6.7 - Final report of research methods and results

Coordinator	NLR
Work Package Manager	Joris.Field@nlr.nl - NLR
Document Editor	Amy Rankin - LiU

Grant Agreement No.	FP7-314765
Document Identification	Deliverable Number
Status	Final
Version	1.0
Classification	Public

This document is produced as part of the FP7 2012 Aeronautics and Air Transport programme under EC contract ACP2-GA-2012-314765-Man4Gen

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium participants (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Keywords: cognitive system engineering, cognitive science, sensemaking, situation(al) awareness, social situation(al) awareness, resilience, methods, analyses

Contributing partner

Company	Name
LiU	Amy Rankin
LiU	Ellen Ekström
LiU	Victor Sjölin
LiU	Rogier Woltjer
ViU	Irene Stepniczka
ViU	Maximilian Eder

Document Change Log

Version No.	Issue Date	Sections modified	Comments
Draft-v1	01-12-2015		Sent out for internal review
Draft-v2	13-12-2015	All sections	Updated with comments from advisors
Final	16-12-2015		Final revision

Approval status

Prepared by: *(name and role)*	**Company**	**Signature and date**
Amy Rankin	Linköping University	16.12.2015
Irene Stepniczka	Vienna University	16.12.2015
Concurred by: *(name and role)*	**Company**	**Signature and date**
Joris Field	NLR	16.12.2015
Edzard Boland	NLR	16.12.2015
Approved by: *(name and role)*	**Company**	**Signature and date**
Joris Field	NLR	16.12.2015

This page is intentionally left blank

EXECUTIVE SUMMARY

Problem Area

The problem statement in the Man4gen project:

"The pilot's task in modern airliners has transitioned from flying the aircraft by means of manual control inputs, to increased programming of automation and monitoring of the cockpit systems and information during most phases of the flight. Despite the substantial and proven safety benefits of automation systems in 3rd and 4th generation aircraft, evidence indicates that when faced with unexpected and challenging situations, pilots sometimes have difficulties in quickly responding to situations which require a rapid transition in their activity from monitors of very reliable systems, to active and authoritative decision-makers exercising manual control of the aircraft."

Description of Work

The final report on research methods and results brings together the main theoretical concepts, methods and findings made throughout the project from an academic perspective. The document includes the results from a cognitive systems engineering approach (sensemaking) and a cognitive science approach (situation(al) awareness). Methodological considerations and operational recommendations from each perspective are presented. In addition, results from the two research strands are jointly discussed to illustrate their individual strengths and how they complement each other. This report should be viewed as a summary document and as a guide on where to find additional information.

Results & Conclusions

In this report the central theories, methods and main outcomes from the two research strands have been presented, including a discussion on the different methodological approaches. New methods to investigate how crews maintain control following surprise have been introduced within the project, including fMRI, innovative forms of communication analyses, evaluation of social situation(al) awareness studies, sensemaking tags analysis, and the operationalisation of the COCOM and ECOM models to describe and assess the joint crew-aircraft system. The findings from the different perspectives and analyses have been linked to each other and to the industry based assessment, demonstrating how different methods can complement and validate each other. The results from the WP2 and WP5 experiments show that there is a great variation in how the crews understand, prioritise, manage uncertainties and take action following the unexpected events introduced in the scenarios. The findings have generated recommendations for operational use (training, procedures and display design) and raise further questions regarding challenges and possibilities for pilots to maintain control in surprise situations. Suggestions for further research are summarized in the final chapter.

Applicability

The findings presented in this report should be used to get an overview of the research carried out in Man4Gen project and as a reference document to get further information. Further, it contains methodological developments and considerations that can be of use for future studies. Operational recommendations include suggestions for the development of training programs, procedures and display design. Suggestions for future research topics conclude the final chapter.

This page is intentionally left blank

TABLE OF CONTENTS

This page is intentionally left blank

1 INTRODUCTION

1.1. Purpose and scope

This deliverable brings together the research findings throughout the project regarding methods for identifying, analysing and assessing sensemaking (SM) and situation(al) awareness (SA) in operational environments. It includes the outcome of the preparatory studies from WP1, the fMRI trial in WP2, and the experiments in the research simulators at NLR and DLR from WP2 and WP5. In the simulator experiments, pilot behaviour during challenging and surprising situations was investigated. This report thus gives an overview of the main concepts and theories from two different but complementary approaches, the methods applied in each approach and a summary of the main outcomes of the different studies.

1.2. Structure of the document

This document contains seven chapters. Following this introduction, chapter 2 presents the cognitive systems engineering approach (sensemaking), its theoretical background, research questions, the methods applied as well as the associated results. Chapter 3 provides information concerning the cognitive science (situation[al] awareness) perspective, its theoretical basis, related research questions, the methods used as well as the results. Then chapter 4 gives an overview of the combined results from both of the approaches and highlights the differences and common grounds. Finally, chapter 5 summarizes the operational recommendations from both perspectives and presents the future research potential.

1.3. Reference documents

This document is a reference document containing an overview of both of the academic research streams applied within the Man4Gen project. Throughout the document references are made to other deliverables and technical notes in which more information can be found. The main deliverables related to this document which provide more details about each approach are the following (the most important reference documents are highlighted):

- D1.1 Literature Overview Report
- **D1.2 Man4Gen automation problem analysis report**
- **D1.3 Sense-making research method report**
- **D1.4 Situation(al) awareness research method report**
- D2.3 Experiment results – NLR

- D2.4 Experiment results – DLR
- **D2.5 Experiment results – MUW**
- **D3.1 Final analysis of research & operational evaluation**
- D5.2 Results of the procedural evaluation
- D5.4 Results of technology evaluation
- **D6.5 Report on Situation(al) Awareness research**
- **D6.6 Report on Resilience research**

2 COGNITIVE SYSTEMS ENGINEERING (SENSEMAKING)

The goal of the sensemaking research in the Man4Gen project it to investigate how crew's adapt their performance to cope with unexpected events. Research focus is on understanding the sensemaking and control processes taking place. For an elaborated overview of the theories, please see D1.1, D1.3 and D6.6.

2.1. Theoretical Background

Cognitive Systems Engineering (CSE) is a discipline devoted to the understanding of how human-machine systems maintain control in complex environments. CSE sets out to investigate the ways in which people work within the applicable context for the work; the flight crew operating in the flight deck in this case. Studying work practice in an operational setting warrants the main contextual factors to be included in the analysis; factors such as the influence of organizational, as well as cognitive and situational demands (Woods & Hollnagel, 2006). By examining crew behaviour in a simulated operational setