental model, and system annunciations requires a new methodology to handle the explosion in the problem space. Fortunately, the FAA recognizes this bias towards system-centric process and has made strides to release human factors regulations intended to promote a more human-centered design approach.

Regulatory guidance for addressing human factors issues are distributed among several advisory circulars, policy memos, orders and notices. Most recently, the FAA (and EASA) has published AC25.1302, which is intended to minimize pilot error and ensure usability of crew interfaces (FAA, 2013). Of particular interest, 25.1302 (b)(3) states that: "flight deck controls and information intended for flight crew use must enable flight crew awareness of the effects on the airplane or systems resulting from flight crew actions." The guidance within AC25.1302 suggests that applicants seeking certification will need to verify that cockpit technologies ensure crew awareness but does not provide specific guidance for its verification.

Aviation Industry Documents Pertinent to Delegation of Authority and Autonomy

The FAA has acknowledged issues pertaining to delegation of authority and autonomy and has taken steps to address them in the documents below:

- 14 CFR 25.1329 ("Automatic pilot system"), which contains FAA's standards for certifying automatic pilot systems on transport category airplanes;
- 14 CFR 25.1335 ("Flight director systems"), which contains FAA's standards for certifying flight director systems on transport category airplanes; and
- AC 25-11A "Electronic Flight Deck Displays"
- AC 25-1322 "Flight Crew Alerting"
- Advisory Circular (AC) 25-1329-1A ("Automatic Pilot Systems Approval," dated July 8, 1968), which describes an acceptable means by which compliance with the automatic pilot installation requirements of § 25.1329 may be shown.

Department of Defense

It can be argued that the science of human factors engineering was founded by the military during World War I to address new demands placed on soldiers as weaponry became more mechanized and complex. In the decades since, technological advancements of military systems have continued to challenge and evolve the science.

The engineering process used to develop military systems is defined in DoD 5000.02 which invokes processes described in industry standards such as ISO 15288 and IEEE 1220. DoD 5000.02 and the related industry documents describe a five-stage systems engineering process which includes: 1) Concept Development, 2) Technology Development, 3) Production, 4) Utilization, and 5) Retirement. The concept development phase defines the operational requirements that will be used to validate the final product. During engineering development, the supporting system, subsystem, and component level requirements leading to preliminary design and critical design will be iteratively verified through various types of testing and analysis during materialization, integration, and testing. The high-level requirements applied during the concept development phase are primarily defined within MIL-STD-1472 [20].

MIL-STD-1472 has defined human engineering requirements for military systems, subsystems, equipment and facilities since 1989. Its thoroughness has made it a widely cited standard for human factors professionals in almost every industry. Now in its seventh revision, the standard has been updated to address contemporary issues posed by complex systems including psychological constructs such as human-automation interaction (Section 4.12) and situational awareness (Section 5.12). Much like the guidance provided by industry regulators, the requirements in MIL-STD-1472 are intended to serve as high level requirements to be applied to various products to be acquired by the DOD.

The inclusion of situation awareness within MIL-STD-1472 was only included in the most recent version which was released in 2012 so the implications for the V&V process are not well known. However, the US Coast Guard is reportedly planning to conduct comparative evaluations in which human performance on existing systems will serve as the baseline for comparison against all new systems. For example, situation awareness provided by a new radar system onboard a Coast Guard Cutter will be compared to the situation awareness of the existing system. If the new system is determined to be equal to or better than the existing system, it will be accepted. This is similar to FAA certification requirement that new systems support human performance that is no worse than that supported by a related, certified system.

Nuclear Regulatory Commission

The Nuclear Regulatory Commission's (NRC's) Office of Nuclear Reactor Regulation (NRR) evaluates human factors engineering (HFE) programs of applicants for construction permits, operating licenses, standard design certifications, combined operating licenses, and for license amendments. The human factors reviews conducted by the NRC verify that accepted HFE practices and guidelines are incorporated into the applicant’s HFE program. The HFE review includes the design process, the final design, its implementation, and ongoing performance monitoring.

NUREG-0700 [21], much like MIL-STD-1472 and NASA-STD-3000, provides high level requirements and detailed human factors design standards to evaluate nuclear power plant control stations. Psychological constructs such as situational awareness, workload and usability are referenced throughout the document. For example, the high level requirement for situational awareness is described in the general display guidelines:

> 1.1-11 Display of Goal Status

> The information system should provide for global situation awareness (i.e., an overview of the status of all the operator's goals at all times) as well as supplying details about the current specific goal.
>
> NUREG-0711 (NRC, 2003) details how to conduct verification and validation of advanced nuclear power plant designs provided in NUREG-0711. NUREG-6393 provides supplementary guidance for NUREG-0711 and dedicates large sections to the measurement of psychological factors. In particular, section 5.6.2.3.1 is dedicated to defining situation awareness as well as describing techniques to measure it and the advantages and disadvantages of each technique.

While the NRC literature is the most comprehensive in addressing V&V of psychological constructs, much like MIL-STD-1472, it has been put to little practice as no new nuclear facilities have been built in over two decades. It should be noted that guidance provided in NUREG-6393 and the evaluation of psychological constructs are anticipated to pose many challenges to the V&V process [22].

National Aeronautics and Space Administration

Since the inception of the space program, NASA has made deep investments in understanding human performance and incorporating good human factors principles. NASA publishes several documents to ensure proper HMI principles are incorporated into their programs. For example, NASA-STD-3001 [23] is a widely used standard that documents HMI considerations and requirements. Despite the focus on human performance and HMI, the standard does not address psychological constructs of human performance like situational awareness.

Recent human-system integration requirements written for the Constellation Program [24] and the commercial space systems [25] include functional requirements for workload and usability. These documents are unique to other industry regulatory documents in that the requirements specify acceptable thresholds for workload and usability and specify how the requirements will be verified. Interestingly, the introduction for the chapter on display format design (user interfaces), states that the "Display formats must provide situational awareness," yet there is no formal requirements (shall statements) for situational awareness within the document.

NASA's human factors requirements are unique in their prescriptive nature in that they specify success criteria as well as the methods and tools to be used during the V&V process. It could be argued that NASA very rarely commissions a new spacecraft and when they do, very few are built; therefore, the repercussion of prescriptive requirements is less burdensome than other domains. Nonetheless, these types of requirements do pose a certain amount of programmatic risk that must be acknowledged and accounted for. Most notably, the methodologies defined in NASA's requirements to measure workload and usability are subjective, making them more susceptible to biases of the participants and inherent variability of human behavior. The inability to mathematically predict the outcome of human-in-the-loop evaluations may create unease among program managers and increase the risk of the program failing verification evaluations. To be clear, these issues are not insurmountable, but they do require vigilance and close collaboration between the developer and the evaluator.

Authority Sharing Cue Sufficiency

Based on our review, current industry practices do not specifically and systematically address awareness of authority/autonomy modes. While AC25.1302 requires certified flight deck

systems to "enable flight crew awareness", it is too vague to realistically address the subset of awareness related to authority and autonomy sharing or management.

To investigate this, several incidents/accidents were selected for review to identify potential gaps in the V&V process, see Appendix A: Accident/Incident Reports. Official accident reports and Aviation Safety Reporting System (ASRS) reports were reviewed to understand the involvement of pilot state awareness. It is our conclusion that the automation mode cues operated as designed in all incidents/accidents. In spite of this, it appears that pilots understanding of the airplane state did not reflect the actual state, often contributing to the accidents. This was also the case for incidents reported in ASRS such as altitude busts (ASRS 113722, 1989), as well as automation surprise research [26][27]. In these cases, it was confirmed that the interaction and display between the flight crew and the flight deck did not sufficiently support awareness of A&A sharing.

Results and Discussion

The challenge of maintaining situation awareness in work domains that include complex automated systems is a ubiquitous one. Government agencies across several domains have recognized the issue and are attempting to address it by incorporating human factors principles into their regulations and system engineering requirements. The nature of this challenge was confirmed during a Human Factors and Ergonomics Society (HFES) Panel discussion chaired by the Honeywell team in October 2013[1], where it was recognized that existing verification and validation practices are generally not intended to address the system's influence on human performance or the influence of human performance on the system, and that new engineering V&V processes may be needed to accommodate human factors requirements.

Regulations intended to address psychological constructs (e.g., situation awareness, workload, usability, etc.) create new opportunities and challenges. On one hand, the regulations provide authority for human factors engineers to enforce good human factors principles. On the other hand, such regulations pose a certain amount of risk to the V&V process which must be acknowledged and accounted for.

Verifying and validating human factors requirements can be costly and time consuming to perform. They require additional analyses to be performed, mockups and simulators to be built, and ideally, the employment of human factors specialists. In addition, V&V of psychological constructs like situation awareness require human-in-the-loop evaluations which incur additional complexity and cost. The additional time and expense associated with human factors evaluations are often unwelcome burdens to program managers, especially given the potential risk the requirements pose.

Human factors requirement also create requirements traceability challenges for the hardware and software designers. The high-level requirements defined in the various regulatory documents discussed above are decomposed into functional requirements and applied to hardware components and software specifications in a manner that can be traced from each element back to the high level requirement. However, human factors requirements, especially those addressing psychological constructs, can only be realistically/meaningfully tested at an integrated level. This poses challenges for the deconstructive V&V process at the lowest levels

[1] SD5 – VERIFICATION AND VALIDATION: HUMAN FACTORS REQUIREMENTS AND PERFORMANCE EVALUATION (http://www.hfes.org/web/HFESMeetings/HFES_2013_AM_Program.pdf)

of decomposition. This is especially the case for constructs such as situational awareness, which emerges from the perception, comprehension, and projection of elements within the environment [1]. For instance, how does one design a test to verify that an individual icon, light, button, etc., improves situational awareness or reduces workload? At some point in the decomposition process, psychological constructs are no longer verifiable creating traceability issues for requirements managers and V&V engineers.

The evaluation tools and methodologies used during verification are an additional challenge. Performance-based evaluations are viewed to have the most external validity but are typically cost prohibitive and reserved for final test of only the most critical systems. Most often, the tool used to assess psychological constructs are subjective measures as they offer the greatest efficiency. Tools such as NASA-TLX [28], Bedford Workload Scale [29], SART [14], SA-SWORD [15], etc. are all commonly used tools; however, they are subjective measures and thus prone to the inherent variability of human judgment and biases. Methodology issues combined with relatively small sample sizes used during the evaluations can impose a great deal or risk to the verification and validation of a product. Thus, a poorly designed questionnaire, misapplied tool, or even a disapproving test pilot can jeopardize an evaluation and the entire V&V effort.

Regardless of the methodology used, or the outcome, human factors evaluations are often viewed with skepticism by traditional engineers based solely on statistical issues. At their best, behavioral statistics can provide statistical significance levels of 10^{-3} which pales in comparison to engineering evaluations that measure reliability on the order of 10^{-7} or 10^{-9}. This sense of scale can make it difficult to convince authorities that results from a human factors evaluation will generalize to the equivalent performance during day-to-day operations.

Enhanced Methods and the Role of Testing

In the previous section we reviewed the current practices for systems development, the relevant guidance material, and V&V practices in aerospace regarding ASA and A&A management. A common feature of current and upcoming human factors (HF) V&V practices is lack of formalism. They rely mostly on the expert judgment of users (pilots) and engineers to establish what activities (i.e. verification scenarios) are required, what pass/fail criteria are appropriate and when those activities can be considered to be adequately complete. Typically, HF engineers are not included in the development and validation of formalized, system requirements or to the development of verification test cases except in an ad hoc fashion. Whilst this is adequate for evaluating new systems; however, it would not seem adequate for evaluating the system's effect on human performance or influence of human performance on the system.

This section comprises an analysis to determine the role of linked ground/flight testing during V&V of new A&A constructs. The analysis begins with a discussion of a generalized structured, formalized, and repeatable process for evaluating A&A issues. This process augments current methods and provides a more rigorous and comprehensive methodology for the design of such systems (including the HF requirements). It is within the context of this presumed process, that the role of modeling, simulation, and testing is discussed.

Examples of ASA Related Accidents/Incidents

The list of incidents in Table 2 is representative of incidents where loss of airplane state awareness was a contributory factor. We focus on events where on-board system failure was not a direct, proximate cause. In some cases system failure (those marked with a *) contributed to the loss but the loss was avoidable had the crew had proper ASA and thus possibly could have taken timely corrective action. Incorrect, missing or delayed actions can, in part, be attributed to a lack of understanding by one or more crew members of the automation's current state. This loss of ASA is variously caused by loss of environment awareness, mode confusion, automation surprise or loss of aircraft state awareness.

Table 2 – Aircraft Incidents Related to ASA

Incident	Airplane State Awareness Issue
China Airlines Airbus A300B4-622R, on approach to Nagoya Airport, Japan, April 1994 [30]	Crew was unaware of autoland mode
*Predator B UA Crash, Nogales, AZ, April 2006 [31], [32]	Crew was unaware that the fuel supply had been accidentally cut
American Airlines Flight 965 B757 near Cali, Bogota, December 1995[33][34]	Crew was unaware that that the FMS had put them on an offset parallel track
Comair Flight 5191 Bombardier CL-600-2B19, Attempted Takeoff from Wrong Runway, Lexington, Kentucky, August 2006 [35]	Crew were unaware they had lined up for takeoff on a wrong runway too short for takeoff
Northwest Airlines A320 Flight No NW188, N374NW, Overflight of Minneapolis Airport, October 2009 [36]	Crew were unaware that 1) the ATC radio channel was mistuned, 2) ATC could not contact them and 3) they were unaware they had over flown a waypoint
Colgan Air Flight 3407 Bombardier DHC-8-400, Loss of Control on Approach under icing conditions, Clarence Center, New York, February 2009 [37]	Crew did not report icing conditions so Colgan AOC was unaware of icing conditions and provided incorrect approach speed recommendation
Kenya Airways B737-800, Douala, Cameroon, May 2007[38]	Crew was unaware of a gradually increasing roll angle
*Air France Flight AF 447 Airbus A330-203 loss of control, Rio de Janeiro to Paris, Atlantic Ocean, June 2009 [39]	Crew were confused by 'unreliable airspeed' warnings and ignored or were unaware of standby indicators
*Turkish Airlines Boeing 737-800 Crashed during approach, near Schiphol Airport, Amsterdam, Holland , February 2009 [40]	Crew did not react in a timely manner to premature autothrottle thrust reduction following failure of one radar altimeter
*XL Airways delivery flight Airbus A320-232, Accident off the coast of Canet-Plage, France, November 2008 [41]	Crew were unaware that 2/3 AOA sensors were frozen and of the resulting AFCS mode
Many cases of 'altitude bust' in ASRS where FMC cannot fly to constraints, busts flight plan constraints, e.g. ASRS 113722, 1989 and [26][27]	Crew did not comprehend the mode logic operation of the autoflight system

The common conclusion from these is that additional system requirements could have improved crew ASA and thereby possibly avoided or mitigated the resulting loss. The question then arises; what additional requirements could have been specified during the early stages of system

definition and how could those requirements have then been verified during detailed system design and evaluation?

Several tools and methodologies may be used during verification of psychological constructs. The most commonly-used techniques are summarized below.

Human Reliability Assessment (HRA)

HRAs apply engineering reliability analysis to the human operator to identify potential opportunities for human error and quantify the probability of their occurrence. HRAs can be used as a design tool or an accident investigation tool. *Prospective* HRAs are conducted during the design phase and is used to assess the probability of an event happening and allow designers to improve the design accordingly. *Retrospective* analyses evaluate systems that have already been designed or evaluate events that have already occurred in order to determine the likelihood that something could or should have happened. Prospective and Retrospective analysis use the same methodology. Retrospective analyses are used for accident investigation and therefore have the advantage of hindsight in that they know what the outcome was. Likewise, the conditions of the accident and the design of the systems are already established. Prospective analyses are used to make guide systems design decisions. During systems design the analysts may be faced with many unknowns and therefore and must rely on the foresight of the analyst to predict potential failure modes and estimate error probabilities. Since this report is intended to address the certification of new products, the reviews below focus on prospective HRAs.

HRAs also provide the ability to explore and analyze different high-risk scenarios. This is especially desirable when considering *"edge of the envelope"* scenarios which may be highly improbably but extremely critical situations or conditions which could be difficult or impossible to evaluate with human-in-the-loop (HITL) experiments. Iterative use of HRAs can be used to determine the sensitivity or "brittleness" of systems and tasks to human error and potentially identify opportunities to make them more resilient.

While the use of HRAs provides advantages, they do have limitations and pose potential pitfalls if the methods are not performed correctly or the results are used inappropriately. Most notably, the quantification of human error probabilities should be interpreted with caution. One of the primary arguments against the use of HRAs is that the resulting probabilities do not accurately reflect actual human error rates or system safety. Another concern is that HRA methods assume that humans fail in the same manner as systems or their components. Considerable evidence has shown that this is not the case [42].

Another criticism of HRA methods is that they can only analyze failures that the analysts can foresee. Unfortunately, humans are the most complex and least understood part of any engineered system, susceptible to a more diverse range of failure modes than any other component, making it impossible to foresee and analyze every possible failure.

Finally, HRA methods see humans as a source error and conclusions from analyses often seek to eliminate or minimize the human's interaction with the system. This bias fails to acknowledge that positive human intervention can prevent a system from failing, recover a failed system, or manage a failed system. HRA methods could provide insight into "edge of the envelope" scenarios, as long as its results are interpreted with caution and used in conjunction with complementary methods.

The Use of Subjective Measures

Most often, the tool used to assess psychological constructs are subjective measures as they offer the greatest efficiency. Tools such as NASA-TLX, Bedford Workload Scale, SART, SWORD, etc. are all commonly used tools; however, they are subjective measures and thus prone to the inherent variability of human judgment and individual opinion. Methodology issues combined with relatively small sample sizes can impose a great deal or risk to the certification of a product. Thus, a poorly designed questionnaire, misapplied tool, or even a disapproving test pilot can jeopardize an evaluation and the entire certification effort.

Current guidance does provide some correlation between the expected outcome of a failure and the probability of such a failure, e.g. AC 25.1309 [43]. For safety critical systems, a loss of function probability of 1E-9/fh is the guidance figure. It must be remembered that this is a guidance figure for engineering purposes and the real airworthiness requirement is that a single point failure shall not, under any conceivable circumstances, lead to a catastrophic loss. We therefore view the standard probabilistic methods for safety assurance as necessary but not sufficient and think that a process orientated towards discovering 'conceivable circumstances' to be more promising for the discovery of HF system requirements.

In the development of any system, a well-worn process of developing the requirements and then verifying that the implementation has correctly implemented all of them is generally followed. The 'develop requirements' step also needs a process to establish that the requirements so developed are complete and consistent to an adequate extent. This latter step is often referred to as requirements validation or validation for short. Thus validation and verification are distinct and separate activities with different objectives. These principles are enshrined in the 14CFR airworthiness regulations and the structure of their associated guidance documents as we reported previously.

In the development of hardware and software for safety critical systems, it is apparent that validation is a critical step upon which all subsequent design and verification activities depend. An examination of various accident reports, such as those in the previous section, indicates that at least some can be attributed to faulty or missing requirements rather than faulty design or verification. A subset can be traced to unforeseen human error which in turn may be regarded as faulty or missing HF requirements that should be levied on the system itself.

System-level Human Factors Requirements

With regard to certification, 14 CFR 25.1302 (c) includes a situation awareness requirement that states: "flight deck systems shall enable flight crew awareness". This is not specific enough to address the challenges related to airplane state awareness. While proposed new system requirements could technically be covered by existing, generic requirements, the level of detail is not sufficient for use in design and evaluation. To augment this generic requirement, a proposed approach should include generating more detailed system requirements based on a wealth of HF research into automation awareness in particular. For example:

All flight deck systems given a degree of authority and/or autonomy for changing the trajectory of the airplane, shall:

- Annunciate all normal and abnormal disconnects and disengagements (does not fail silently).
- Prominently annunciate all mode transitions.

- Provide indication of most recently changed mode setting.
- Provide immediate feedback if pilot is providing opposing control input to autoflight system.
- Flight mode annunciation panel shall efficiently convey relevant mode states and be comprehensible "at a glance"
- Where practicable, identify reasons why mode activation is not allowed.

System level requirements of this type would not suffer from the challenges of most HF performance requirements which tend to be unverifiable or rely on subjective measures.

Mechanized Approaches

In addition to these analytic HF methods, a comprehensive approach should include more mechanized approaches that systematically and objectively evaluate A&A based on HF principles related to ASA. Current practice relies heavily on the judgment of system engineers and test pilots to identify unsafe scenarios and actions; however, their analysis can be biased by their expertise leading them to discount scenarios that may be catastrophic but highly improbable. Mechanized approaches dispassionately evaluate system characteristics, ignoring the relative frequency of certain system states and without human biases. For example, [44] suggests a model checking method to identify scenarios where there is likely to be a mismatch between the pilot's mental model and airplane state for the 'kill the capture' automation surprise in the MD-88 autopilot. A possible mechanized approach is to identify test scenarios where disconnect is likely and map out the mode logic and all of its transitions in a state machine representation for a given automated system and then 'run' a set of pilot mental model heuristics against it to identify violations. Experts would develop heuristics that are representative of pilot expectations and biases while being diagnostic to identify areas of disconnect. The assumption is that, even under ideal conditions, some automation logic or mode transition behavior could violate pilot expectations.

Computed-based Modeling Methods

When studying human-system interaction, the number of variables that can influence human performance is often too large to allow empirical assessment of all the possibilities. This is especially true when studying complex systems or complex cognitive constructs like SA. Computer-based modeling of human performance has been proposed as an alternative method to HITL evaluations to help explore a vast problem space [45].

Modeling tools and techniques have been applied to replicate and study various aspects of cognition including perception, motor control, learning, and decision making. Modeling tools have also been developed to analyze and evaluate human performance during various tasks. However, most modeling methods have failed to successfully replicate the cognitive aspects of SA. Likewise, task analysis models that are intended to evaluate work allocation often fail to model the dynamic nature of authority- and autonomy-management in the aviation environment.

One modeling method is addressing this shortcoming by applying dynamic computational modeling to simulate work by multiple agents in complex dynamic systems. Agents (human or mechanical) are modeled as responding to, and changing, their environment [46][47]. The authors describe their simulation as models doing work where work is defined as "purposeful activity of acting on, and responding to, the environment as required by the situation". This work is performed by automated and human agents and involves both cognitive and physical activity

whereby agents evaluate the situation to select the appropriate sets of actions. Work is thus a response to the situation, with strategies chosen in response to the physical environment, the allocation of responsibility within the team; and agent status including expertise, the demands on the agent, and resources available to the agent such as time and information. As such, this methodology provides the ability to dynamically simulate A&A issues as situated within a dynamic environment driving and responding to human or mechanical agent activity[48].

Synergistic Modeling and Simulation

We recognize that computer-based modeling of human behavior is very difficult and itself requires verification, therefore it can be advantageous to add multiple levels of simulation fidelity to the computer based modeling methods. This allows scenarios to be postulated, possible crew errors of timing, omission or commission to be postulated, initially analyzed through safety assessment methods and finally verified within a simulated environment. Throughout this process, additional requirements will emerge which can be considered for removal through equipment design change or mitigation through crew training or procedural modifications.

A synergistic approach to evaluating ASA and A&A-management issues fuses two methods, 1) model-based design and safety analysis and 2) simulation within a simulation facility or group of linked facilities that include human-in-the-loop. Recognizing that simulator time can be expensive, this scheme puts as much as possible of the load into the lower cost computer-based environment and therefore maximize the productivity of the high level simulators. The general scheme proposed is shown in Figure 2. The process is iterative and may be started and stopped at any point depending on the level of assurance required of the system.

The selection of model or simulator is specific to the hazard under investigation. It is the initial safety analysis that provides the initial hazards which are then refined by process iterations, employing the appropriate models and simulators as required.

The use of models provides three key features; 1) it allows for the representation of rare events that could not be realistically or safely reproduced in test flying and 2) it permits the generation of test cases that can be presented to simulation facilities with multiple and arbitrary degrees of closeness to actual operations and 3) allows rapid iteration of multiple variables that may detract or contribute the pilots' performance. Executing the models and scenarios within a simulation environment allows confirmation of the predicted effect and also generates test results that can be used to refine the models and create additional scenarios.

The major benefit of this approach is that it builds on itself; system safety analysis can create rare event scenarios which in turn reveal missing requirements. By an iterative process, the model is extended and improved, new requirements are discovered and generated and appropriate test cases developed for use in later verification activities.

Traditional system safety analysis methods such as the example PRA methods in ARP-4761 [8], now incorporated in 14CFR guidance, are somewhat weak in representing complex interactions, feedback, and a degenerating safety state that are typical of the rarer type of safety event as typified in the accidents we reviewed. The safety analysis step in this process should be extended with additional safety analysis (e.g. STAMP/STPA [49]) be performed in addition to the industry standard PRA. Further work is required to include HF considerations within the safety analysis

methods, since none of those presently known are capable of generating HF related system safety requirements.

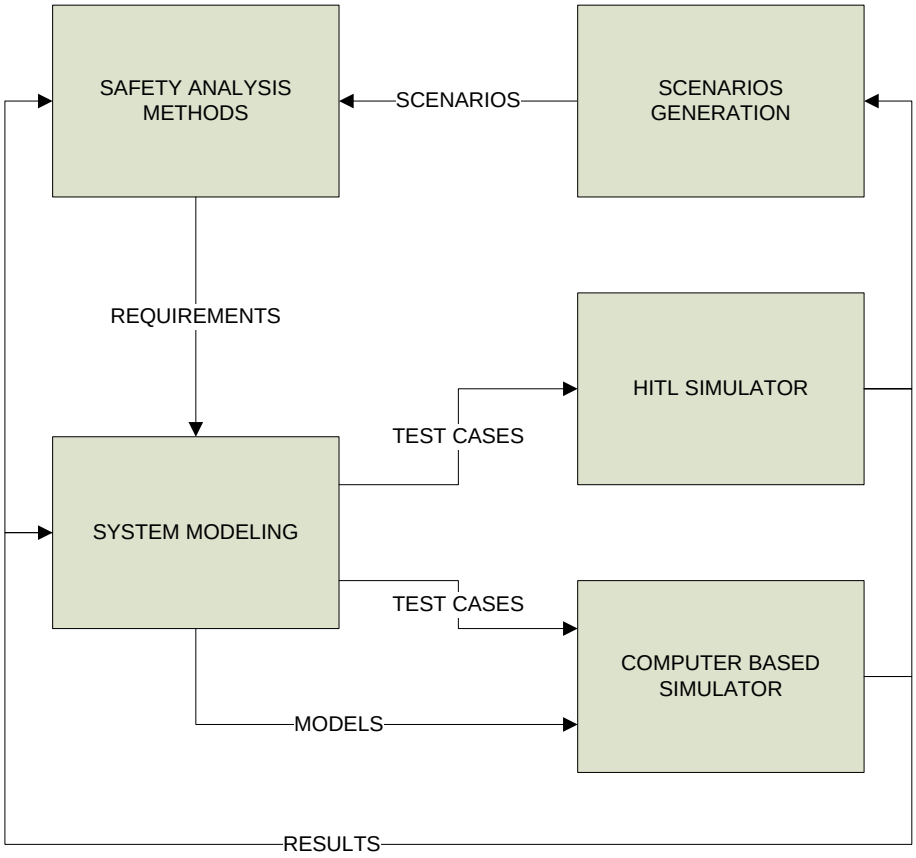

Figure 2 – Model Based Design and Simulation

The notion of model in the above is quite general. In our parlance, a model is simply some representation of the object being modeled. By this definition, a model may be formalized mathematical construct such as a mode logic state machine, some representative hardware incorporating some elements of aircraft and airborne equipment such as a full motion simulator, or the aircraft and all its systems A safety model may represent the closed loop response model of an airspace procedure, e.g. ITP [50].

The generation of requirements through this process results in modifications to the design of equipment, procedures and training to eliminate or mitigate the potential for the occurrence of safety events. Such requirements include the HF cues provided to the crew to inform them of system state and the backup warnings should a crew action (or inaction) be incompatible with the system state, thereby driving the system into a hazardous state. This is the approach we recommend to infer crew SA since methods for direct modeling of crew SA constructs have proved problematic in the past.

We note that expert judgment still maintains a place in this methodology since the iterative process cannot be known to be complete. The main benefit is that progress is made in incremental steps, documented along the way, each step building on the previous one. System modifications made as the process iterates are more easily rechecked as design proceeds without having to redo large parts of the safety analysis. Since the larger ‘systems’ will be in a state of

almost continuous development as NextGen continues to evolve, this methodology provides for system modifications to be easily re-checked for hazards without having to go back to 'square-one' on each change.

High Level Requirements for Simulation Facilities

Simulation facilities need to be selected and configured to address specific hazards; there is no generic simulation specification that covers all possible hazards. The use of simulators is useful for confirmation that a given test case will produce a system hazard. Simulator is a broad term that encompasses the small and focused, e.g. a symbology simulator, to large airspace simulators that may include both ground based and airborne assets. The intent is that low level modeling tools would perform in 'fast time' and the complex and expensive 'slow time' simulators would be confined to a confirmatory role.

The Figure 2 illustrates two basic classes of simulation facilities:

- A computer based tool that executes a given model and provide results from given test cases.
- A high-level simulator that represents a typical cockpit including pilots and external feeds of operational data. This class of simulator requires that it be presented with a test case in the form of a scenario definition. Such simulators may be "linked" either to other simulators, or to operational platforms (e.g. aircraft or ATC towers)

A review of NASA and FAA data has located a considerable number of US facilities that offer high level simulation facilities. These are listed in Appendix C: Selected Simulation Facilities at FAA and NASA. The following are representative capabilities that are suggested but the subset of these required will be dependent on the precise problem being investigated.

- Pilot symbology generator.
- Formal methods analysis of state machine representations of a multi-LRU system, e.g. theorem prover, model checker, abstraction.
- Full motion 6-DOF large transport cockpit with representative equipment, e.g. displays (EFB, SVS, HUD), mode control panel, standby instruments.
- Weather data feed.
- ATC communications utilizing voice and CPDLC.
- Multi-target generator for TIS-B and ADS-B traffic data, live from airborne assets or from other linked simulation facilities.
- Flight data, communications and voice recording.
- Pilot, co-pilot eye tracking.
- Fault injection to simulate functional failure of individual high level functions, e.g. blank displays, primary power loss, hydraulic pressure loss, sensor fault/icing.

Anticipated Impact of Using Linked Ground/Flight Testing

Although flight simulators are typically thought of as pilot training devices, they also play a key role in aircraft systems research and development (including V&V). Simulators provide the unique ability to test the impact of a new system in a controlled environment and better understand the capabilities and limitations of a system before it is fully developed and deployed.

As complexity and level of integration of aviation systems increases, the role of simulators as tools within the V& V process will most likely increase.

One issue that must be addressed is the level of fidelity the simulator must provide in order to address the testing objectives at hand. The simulator must provide an adequate level of fidelity to replicate the critical aspects of the flight in order to ensure that the test findings generalize to the actual flight environment with an acceptable level of confidence. If the goal of the simulator is to evaluate systems impacts on a pilot's ASA, then it is essential to understand what perceptual and cognitive experiences the simulator must provide in order to elicit behaviors that may occur in the actual flight environment. In the training domain, the generalization of behaviors from a simulator to the actual operational environment is referred to as *transfer*.

Flight Simulator Fidelity

The amount of simulator fidelity required to transfer behavior from one environment to the next has been debated since it first received attention in the beginning of the 20th Century (See Thorndike and Woodworth, 1901; Judd, 1908 for early debates). The debates have primarily focused on the types of fidelity a simulator can provide and their effect on learning and transfer of training. The overall fidelity of a simulator is defined by four variables. Physical fidelity refers to how closely physical components look and feel like the actual aircraft. Visual fidelity most often refers to the realism of the environment when the pilot looks out the windows. Motion fidelity refers to the extent to which the motion forces of the simulator match those of the actual flight environment. Last, cognitive fidelity refers to extent to which the simulator engages the same cognitive processes (e.g., attention, workload, situational awareness, etc.) as the actual flight deck in an operational environment.

Studying pilot awareness, or the loss thereof, presents several unique issues to consider when determining the level of simulator fidelity required. Situation awareness is a cognitive phenomenon that the pilot develops by collecting information from the environment and integrating into a cognitive model of the situation. Pilots may integrate information from a variety of sources. In terms of simulator fidelity, physical fidelity may be critical for determining how the location of information in the flight deck influences situation awareness as displays and controls not within the primary field of view may be attended to less often or may not be attended to during periods of high stress or workload. In some scenarios the use of motion may help understand the role of proprioception in building or maintaining aircraft state awareness (e.g., attitude changes, turbulence, stalls, etc.). Visual fidelity can be important for scenarios involving environmental factors (e.g., terrain, traffic, weather, etc.). Last but not least, cognitive fidelity is of the upmost importance as situational awareness can be volatile and easily effected by stress, workload, attentional demands etc.

There is a dramatic range of flight testing/simulation platforms, including PC-based desktop trainers, full motion certified simulators (e.g. Level D), modified operational aircraft, and network-based connected simulators that link operational aircraft to ground-based simulators. In general, ground-based simulators allow a level of safety and experimental control along with reduced cost; while flight evaluations provide higher fidelity but at a higher cost with less safety and less experimental control. Both airborne test aircraft and high-fidelity simulators can engender realistic levels of pilot workload and stress and expose the pilot to realistic environmental dynamics, thus maximizing cognitive fidelity.

Cognitive Fidelity

To highlight the importance of cognitive fidelity, under realistic workload levels, pilots are less likely to monitor the output and state of automated systems, thus increasing likelihood that their awareness de-couples from the actual system state. For this analysis, maximizing cognitive fidelity would be a benefit since it is the dimension that most impacts ASA in operations. Some high level cognitive factors that contribute to loss of ASA include:

- Under sampling sources that provide awareness (narrowing of attention and degradation of task management) due to:
    - Low workload (loss of vigilance)
    - High workload
    - Stress
    - Distraction
- Inadequate mode annunciation indications.
    - Complex displays that are difficult to parse quickly
    - Low salience of mode change indications relative to flight deck visual environment.
- Inaccurate/incomplete mental model (pilots infer airplane state based on their experience and/or understanding of automation mode logic).
    - Automated system mode logic complexity & intuitiveness
    - Pilot knowledge/skill base (e.g., level of proficiency, overall knowledge, practical experience, familiarity with aircraft, exposure to unusual situations, etc.).
- Conformity of flight deck interface

To illustrate the importance of cognitive fidelity, we consider the narrowing of attention phenomena. It is well known that workload and stress can narrow pilots' attention which could increase the likelihood that they fail to fully sample the flight deck environment to maintain airplane state awareness, as illustrated by Figure 3. Accordingly, varied and realistic workload and stress levels could be considered a highly beneficial element for HITL evaluations, to be weighed against the cost to create a test setup that provides adequate realism.

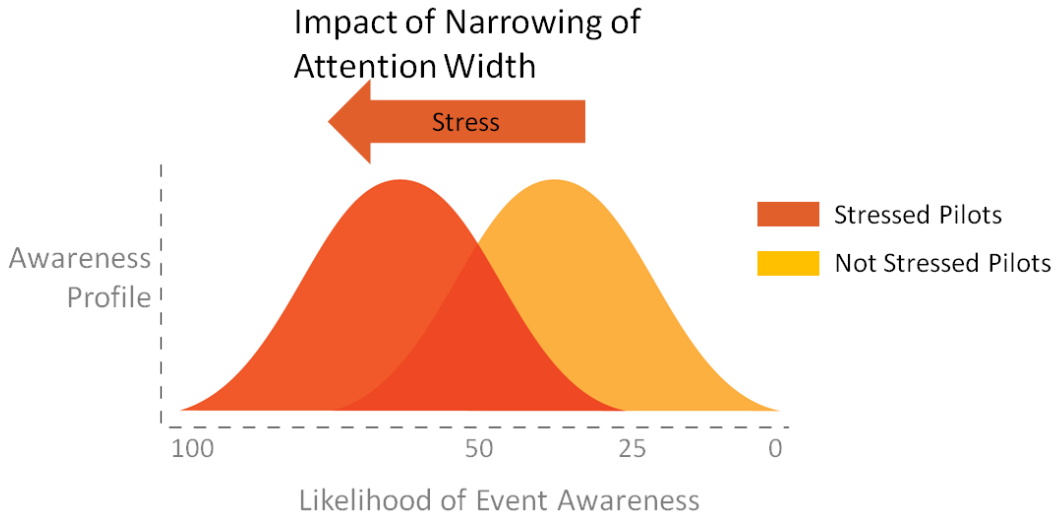

Figure 3: Attention under Stress

This also illustrates the risk of failing to identify an ASA issue due to limitations in the test setup. If the testing scenarios do not adequately induce workload and stress, test pilots could exhibit higher levels of awareness than would be expected under realistic conditions. This could lead the testers to inappropriately conclude that there are no ASA issues with the evaluated design. Consequently, the issue may not be identified until later in the development and certification process, incurring greater re-design, development, and certification schedule and monetary costs. Worse yet, the issue could be first identified as the result of an incident or accident once fielded, incurring very expensive re-design efforts, damage to reputation, and exposure to legal liabilities.

Costs and Benefit for Testing Options

A cost/benefit analysis is integral to decisions regarding the required level of test environment fidelity. Once you have identified system elements involved with A&A management that are of concern and should be tested, a cost/benefit analysis can provide some guidance on how to configure the testing setup. This can help streamline the process by scoping the testing elements that are integrated to be commensurate with testing objectives, in this case the cognitive fidelity related to managing A&A and maintaining ASA.

Costs include the development time, evaluation schedule, personnel and equipment, and operating costs such as fuel. The key advantage to linked facility testing is that it provides another layer of fidelity; specifically, the ability to evaluate large scale A&A issues and their effect on the air traffic system. This will be needed to fully evaluate and test NextGen technologies and operations where roles and responsibilities between pilots and ATC will become more co-dependent and traffic is anticipated to be more dense. To enable comparison of different testing setup options, some quantification or categorical assessment is required for the cost and benefits elements. In general, cost is operationalized as a categorical, rough order of magnitude estimate of dollars costs of operating costs of, personnel time to integrate with, accessibility to simulation capability, and project schedule impact. Another testing cost is exposure to unsafe conditions. While it is difficult to quantify this in terms of dollar costs, we can assume it is desirable to avoid expose testing personnel and the public to unsafe operating situations.

With regard to estimated costs, Honeywell's experience is that flight test operations cost range from $3.5K to $5.0K an hour depending on the air frame, while FAA certified Level D simulator costs approximately $2.0K an hour. These do not include the substantial personnel cost for designing, executing, and analyzing results from tests. For PC-based simulators, the operating costs are negligible but the modification and integration costs could be similar to flight tests.

Table 3 captures rules-of-thumb estimates of relative costs for different testing platforms.

	Flight Test	High Fidelity Simulator	PC-based Simulator
Operating Costs	High	Medium	Low
Personnel Costs	High	Medium	Low
Schedule Cost	High	High	Low

Table 3: Estimate of Relative Cost Impact

Likewise, some benefits, such as support of certification efforts, can also be estimated in terms of dollars savings and possibly reduced time to market. This assumes that an applicant would be

expending financial, personnel, and time to conduct similar evaluation and documentation efforts if not done in the proposed V&V evaluation and testing process. There is a less direct mapping between cognitive fidelity and monetary estimates. In including testing elements germane to ASA for a given system or subsystem, the primary, direct benefit is to increases confidence that the testing findings are ecologically valid, and thus can be generalized to real-world operational settings.

A less direct mapping could be between an unrealized benefit, from a course of action not taken, that could become a cost should the applicant fail to find an ASA issue and continue to mature the design. In this scenario, the issue would be found later in the certification process or after fielded, thus dramatically increasing the cost to address this. The worst case, and less direct mapping, would be an ASA issue that contributes to some incident or accident, resulting in damage to reputation, legal liability risk, and possible survivor benefits.

Relevance to FAA AC 25.1302

The potential for certification benefits from this proposed new V&V process has increased following the issue of FAA AC 25.1302: Installed Systems and Equipment for Use by the Flight crew on May 3, 2013 [11]. This AC includes design guidance explicitly for supporting flight crew awareness as well as specifying how applicants can demonstrate compliance. The guidance within the AC includes an emphasis of the use of HITL evaluations as a means of showing compliance with the human factors related requirements.

Examples of such design guidance include the following:

- Uncommanded mode changes and reversions should have sufficient annunciation, indication, or display information to provide awareness of uncommanded changes of the engaged or armed mode of a system (§ 25.1302(b)(3),5-6: System Behavior: C: System Functional Behavior: 3:b:4).
- The automated system must, per § 25.1302(b) (3), support flight crew coordination and cooperation by ensuring shared awareness of system status and flight crew inputs to the system, if required for safe operation (5-6: System Behavior: B: System Function Allocation: 7: c).Section 25.1302 (b) requires flight deck controls and information intended for the flight crew use be provided in a clear and unambiguous form, at a resolution and precision appropriate to the task. The flight deck controls and information must be accessible and usable by the flight crew (e.g. including all lighting conditions and all phases of flight) in a manner consistent with the urgency, frequency, and duration of their tasks, and must enable flight crew awareness, if awareness is required for safe operation, of the effects on the airplane or systems resulting from flight crew actions (5-1: Overview: f: 2).

When weighing the costs and benefits of different testing elements, one should keep in mind that some options, especially those including conformal flight deck interfaces, could support compliance per AC 25.1302. However, conformal flight test hardware and software are not available until late in the design cycle during which design changes can be quite costly.

HITL evaluations should be documented to serve as a means of compliance. In some cases, testing involving conforming components (product/system, flight deck, and/or system interface) could also serve as a means of compliance. However, given the range of evaluation scope

options and that conformity is not required; evaluations would likely be the most selected course of action and therefore a common benefit that need be considered.

For those testing configurations that would meet the requirements for compliance, benefits could include reductions in certification duration and costs. This assumes that most applicants would be conducting various HITL evaluations as part of their standard product evaluations processes, so this would then reduce the scope of certification tasks, thus reducing their duration and cost.

Examples

It is a common question as to what level of hardware conformance is sufficient for HITL evaluations. For example, evaluating autoflight functionality would require feedback on the presentation of active autoflight modes on a Mode Control Panel (MCP). It is an open question whether the exact A/C specific panel is needed or whether a software display-based virtual panel would suffice. Human Factors experts within test teams would do a cost/benefit on the inclusion of conformal hardware panels, resulting in output that could resemble the following:

Costs:

- Schedule delay—until hardware panel is available
- Integration cost—install hardware panel in simulator cab, integrate with simulation software

Benefits:

- Minimal increase in ecological validity between hardware panel presentation and virtual panel; software display panel, if in same location, would engender equivalent cognitive fidelity with regard to ASA as hardware panel.

Conclusion

- For ASA evaluations, the costs of hardware panel integration do not warrant the minimal benefits to ecological validity.

Another question is whether the exact panel needs be within an airborne platform or in a ground-based simulator. Given the relative simplicity of a MCP design, it is unlikely that airborne testing would be required to evaluate the panel itself. However, if the test relates to dynamic airplane state awareness, presentation of MCP information is a necessary but not sufficient condition. The Human Factors perspective looks at it three ways 1) does it behave in a meaningful way, 2) is the signal detectable under conceivable operational environments and 3) is the signal clear and unambiguous. For this evaluation, ground-based simulator would be sufficient provided the testing scenarios include realistic levels of workload to assess whether the presentation of autoflight state satisfies 2 & 3.

Table 4 depicts the outcome of a cost/benefit analysis for the MCP evaluation example. The full table with assessments of factors, such as workload, distraction, and complex automation logic, can be seen in Appendix B: Cost/Benefit Analysis for MCP Testing Example. The results of the assessment can be seen in the rightmost column which reflects the number of ASA-related factors on which the testing component has a medium or high impact.

Testing Component	Type	Test Setup Option 1: MCP	Cost Estimate	# Med or High

		Evaluation		Impact
Automation Simulation	Test Software Platform	R	M	6
Configurable All Glass Flight Deck Interface	Test Software Platform	R	M	5
Conformal flight deck user interfaces	Test Software Platform	O	H	6
Part-Task Simulation	Test Scenario Element	R	M	5
Full Mission Simulation (high fidelity in procedures, actors, and roles)	Test Scenario Element		H	5
Realistic Flight Deck Workload Support	Test Scenario Element	R	M	4
ATC Operator Station	Test Hardware Platform		n/a	0
ATM HIL Simulator	Test Hardware Platform		n/a	0
Conformal physical hardware interfaces	Test Hardware Platform		M	1
Desktop Simulator	Test Hardware Platform		L	0
High Physical Fidelity (displays, controls, AC dynamics)	Test Hardware Platform		H	2
Fixed Based Simulator	Test Hardware Platform	R	M	3
Motion Platform	Test Hardware Platform		H	1
Flight Data	Test Data		H	0
Airport conditions	Simulated Test Scenario Element		M	0
ATC Simulation	Simulated Test Scenario Element		M	2
ATM Simulation	Simulated Test Scenario Element		M	0
Traffic Simulator	Simulated Test Scenario Element	O	L	3
Voice Communications	Simulated Test Scenario Element	O	H	3

Weather Simulator	Simulated Test Scenario Element	O	L	3
Connectivity to Remote Simulation Capabilities	Connectivity	O	M	3
Linked Air to Ground via AC Telemetry	Connectivity		M	0
Live Traffic Linked to Ground Station	Connectivity		M	0

Table 4: Results of Cost/Benefit Analysis for MCP Testing example

For this example, the following test setup would result from the cost/benefit analysis:

- Purpose: evaluating autoflight mode presentation on MCP
- Simulation environment: ground, fixed-based, part-task simulator with configurable glass cockpit (e.g. NASA LaRC Integrated Flight Deck Simulator)
- Testing scenarios: induce realistic levels of pilot workload:
    - Traffic: option--simulated is sufficient
    - Weather: option that could further increase workload
    - Two-crew setup: include Crew Resource Management (CRM) to induce workload overhead and AC state monitoring is pilot monitoring responsibility
    - Scope: focus on phases where there are mode transitions, such as take off and approach; include off-nominal scenarios identified by earlier hazard analysis, such as Go-Around, subsystem failures, etc.
    - Workload: realistic level

Given the added complexity of resource dependencies and system integration, Linked air-ground testing will almost always increase the cost of a test, relative to either ground-based only and airborne only, in terms of schedule, budget, resources availability, and personnel. Across the range of HITL evaluations that can vary greatly in complexity and realism, a case could be made that for most ASA testing scenarios, ground-based simulation tests should be sufficient when considering the distributed capabilities available. This is especially true given the conventional flight test setup which is highly constrained by safety restrictions, typically only expose pilots to nominal situations, and involve test pilots who are very familiar with the system under evaluation, if not part of the product team. This conventional flight test setup essentially limits the scope of benefits since ASA issues often occur at the edge of the operational environment, including subsystem failures, higher risk phases of flight, and pilot errors—all elements that are not available in flight tests. Engineers have more latitude in ground-based simulators to simulate subsystem failures and induce off-nominal operations. Given these factors, it is hard to imagine a cost/benefit analysis that favors linked ground to air testing for broad application.

However, there are some noteworthy exceptions where classes of benefits have been identified for linked air-ground testing paradigm. The following classes are worth consideration relative to cost of linking:

- Support testing conditions to possibly enable certification credit for V&V artifacts developed during flight tests

- Introduce environmental elements, such as traffic and weather, with increased control and safety (e.g., FAA William J. Hughes Technical Center Target Generation Facility (WJHTC TGF).
    - Experimental control to manipulate test scenarios in ways that are impossible or impractical in the real world.
        - Virtual traffic generated by ground-based traffic simulator
        - Simulated weather
        - Manipulate ATC voice and/or data communications.
    - Increase safety margin by using ground-based simulated weather and traffic, exposing airborne test platform to fewer hazards.
- ATM automation design evaluation
    - Airborne platform could support realistic evaluation by acting as intruder within live traffic, provide measure of control in generating alert conditions that would unlikely happen by chance with unlinked evaluation
- Linking ground ATC test station (WJHTC Experiment Operator Station (EOS) to live traffic to evaluate new operator tools and displays):
    - Evaluation of real-time operational data from airborne platforms (e.g. energy state) on controller ASA
    - Highest level of realism of traffic dynamics and weather impact.
- Unmanned Air System (UAS) within US National Airspace (NAS) evaluations
    - Manned airborne platform could link with ground control and simulation to provide critical safety oversight (line of sight),including remote control
    - E.g., UAS cargo flights in Class B airspace
- Special case: substitute simulated environmental reality to pilot via flight deck while obscuring out the window (OTW) view:
    - Limiting the field of vision of pilots and forcing them to use only the flight instruments simulating the conditions of low ceiling of clouds, heavy fog, night, and other instrument meteorological conditions (IMC).
    - Pilots must rely on instruments for awareness.
    - Can increase safety margin of evaluating approach scenarios at a much higher altitude—while presenting a lower altitude reality for the pilot via the flight deck
    - As a form of experimental deception, expect increased scrutiny from internal and external review boards.
    - Presenting virtual terrain and weather on airborne flight deck.

We anticipate NextGen changes to create additional benefit cases for linking facilities and aircraft since roles and responsibilities between pilots and ATC will become more co-dependent and traffic is anticipated to be more dense. For NextGen operations, the interaction between the pilots, aircraft and ATC will become much more interdependent. Further, it is envisioned that some of the interaction will be mediated by ground and airborne automated systems, adding a new class of A&A-management systems and interactions for which the pilots must maintain awareness and proficiency.

Linked simulators allow researchers to evaluate the 2nd or even 3rd order effects of a more condensed and interdependent air traffic management system. For example, the repercussions of a delayed ATC clearance or a pilot's failure to notify ATC of a deviation may be far more disruptive in NextGen operations; if a pilot deviates from a clearance in NextGen operations, it will most likely have a far more intrusive effect on the neighboring traffic and the air traffic

controller trying to manage them. Traditional (e.g., unlinked) simulators can be used to evaluate human-machine interaction issues (SA, vigilance, workload, usability, etc.), and human-human issues (CRM, etc.). The key advantage to linked facility testing is that it provides another layer of fidelity with a broader scope that is commensurate with the future vision for a more "connected" and interdependent NAS. Specifically, the ability to evaluate large scale A&A-management issues, related to distributed automation and control, and their effect on the entire air traffic system.

Determining Simulation Requirements and Test Planning

Determining simulation requirements is based upon an assumed approach for identifying and testing pilot awareness of complex subsystem states involving A&A management. The approach includes: 1) identifying hazardous scenarios, 2) identifying fragile human-system interaction points that could compromise awareness and 3) considering the span of modeling and simulation environments that may be used to examine scenarios prior to operational service. The approach is expanded on below and illustrated by two examples from published accident reports and one example of future operations.

Identifying Hazardous Scenarios

The analysis follows the STAMP methodology [49] of first identifying high level hazards (collision, Loss of Control (LOC), Controlled Flight into Terrain (CFIT) etc.) and then identifying control actions that could lead to them. This is done for each of the flight phases (takeoff, climb, cruise etc.) and considering the major aircraft systems that are relevant to that phase in order to narrow the problem scope. For example, the braking system is not relevant during the climb/cruise decent phases but is during takeoff and landing. Conversely, incorrect thrust reverse actuation is potentially hazardous during all phases. The analysis leads to the identification of hazardous scenarios that determine the required simulation facility capabilities and the span of test cases and conditions. The simulation will then have two roles, 1) to first confirm the hazard potential and 2) to test system modifications (i.e. additional requirements) designed to mitigate the identified hazards.

During this analysis, the pilot is considered as one node of the system; the pilot is able to provide/not provide, provide too early/too late the control actions that could lead to the high level hazard. Such 'errors' by the pilot can be caused by loss of SA, mode confusion; or stated more generally, the loss of synchronization between the actual system state or mode and the pilot's mental model of system state, such as can occur during A&A transfers.

Identifying Fragile Human-System Interaction Points

Current practice relies heavily on the expert judgment of system engineers and test pilots to identify unsafe scenarios and actions; however, this method is inherently limited by biases which can lead them to discount highly improbable but critical tasks such as those identified in the JCAST report (2008). Mechanized approaches dispassionately evaluate system characteristics, ignoring the relative frequency of certain system states and without human biases. For example, Rushby suggests a model checking method to identify points where there is likely to be a mismatch between the pilot's mental model and airplane state, inducing automation surprise in the MD-88 autopilot [44]. A possible mechanized approach is to identify test scenarios where disconnect is likely and map out the mode logic and all of its transitions in a state machine

representation for a given automated system and then 'run' a set of pilot mental model heuristics against it to identify violations. Experts would develop heuristics that are representative of pilot expectations and biases while being diagnostic to identify areas of disconnect. The assumption is that, even under ideal conditions, some automation logic or mode transition behavior could violate pilot expectations.

The output of this approach would identify points where awareness (e.g. the perceived state) is likely to deviate from the actual state. The project team would evaluate the results to determine:

1. For each point, do they believe these are likely to pose awareness problems? If not, document rationale. If so, proceed to Step 2.
2. Whether new requirement(s) would be warranted. If so, develop a new system requirement that addresses the identified issue. If not, proceed to Step 3.
3. Consider a HITL evaluation to test whether issue does in fact pose an awareness issue for pilots.

Modeling and Simulation

When studying human-system interaction, the number of variables that can influence human performance is often too large to allow empirical assessment of all the possibilities. This is especially true when studying complex systems or complex cognitive constructs. Computer-based modeling of human performance could identify edge of the envelope situations that are unlikely to be considered by traditional analyses.

Computer-based modeling of human behavior is very difficult and itself requires verification, so typically multiple levels of simulation fidelity are employed to achieve confidence in the computer based modeling results. The selection of model or simulator is specific to the hazard under investigation. It is the initial safety analysis that provides the initial hazards which are then refined by process iterations, employing the appropriate models and simulators as required.

A synergistic approach to studying A&A transfer issues is assumed here that includes: 1) model-based design and safety analysis and 2) simulation within a simulation facility or group of linked facilities that include HITL evaluation. Recognizing that high fidelity simulator time is expensive, as much of the V&V as possible is performed in the lower cost PC-based environments; this also serves to maximize the productivity of the high fidelity simulators. The generic scheme is shown in **Error! Reference source not found.**. The process is iterative in that t may be entered and left at any point. Throughout this process, additional requirements will emerge which can be considered by the project team in a process similar to the one described above.

Cost/Benefit of Simulator Testing

A cost/benefit analysis is integral to decisions regarding required level of test environment fidelity (see Appendix C: Selected Simulation Facilities at FAA and NASA). See Section: Anticipated Impact of Using Linked Ground/Flight Testing , for a detailed description of the cost/benefit analysis. Once system elements involved with A&A management and that are of concern have been identified, they should be tested. A cost/benefit analysis can provide some guidance on how to configure the testing setup. This can help streamline the process by scoping the testing elements that are integrated to be commensurate with testing objectives, in this case the cognitive fidelity related to maintaining SA (for A&A management elements).

Example Applications of the Methodology

Table 2 provides a list of selected accidents where loss of airplane awareness was a proximate cause. We expand on two representative examples and use these to discuss simulation design issues and considerations.

China Airlines Airbus A300B4-622R, Nagoya Japan (1994)

The first example [30] illustrates a case where the crew lost awareness of the autopilot mode during a landing, causing them to attempt to manually oppose the autopilot. No equipment failures contributed to this accident.

Accident Summary: On approach to Nagoya airport, the go-around lever was accidentally and unknowingly engaged. This accidental mode selection caused the autopilot to command the Trimmable Horizontal Stabilizer (THS) to apply full nose up trim and increase thrust. Unaware of the go around mode being selected or the full nose up trim applied by the THS, the First Officer (F/O) attempted to resume the expected attitude and flight path by applying forward pressure to the control column. In the go around mode, the aural warning of THS motion was inhibited by design and therefore the crew did not get notification of the mode selection. The aircraft continued to climb with decreasing speed and increasing angle of attack (AOA) until it stalled and then rapidly descended and crashed. The aircraft was a total loss and 264 of the 271 occupants were killed.

A previous Airbus service bulletin (SB) had been issued on the autopilot to disconnect if a large force was applied to the control column whilst above 400 ft and in go-around mode. The SB had not been implemented on this aircraft and did not identify itself as safety related and was therefore categorized by China Airlines as an ‘on maintenance’ item.

This accident has elements of loss of awareness (go-around mode confusion) and A&A transfer conditions (automatic full nose up trim in go-around mode).

Following the steps previously discussed:

1. Identifying hazardous scenarios: Hazard analysis for autopilot would identify control actions across phases of flights. For approach phase, incorrect pilot input would be accidental actuation of go-around lever, actuating TOGA autopilot mode.
2. Identifying Fragile Human-System Interaction Points: Mapping of mode logic space would reveal inhibition of aural THS alert during approach as a fragile state. Mechanized approach would more likely identify this than SME since it involves an accidental actuation that most expert pilots would not even consider.
3. Modeling and Simulation Scenario generation (approach, go around selection, over-ride of THS pitch up). Two crew with one being experimental confederate, the other being the pilots whose responses are under evaluation. Confederate is necessary to “accidentally” activate go-around lever.
4. Cost/Benefit Analysis: Simulator elements critical to A&A aspects of ASA would include conformal visual and auditory representation of flight deck, full motion to provide some feedback regarding trimmable horizontal stabilizer (THS) motion, part-task to support approach phase evaluation; Results: part-task, ground-based, full motion simulator with configurable glass would be sufficient.

Simulation Environment: The scenario leading to this accident could be simulated in a full motion simulator such as those routinely used for pilot training. There would not be any need for external inputs such as traffic or weather. The results from this test would likely indicate the hazard potential of this scenario and highlight that 1) the aural warning inhibit of THS motion in go around mode and 2) lack of automatic autopilot disconnect when excessive forward pressure was applied to the control column in go around mode were potential hazards. This realization could then lead to a reconsideration of system requirements and inform what other requirement(s) would mitigate the hazard. At a next higher level, the scenario could also be confirmed to be hazardous on a real aircraft by artificially moving ground level to a safe altitude and performing the identified scenario.

Turkish Airlines Boeing 737-800, Schiphol Airport, Holland (2009)

The second example [40] concerns a situation that was initially caused by a system failure in turn causing the crew to lose SA. This accident also revealed a previously unknown flaw in the design of the autopilot in relation to how primary/secondary radar altimeter information is used.

Accident Summary: During an approach at 1950 feet, a faulty captain-side (left) radio altimeter suddenly failed, causing the autopilot to believe the altitude was -8ft. This caused the autothrottle to (correctly) decrease engine power to the 'retard flare' low power setting. This should only occur during flare-out at 27ft, just prior to touch down. This failure should have logged an error and transferred data sourcing to the right-side radio altimeter, which was reading correctly. It did not and continued to be the source of altitude data for the autothrottle and other systems. The crew had no understanding (loss of SA) of the conflicting effect on the autothrottle of an undetected failure of one radio altimeter whilst the other continued to perform correctly.

The PF was the F/O (right-seat) and therefore the autopilot in control was also the right-side. The right-side autopilot received altitude data from the still functioning right-side radio altimeter and thus attempted to keep the aircraft flying on the glide path for as long as possible. This meant that, unnoticed by the crew, the aircraft's nose continued to rise, creating an increasing AOA as the autopilot attempted to maintain lift as the airspeed reduced.

The aircraft continued to descend and slow until the stick shaker stall warning activated. When the stick shaker went off, the captain took control. There seemed to be some confusion between the captain and F/O in this transfer of control (authority). The F/O released the throttles, which caused the autothrottle to return to the idle setting, correctly according to its design. There was some delay until the captain disconnected the autothrottle and commanded full thrust. By this time there was insufficient altitude (350ft) or forward speed available to effect recovery.

The basic cause was that the left-side radio altimeter failed undetected thus defeating the dual redundancy. The deeper cause was the system design that allowed the autothrottle and autopilot to operate from a failed radio altimeter sensor and the consequential loss of crew SA following this failure. Had the crew been aware of the failure mode and its implications for this particular system design, recovery action could have been taken in time to avoid the accident. The same analysis steps would be followed as in the previous example.

Simulation Environment: The scenario leading to this accident could be simulated using the same process described in the previous example. A safety analysis and scenario generation would be necessary to configure the simulation setup for the necessary test conditions.

Trajectory Based Operations (TBO)

As an example of simulation-assisted safety assessment of A&A-management scenarios, we consider trajectory based operations (TBO) [51][52]. TBO is a major component of future NextGen and SESAR that frees operators from many of the constraints of current operations and flow management. Pre-defined 4D (i.e. position, height and time) trajectories are defined for aircraft to fly based on the operator's view of an optimum (e.g. lowest cost) trajectory. This functionality is provided by the use of existing FMS systems installed on most Part 23 aircraft and a considerable proportion of Part 25 aircraft. TBO is primarily aimed at en-route operations, though it has potential to be extended in TMA operations in conjunction with planned merging and spacing operational improvements.

TBO will allow operators to define pre-planned trajectories that are coordinated by air traffic management (ATM) using system-wide information management (SWIM) to assure conflict free trajectories. Trajectories will be based initially on nominal conditions, not allowing for external events, e.g. weather, airport closure, in flight emergencies. Off-nominal conditions will require that the pre-defined trajectories of many aircraft be amended in a coordinated fashion to remain conflict free. We may therefore consider that if one aircraft requires a trajectory change then several others may be required to alter their trajectories also.

This implies that aircraft automation may perform the trajectory change and that all affected pilots must retain SA throughout. Trajectory changes may be temporary e.g. a response to a TCAS RA or permanent such that the aircraft follows a new trajectory to its destination. There is obvious potential for A&A transfers that could lead to loss of SA under this circumstance.

Scenarios of this type are potential candidates for the use of linked simulation facilities during the development of aircraft systems, TBO procedures and ground based ATM facilities such as SWIM.

To investigate scenarios that could lead to loss of SA under off-nominal TBO situations, an analysis and simulation setup would be needed that provides or links to the following facilities.

- ATM and Traffic simulation incorporating SWIM.
- Weather feed.
- OTW view, SVS, HUD as required.
- Navigation equipment (e.g. GPS, FMS, AFCS).
- Separation assurance equipment (e.g. ADS-B, TCAS/ACAS).
- Communications equipment (e.g. Voice, CPDLC).
- Representative crewed flight deck with supporting autoflight components.

As in the previous examples, identifying hazards generates specific scenarios that are confirmed in the simulation and lead to additional safety requirements for implementation in systems and procedures.

Many scenarios could be generated, e.g.

- A TCAS RA causes the aircraft to deviate from the approved trajectory.
- Coordinated avoidance action by other aircraft in the vicinity in order to maintain minimum separation.
- Re-planned trajectory to accommodate an off-nominal situation (e.g. airport closure, airspace avoidance) with concomitant changes to other affected aircraft trajectories.

After these assessments, a cost/benefit analysis can inform the decision on which simulation components are appropriate for the different scenarios. For example, in the case of the TCAS RA scenario, it could be determined that the potential safety cost of live traffic does not warrant the benefit of realism. Accordingly, ground-based traffic simulators could provide virtual traffic for an airborne AC for the HITL evaluation. In addition to the safety afforded, more experimental control could be exercised over the simulated traffic and subsequent coordinated response to a TCAS RA.

Requirements for Linked Simulation Facilities

Clearly simulation facilities need to be selected and configured to test for vulnerabilities to specific hazards; there is no generic simulation specification that covers all possible hazards. The use of simulators is beneficial for confirmation that a given test case will produce a system hazard. Simulator is a broad term that encompasses the small and focused, e.g. a symbology simulator, to large airspace simulators that may include both ground based and airborne assets. The intent is that low level modeling tools would perform in 'fast time' and the complex and expensive 'slow time' simulators would be confined to a confirmatory role.

Figure 2 illustrates two classes of simulation facility spanning fidelity and missions:

- A computer based tool that executes a given model and provide results from given test cases.
- A high-level group of linked simulators that represent one or more cockpits including pilots, at least one ATC/ATM facility including controllers, and external feeds of operational data. This class of simulator requires that it be presented with a test case in the form of a scenario definition.

A review of NASA and FAA data has located a considerable number of US facilities that offer high level simulation facilities of the types mentioned above. These are listed in Appendix C: Selected Simulation Facilities at FAA and NASA.

The following are representative capabilities that are suggested but the subset of these required will be dependent on the precise problem being investigated.

- Pilot symbology generator.
- Formal methods analysis of state machine representations of a multi-LRU system, e.g. theorem prover, model checker, abstraction.
- Full motion 6-DOF large transport cockpit with representative equipment, e.g. displays (EFB, SVS, HUD), mode control panel, standby instruments.
- Weather data feed.
- ATC communications utilizing voice and CPDLC.
- Multi-target generator for TIS-B and ADS-B traffic data, live from airborne assets or from other linked simulation facilities.
- Flight data, communications and voice recording.
- Pilot, co-pilot eye tracking.
- Fault injection to simulate functional failure of individual high level functions, e.g. blank displays, primary power loss, hydraulic pressure loss, sensor fault/icing.

When considered for use during V&V, simulator capabilities should be evaluated based on their costs and benefits. Costs include the development time and resources required to integrate some simulator capability; benefits include improved performance for a dimension of interest such as cognitive fidelity, coverage of problem space, and safety. In general, cost is operationalized as a

categorical, rough order of magnitude estimate of dollars costs of operating costs of, personnel time to integrate with, accessibility to simulation capability, and project schedule impact. Another testing cost is exposure to unsafe conditions. While it is difficult to quantify this in terms of dollar costs, we can assume it is desirable to avoid expose testing personnel and the public to unsafe operating situations.

The overall setup of the suggested scheme is shown in Figure 4. The boxed section represents conventional current practice except that we recommend that STAMP be additionally used to augment the current safety assessment processes suggested in ARP-4761 in order to improve hazard analysis under off-nominal conditions. Many hazards will be identified and removed within this conventional process without the need for simulation. Problems typified by A&A failures are typically very difficult to find through this conventional process,

The technique is focused on a particular system defined by some system boundary. This is not an overly restrictive constraint since the boundary is selected to contain the system of interest. This is necessary to scope the problem and to generate scenarios that are hazardous for the system so defined. System may here be considered to consist of a subset of crew, LRUs, operating procedures, protocols and if necessary, the surrounding managerial structures. Repeated applications of this process are needed to walk the system boundaries out so far as is considered prudent.

The benefits arising from this suggested approach are that defining the hazardous scenarios results in the creation of specific, focused test cases being provided to the simulation setup so that expected outcomes are known in advance. Results that differ from expected indicate that new system requirements are necessary to avoid or mitigate the hazard.

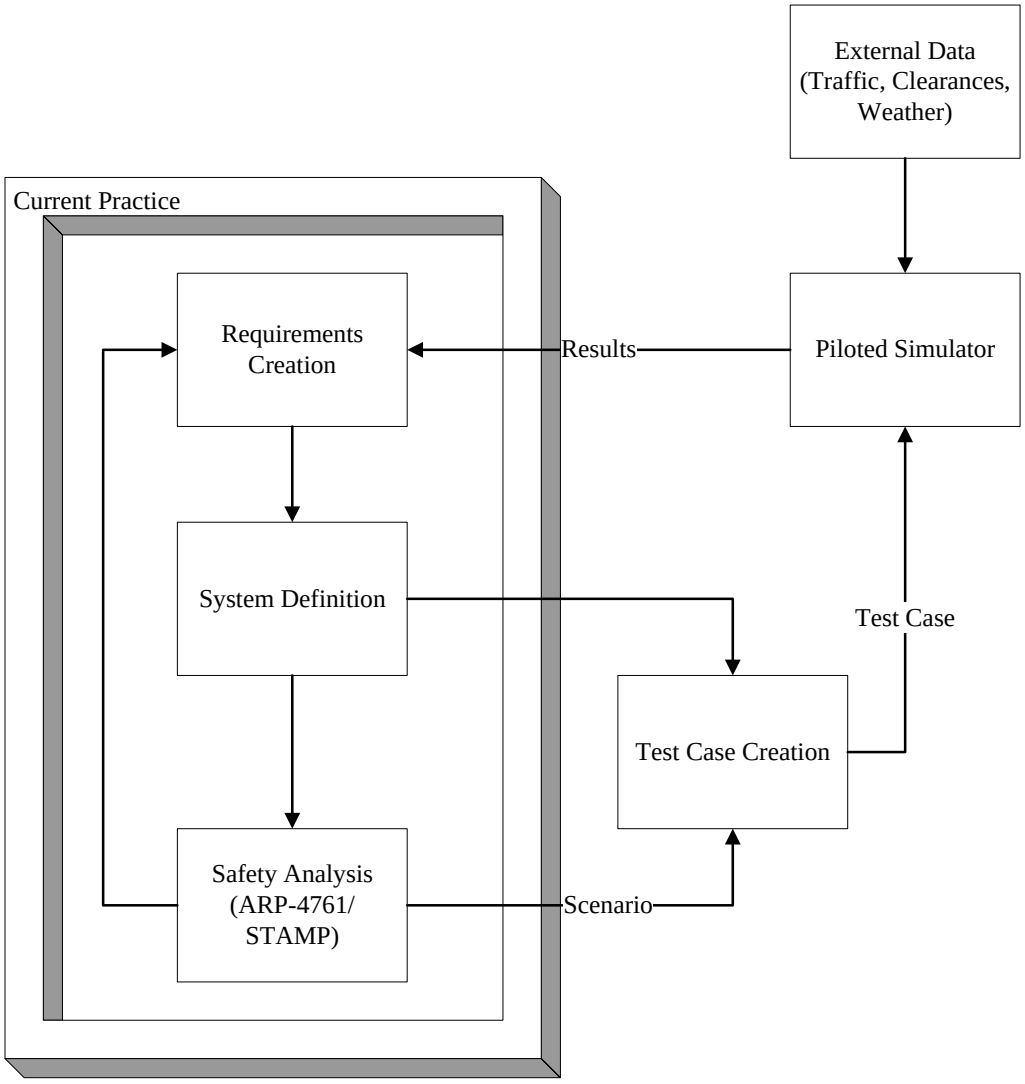

Figure 4 – Simulation Environment

Conclusions

Based on our review, current industry practices do not specifically and systematically address awareness of authority-/autonomy-management dynamics and modes. More importantly, current engineering processes tend to be system-centric only rather than considering human-centric along with a systems perspective. As a result, system designs fail to account for issues that arise due to poor human-system interaction. The commonly held assumption is that, ensuring the safe and reliable operation of system will result in the safe and reliable performance of the operator. Unfortunately, this system-centered approach does not accurately account for the operator's role in the overall safety of the system. An analysis reported in [4] concluded that incidents and accidents can occur even when systems are operating correctly. What is needed is a systematic process for identifying potential human-system interaction failures which can be used to define verifiable human factors requirements.

High level human factors requirements have been invoked by regulatory agencies across a variety of domains (e.g., DoD, FAA, NASA, NRC, FDA). However, human factors requirements, especially those involving cognitive constructs such as situation awareness pose unique challenges that are not currently accommodated by existing V&V processes.

To address this, a more rigorous and comprehensive V&V methodology is discussed for complex automated systems that minimize the potential for loss of airplane state awareness by flight crews. It is a synergistic approach that combines and builds upon two methods, 1) model-based design and safety analysis and 2) simulation within a simulation facility or group of linked facilities that include human-in-the-loop. Recognizing that simulator time can be expensive, this scheme puts as much of the load as possible into the lower cost computer-based environment and therefore maximize the productivity of the high level simulators. The selection of model or simulator is specific to the hazard under investigation. It is the initial safety analysis that provides the initial hazards which are then refined by process iterations, employing the appropriate models and simulators as required.

A cost/benefit analysis is integral to decisions regarding the required level of test environment fidelity. Once you have identified system elements involved with A&A management that are of concern and should be tested, a cost/benefit analysis can provide some guidance on how to configure the testing setup. This can help streamline the process by scoping the testing elements that are integrated to be commensurate with testing objectives, in this case the cognitive fidelity related to managing A&A and maintaining ASA.

Given the added complexity of resource dependencies and system integration, linked air-ground testing will almost always increase the cost of a test, relative to either ground-based only and airborne only, in terms of schedule, budget, resources availability, and personnel. However, we anticipate NextGen changes to create additional benefit cases for linking facilities and aircraft since roles and responsibilities between pilots and ATC will become more co-dependent and traffic is anticipated to be more dense. For NextGen operations, the interaction between the pilots, aircraft and ATC will become much more interdependent. Further, it is envisioned that some of the interaction will be mediated by ground and airborne automated systems, adding a new class of A&A-management systems and interactions for which the pilots must maintain awareness and proficiency. Linked simulators allow researchers to evaluate the 2nd or even 3rd order effects of a more condensed and interdependent air traffic management system. For example, the repercussions of a delayed ATC clearance or a pilot's failure to notify ATC of a deviation may be far more disruptive in NextGen operations

Determining simulation requirements is based upon an assumed approach for identifying and testing pilot awareness of complex subsystem states involving A&A management. The approach includes: 1) identifying hazardous scenarios, 2) identifying fragile human-system interaction points that could compromise awareness and 3) considering the span of modeling and simulation environments that may be used to examine scenarios prior to operational service.

References

[1] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 37, no. 1, pp. 32–64, 1995.

[2] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs," *J. Cogn. Eng. Decis. Mak.*, vol. 2, no. 2, pp. 140–160, 2008.

[3] FAA, "14 CFR Part 121 Air Carrier Certification," 2012.

[4] Commercial Aviation Safety Team, "Mode Awareness and Energy State Management Aspects of Flight Deck Automation: Final Report," Safety Enhancement 30, Revision 5, Aug. 2008.

[5] Joint Planning Development Office, *Concept of Operations for the Next Generation Air Transportation System*. Version, 2007.

[6] SAE, "Certification Considerations for Highly-Integrated or Complex Aircraft Systems, ARP-4754," Warrendale, PA., 1996.

[7] SAE, "Guidelines for Development of Civil Aircraft and Systems, ARP-4754A," Warrendale, PA., 1996.

[8] SAE, "Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment," SAE, Warrandale, PA, ARP-4761, Dec. 1996.

[9] PARC/CAST Flight Deck Automation Working Group (FDAWG), "Operational Use of Flight Path Management Systems: Final Report of the Performance-based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team Flight Deck Automation Working Group," PARC/CAST Flight Deck Automation Working Group, Sep. 2013.

[10] RTCA, "205/EUROCAE WG-71: DO-178C–Software Considerations in Airborne Systems and Equipment Certification," *Draft IP*, vol. 50, 2011.

[11] Federal Aviation Administration, "Advisory Circular: 25.1302: Installed Systems and Equipment for Use by the Flightcrew," FAA, Washington, DC., 2013.

[12] Federal Aviation Administration, "Advisory Circular: 25.1329-1B Approval of Flight Guidance Systems," FAA, Washington, DC., 2006.

[13] Federal Aviation Administration, "Advisory Circular: AC25.1321 Arrangement and Visibility," FAA, Washington, DC., 2010.

[14] R. M. Taylor, "Situational Awareness Rating Technique(SART): The development of a tool for aircrew systems design," *AGARD Situational Aware. Aerosp. Oper. 17 PSEE N 90-28972 23-53*, 1990.

[15] M. A. Vidulich and E. R. Hughes, "Testing a subjective metric of situation awareness," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1991, vol. 35, pp. 1307–1311.

[16] M. R. Endsley, S. J. Selcon, T. D. Hardiman, and D. G. Croft, "A comparative analysis of SAGAT and SART for evaluations of situation awareness," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1998, vol. 42, pp. 82–86.

[17] N. B. Sarter and D. D. Woods, "Situation awareness: A critical but ill-defined phenomenon," *Int. J. Aviat. Psychol.*, vol. 1, no. 1, pp. 45–57, 1991.

[18] M. R. Endsley, "Measurement of situation awareness in dynamic systems," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 37, no. 1, pp. 65–84, 1995.

[19] K. Abbott, S. Slotte, D. Stimson, E. Bollin, S. Hecht, T. Imrich, R. Lalley, G. Lyddane, G. Thiel, and R. Amalberti, "The interfaces between flightcrews and modern flight deck systems," *Wash. DC Fed. Aviat. Adm.*, 1996.

[20] Department of Defense, "MIL-STD-1472G." 2012.

[21] USNRC, "Human-System Interface Design Review Guidelines (NUREG-0700)," Washington, DC., 2002.

[22] R. B. Fuld, “On system validity, quasi-experiments, and safety: a critique of NUREG/CR-6393,” *Int. J. Risk Assess. Manag.*, vol. 7, no. 3, pp. 367–381, 2007.
[23] National Aeronautics and Space Administration, “NASA Space Flight Human System Standard – Vol. 1 (NASA-STD 3001).” 2009.
[24] National Aeronautics and Space Administration, “Human Systems Integration Requirements Revision C (HSIR Revision C).” 2008.
[25] National Aeronautics and Space Administration, “Commercial Human Systems Integration Processes.” 2011.
[26] E. Palmer, “‘Oops, it didn’t arm.’ A case study of two automation surprises,” presented at the Proceedings of the Eighth International Symposium on Aviation Psychology, 1995, pp. 227–232.
[27] E. Palmer, “Murphi busts an altitude: a Murphi analysis of an automation surprise,” presented at the 18th DASC - Digital Avionics Systems Conference, October 24, 1999 - October 29, 1999, 1999, vol. 1, pp. 4.B.3–1 – 4.B.3–6.
[28] S. G. Hart, “Nasa-Task Load Index (NASA-TLX); 20 Years Later,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 50, no. 9, pp. 904–908, Oct. 2006.
[29] A. H. Roscoe and G. A. Ellis, “A Subjective Rating Scale for Assessing Pilot Workload in Flight: A decade of Practical Use,” Mar. 1990.
[30] Aircraft Accident Investigation Commission, “China Airlines Airbus A300B4-622R, Nagoya Airport, Nagoya, Japan, April 26, 1994,” The Ministry of Transport of Japan, Jul. 1996.
[31] G. Carrigan, D. Long, M. L. Cummings, and J. Duffner, “Human Factors Analysis of Predator B Crash,” presented at the Proceedings of AUVSI 2008, Unmanned Systems North America, 2008.
[32] NTSB, “Factual Report - Predator B UA Crash Nogales, AZ, 25th April 2006,” CHI06MA121, Oct. 2007.
[33] P. Ladkin, “American Airlines Flight 965 B757 at Cali, Bogota, December 20, 1995,” Nov. 1996.
[34] NTSB, “American Airlines Flight 965 B757 at Cali, Bogota, December 20, 1995,” DCA96RA020, 2003.
[35] NTSB, “Attempted Takeoff From Wrong Runway; Comair Flight 5191 Bombardier CL-600-2B19, N431CA, Lexington, Kentucky, August 27, 2006,” Jul. 2007.
[36] NTSB, “Northwest Airlines A320 Flight No NW188, N374NW, Overflight of Minneapolis Airport, October 21, 2009,” DCA10IA001, Mar. 2010.
[37] NTSB, “Loss of Control on Approach Colgan Air (operating as Continental Connection) Flight 3407 Bombardier DHC-8-400, N200WQ Clarence Center, New York February 12, 2009,” NTSB/AAR-10/01, Feb. 2010.
[38] Republic of Cameroon, “TECHNICAL INVESTIGATION INTO THE ACCIDENT OF THE B737-800 REGISTRATION 5Y-KYA OPERATED BY KENYA AIRWAYS THAT OCCURRED ON THE 5th OF MAY 2007 IN DOUALA, CAMEROON,” 2010.
[39] BEA, “Final Report - on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris,” BEA, Jul. 2012.
[40] Dutch Safety Board, “Crashed during approach, Boeing 737-800, near Amsterdam Schiphol Airport, 25 February 2009,” May 2010.
[41] BEA, “Accident on 27 November 2008 off the coast of Canet-Plage to the Airbus A320-232 registered D-AXLA operated by XL Airways Germany,” Sep. 2010.
[42] J. Reason, *Human error*. New York: Cambridge University Press, 1990.
[43] Federal Aviation Administration, “Advisory Circular: 25.1309-1B System Design and Analysis,” FAA, Washington, DC., 2002.
[44] J. Rushby, “Using model checking to help discover mode confusions and other automation surprises,” *Reliab. Eng. Syst. Saf.*, vol. 75, pp. 167–177, Feb. 2002.
[45] A. Newell and S. K. Card, “The prospects for psychological science in human-computer interaction,” *Hum.-Comput. Interact.*, vol. 1, pp. 209–242, 1985.
[46] A. R. Pritchett, H. C. Christmann, and M. S. Bigelow, “A simulation engine to predict multi-agent work in complex, dynamic, heterogeneous systems,” presented at the Cognitive Methods in

Situation Awareness and Decision Support (CogSIMA), 2011 IEEE First International Multi-Disciplinary Conference on, 2011, pp. 136–143.
[47] A. R. Pritchett, K. So Young, S. K. Kannan, and K. Feigh, "Simulating situated work," presented at the Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2011 IEEE First International Multi-Disciplinary Conference on, 2011, pp. 66–73.
[48] A. Pritchett, "Simulation to assess safety in complex work environments," in *The Oxford handbook of cognitive engineering*, J. D. Lee and A. Kirlik, Eds. Oxford University Press, 2013.
[49] N. Leveson, *Engineering a Safer World*. MIT Press, 2011.
[50] C. H. Fleming, M. Spencer, J. Thomas, N. Leveson, and C. Wilkinson, "Safety assurance in NextGen and complex transportation systems," *Saf. Sci.*, vol. 55, pp. 173–187, 2013.
[51] Joint Program Development Office, "OPERATIONAL CONCEPT FOR THE NEXT GENERATION AIR TRANSPORTATION SYSTEM (NEXTGEN) Version 3.0," Oct. 2009.
[52] Joint Program Development Office, "Trajectory-Based Operations (TBO) Study Team Report," Dec. 2011.

Appendix A: Accident/Incident Reports

- Aircraft Accident Investigation Commission, "China Airlines Airbus A300B4-622R, Nagoya Airport, Nagoya, Japan, April 26, 1994," The Ministry of Transport of Japan, July 19, 1996. Flight 140
- NTSB, "Factual Report - Predator B UA Crash Nogales, AZ, 25th April 2006," CHI06MA121, October 31, 2007.
- NTSB, "American Airlines Flight 965 B757 at Cali, Bogota, December 20, 1995," DCA96RA020, 2003.
- NTSB, "Northwest Airlines A320 Flight No NW188, N374NW, Overflight of Minneapolis Airport, October 21, 2009," DCA10IA001, March 18, 2010.
- NTSB, "Loss of Control on Approach Colgan Air (operating as Continental Connection) Flight 3407 Bombardier DHC-8-400, N200WQ Clarence Center, New York February 12, 2009," DCA09MA027, February 2, 2010.
- EASA, DCA09RA052, F-GZCP Report, Air France 447, Airbus A330-200, June 1, 2009
- Kenya Minister of Transport Report, Kenya Airways KQA 507 B737-800 Douala; 05-MAY-2007
- NTSB, DCA03IA005, Icelandair 662, 757-200 near Baltimore, MD on 20-OCT-2002

ASRS Reports

- 896575
- 898667
- 932793
- 937132
- 113722

Appendix B: Cost/Benefit Analysis for MCP Testing Example

The following table describes the proposed analytic process to evaluate the benefit of different testing elements, in terms of cognitive fidelity, for an evaluation of an MCP with regard to pilot airplane state awareness. The columns to the right of, and including, High Workload, represent factors known to impact airplane state awareness. The rows include testing elements that project teams could include in the evaluation. Benefit is defined as the number of factors rated as Medium (M) or High (H) impact to the factors, as rated by a Human Factors expert with over 15 years of industry experience. The estimated benefit is considered relative to the Cost Estimate to determine whether the testing element would be Required (R) or Optional (O).

Testing Component	Type	Test Setup Option 1: MCP Evaluation	Cost Estimate	# Med or High Impact	High Workload	Stress	Distraction	Complex Display	Subtle Mode Change	Complex Automation Logic	Test Pilot knowledge/ skill Base	Conformity of flight deck interface
Automation Simulation	Test Software Platform	R	M	6	M	M	M	H	H	H	n/a	n/a
Configurable All Glass Flight Deck Interface	Test Software Platform	R	M	5	M	M	H	H	M	n/a	n/a	n/a
Conformal flight deck user interfaces	Test Software Platform	O	H	6	M	M	H	H	M	n/a	n/a	H
Part-Task Simulation	Test Scenario Element	R	M	5	M	M	M	M	M	n/a	n/a	n/a
Full Mission Simulation (high fidelity in procedures, actors, and roles)	Test Scenario Element		H	5	M	M	H	M	M	n/a	n/a	n/a
Realistic Flight Deck Workload Support	Test Scenario Element	R	M	4	H	H	H	n/a	H	n/a	n/a	n/a
All Glass Flight Deck Interface	Test Hardware Platform				M							
ATC Operator Station	Test Hardware Platform		n/a	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
ATM HIL Simulator	Test Hardware Platform		n/a	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Conformal physical hardware interfaces	Test Hardware Platform		M	1	L	L	n/a	L	L	n/a	n/a	H
Desktop Simulator	Test Hardware Platform		L	0	n/a	n/a	n/a	L	n/a	n/a	n/a	L
High Physical Fidelity (displays, controls, AC dynamics)	Test Hardware Platform		H	2	L	L	L	M	n/a	n/a	n/a	M
Fixed Based Simulator	Test Hardware Platform	R	M	3	M	M	M	n/a	n/a	n/a	n/a	n/a
Motion Platform	Test Hardware Platform		H	1	L	M	L	n/a	n/a	n/a	n/a	n/a
Flight Data	Test Data		H	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Airport conditions	Simulated Test Scenario Element		M	0	L	L	L	n/a	n/a	n/a	n/a	n/a
ATC Simulation	Simulated Test Scenario Element		M	2	L	M	H	n/a	n/a	n/a	n/a	n/a
ATM Simulation	Simulated Test Scenario Element		M	0	L	L	L	n/a	n/a	n/a	n/a	n/a
Traffic Simulator	Simulated Test Scenario Element	O	L	3	M	H	M	n/a	n/a	n/a	n/a	n/a
Voice Communications	Simulated Test Scenario Element	O	H	3	H	H	H	n/a	n/a	n/a	n/a	n/a
Weather Simulator	Simulated Test Scenario Element	O	L	3	M	H	M	n/a	n/a	n/a	n/a	n/a
Connectivity to Remote Simulation Capabilities	Connectiviy	O	M	3	M	M	M	n/a	n/a	n/a	n/a	n/a
Linked Air to Groud via AC Telemetry	Connectiviy		M	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Live Traffic Linked to Ground Station	Connectiviy		M	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Appendix C: Selected Simulation Facilities at FAA and NASA

Location	Lab	Component	Capabilities
FAA	Airport Traffic Control Tower Simulation Platform (ATCT Simulator Platform)		Based on DESIREE simulator infrastructure Can connect to other simulations 9.73 inch HD TV for 270 deg OTW view
FAA	Airport Traffic Control Tower Simulation Platform (ATCT Simulator Platform)		Realistic airport conditions (time), visual conditions, weather, and AC emergencies Configurable tower controller table with their tools (ASDE-X or D-BRITE display)
FAA	Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE)	Rapid engineering of UI and functionality	
FAA	Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE)	Simulation Engine	Most realistic and advanced simulator of en route and terminal ATC systems Replicates functions and user interfaces of Display System Replacement (DSR) and Standard Terminal Automation Replacement System (STARS)

Location	Lab	Component	Capabilities
FAA	Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE)	Simulation infrastructure	
FAA	NextGen Integration and Evaluation Capability (NIEC)	MIT Lincoln Lab Tower Flight Data Manager	Prototype
FAA	NextGen Integration and Evaluation Capability (NIEC)	NextGen Integration and Evaluation Capability (NIEC)	End to end NAS ATM environment consisting of numerous integrated legacy and NextGen ATM capabilities and simulations UAS, Tower, Air Traffic, Surface Mgt System, WXR, TMU, AOC, and Research Cockpit Simulator
FAA	NextGen Integration and Evaluation Capability (NIEC)		
FAA	Technical Operations Human-in-the Loop Simulator		High fidelity, HITL simulator to examine human performance in Operations Control Centers
FAA	William J. Hughes Technical Center	Research Development and HF Laboratory (RDHFL)	Experiment Operator Station (EOS) Can be linked with other EOS

Location	Lab	Component	Capabilities
FAA	William J. Hughes Technical Center	Target Generation Facility (TGF)	Simulation Engine: Dynamic real-time air traffic simulator capability to generate realistic AC trajectories and associated digital radar message for AC in simulated airspace environment Up to 600 targets (400 piloted) can be generated in one or more concurrent simulator environments Multiple terminal, en-route, and Oceanic airspaces can be simulated individually or simultaneously
FAA	William J. Hughes Technical Center	Target Generation Facility (TGF)	Simulated Weather Inject standard day atmospheric model WXR into aircraft dynamics model (ADM) will affect the dynamics and movement of AC throughout the simulated airspace Winds at different altitudes, precipitation
FAA	William J. Hughes Technical Center	Target Generation Facility (TGF)	Primarily for HITL simulator
FAA	William J. Hughes Technical Center	Target Generation Facility (TGF)	Realistic traffic flows and voice communications created in real time by pilots operating the simulated TGF AC in response to ATC instructions
FAA	William J. Hughes Technical Center	Target Generation Facility (TGF)	Support all major Air Traffic Labs of The Tech Center are supported including: E-Route DSR Lab, Stars Terminal Lab, EN-Route Integration and Interoperability Facility (IIF), and the RDHFL)
FAA/ ERAU	Florida NextGen Test Bed	ATC Simulators	Connectivity to NIEC via Aviation SimNet

Location	Lab	Component	Capabilities
Govt, University, Industry	Various	Weather Data	Sharing via Aviation SimNet
JPDO		Test bed	
Mitre	Mitre	ATC Simulators	Connectivity to NIEC via Aviation SimNet
NASA	Aeronautics	UAS Integration in the National Airspace System (NAS) Project, Integrated Test and Evaluation (IT&E) subj project	The IT&E sub-project is building a combined live and virtual (simulated) real-time human-in-the-loop distributed test environment in order to facilitate the evaluation of candidate technologies that will enable more routine UAS operations in the NAS
NASA	Flight Deck Display Research Lab (FDDRL)	ATC center	Simulations of flight deck automation tools w/o adding physical components and personnel
NASA	Flight Deck Display Research Lab (FDDRL)	Pseudo pilot station (Confederate)	
NASA	SimLabs	ATC Simulators	Connectivity to NIEC via Aviation SimNet

Location	Lab	Component	Capabilities
NASA	UAS Integration in the National Airspace System (NAS) Project, Integrated Test and Evaluation (IT&E) subj project	This capability will be used in future tests to reduce technical barriers related to the safety and operational challenges associated with enabling routine UAS access to the NAS. The Project will continue to expand its LVC test capability by extending its interface to facilities at NASA Langley and Glenn Research Centers and possibly the FAA's William J. Hughes Technical Center.	
NASA Ames Research Center	Airspace Operations Laboratory	MACS/ ADRS Simulates Architect: Aeronautical Datalink and Radar Simulator (ADRS)	The ADRS is the central communication process enabling information sharing between MACS stations and other "external" simulation components

Location	Lab	Component	Capabilities
NASA Ames Research Center	Aviation Systems Division	Air Traffic Control (ATC) Simulation laboratories	The ATC Simulation Lab enables NASA researchers to perform complex human-in-the-loop simulations to evaluate the performance of new concepts, procedures, and technologies and determine how well such technologies perform with the addition of humans in the decision-making loop
NASA Ames Research Center	Aviation Systems Division	Air Traffic Management Automation Laboratory (ATMAL)	The Air Traffic Management Automation Laboratory (ATMAL) is a facility designed to support air traffic management research including the development and testing of Center TRACON Automation System (CTAS). At the heart of the ATMAL is a large multi-user computational environment consisting of over 100 UNIX workstations.
NASA Ames Research Center	Aviation Systems Division	Verification and Validation (V&V)	CTAS as a research platform is continuously being improved. NASA and the FAA have installed prototype CTAS tools in several stages at air traffic control facilities serving the Dallas/Fort Worth airports. To support the use of CTAS at these field sites and confirm the functionality of the research software, NASA has developed a software release process to introduce new and improved CTAS functionality. As new CTAS functionality is developed in the Air Traffic Management Automation Laboratory (ATMAL), it is periodically captured and "downloaded" into the V&V Laboratory.

Location	Lab	Component	Capabilities
NASA Ames Research Center	Aviation Systems Division	Virtual Airspace Simulation Technologies PROJECT - Real Time (VAST-RT)	Simulation and modeling for Air Traffic concepts
NASA Ames Research Center	Flight Deck Display Research Lab (FDDRL)	R&D of airside displays and interfaces	
NASA Ames Research Center	SimLabs	Air Traffic Labs (ATM)	Simulate air traffic operations
NASA Ames Research Center	SimLabs	Crew-Vehicle Systems Research Facility (CVSRF)	Realistic interfaces
NASA Ames Research Center	SimLabs	Distributed experiments with UAS	
NASA Ames Research Center	SimLabs	FutureFlight Central (FFC)	Immersive visual environment for ATC/ATM simulations
NASA Ames Research Center	SimLabs	High level Architecture	Connects Simulation components

Location	Lab	Component	Capabilities
NASA Ames Research Center	SimLabs	Linked Air to Ground	SimLabs integrated AC telemetry data from NASA Dryden's Ikhana Unmanned AC System into a Live, virtual, and constructive (LVC) flight test environment
NASA Ames Research Center	SimLabs	Vertical Motion Simulator (VMS)	Unsurpassed motion
NASA Ames Research Center		Remote Cockpit Simulator	Connectivity to NIEC via Aviation SimNet
NASA Dryden Flight Research Center	Gulfstream III: G-III C-20A Research Test bed	Data Collection and Processing System (DCAPS)	Enables processing, distributing, displaying and archiving AC flight data and customers' experimental data in real time
NASA Dryden Flight Research Center	Gulfstream III: G-III C-20A Research Test bed	Embedded instrumentation system	Automated configuration setups to reduce engineering support for each mission
NASA Dryden Flight Research Center		Ikhan Unmanned AC System (UAS)	
NASA Langley Research Center	The Flight Simulation Facilities	Cockpit Motion Facility	Motion and fixed based--4 fixed sites + 1 motion site-- 6 DOF 76 inch synergistic motion system; simulators are moved to different sites with overhead crane

Location	Lab	Component	Capabilities
NASA Langley Research Center	The Flight Simulation Facilities	Development and Test Simulator	fixed-base, advanced all glass transport with programmable sidestick
NASA Langley Research Center	The Flight Simulation Facilities	Differential Maneuvering Simulator	Simulating two fighter or spacecraft maneuvering with respect to each other
NASA Langley Research Center	The Flight Simulation Facilities	Generic Flight Deck	All glass reconfigurable cockpit with programmable side-stick
NASA Langley Research Center	The Flight Simulation Facilities	Integrated Flight Deck Simulator	Full mission simulation capability
NASA Langley Research Center	The Flight Simulation Facilities	Link to other simulation facilities at other NASA Centers, DOD facilities, FAA facilities, commercial facilities, and university facilities.	
NASA Langley Research Center	The Flight Simulation Facilities	Research Flight Deck Simulator	All glass reconfigurable cockpit with programmable side-stick

Location	Lab	Component	Capabilities
NASA Langley Research Center	The Flight Simulation Facilities	Sims with one-of-a-kind oculometer (eye tracking) technology for all classes of AC and spacecraft	
NASA Langley Research Center	The Flight Simulation Facilities	Test and Evaluation Simulator	Reconfigurable to represent any type of vehicle; Orion Capsule or Lunar Lander recently
NASA Langley Research Center		Remote Cockpit Simulator	Connectivity to NIEC via Aviation SimNet
NASA/ FAA	North Texas Research Station (NTX)	Collaborative effort between NASA Ames and several GAA organizations to support NextGen research	Field evaluations
NASA/ FAA	North Texas Research Station (NTX)	existing connection to the FAA WJHTC NextGen External Enclave	
National Weather Service (NWS)	National Centers for Environmental Prediction (NCEP)	Rapid Update Cycle (RUC-2) weather forecast model	Inject weather into FAA Target Generation Facility (TGF)

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01- 05 - 2014	Contractor Report	

4. TITLE AND SUBTITLE

Assessing V&V Processes for Automation with Respect to Vulnerabilities to Loss of Airplane State Awareness

5a. CONTRACT NUMBER
NNL06AA05B

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Whitlow, Stephen; Wilkinson, Chris; Hamblin, Chris

5d. PROJECT NUMBER

5e. TASK NUMBER
NNL12AC67T

5f. WORK UNIT NUMBER
534723.02.02.07.20

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Langley Research Center
Hampton, Virginia 23681

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA/CR-2014-218246

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 06
Availability: NASA CASI (443) 757-5802

13. SUPPLEMENTARY NOTES

Langley Technical Monitor: Steven D. Young

14. ABSTRACT

Automation has contributed substantially to the sustained improvement of aviation safety by minimizing the physical workload of the pilot and increasing operational efficiency. Nevertheless, in complex and highly automated aircraft, automation also has unintended consequences. As systems become more complex and the authority and autonomy (A&A) of the automation increases, human operators become relegated to the role of a system supervisor or administrator, a passive role not conducive to maintaining engagement and airplane state awareness (ASA). The consequence is that flight crews can often come to over rely on the automation, become less engaged in the human-machine interaction, and lose awareness of the automation mode under which the aircraft is operating. Likewise, the complexity of the system and automation modes may lead to poor understanding of the interaction between a mode of automation and a particular system configuration or phase of flight. These and other examples of mode confusion often lead to mismanaging the aircraft's energy state or the aircraft deviating from the intended flight path. This report examines methods for assessing whether, and how, operational constructs properly assign authority and autonomy in a safe and coordinated manner, with particular emphasis on assuring adequate airplane state awareness by the flight crew and air traffic controllers in off-nominal and/or complex situations.

15. SUBJECT TERMS

Aircraft safety; Airplane state awareness; Autonomy; Flight deck systems; Human factors; Pilot-vehicle interface; Situational awareness requirements

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	62	(443) 757-5802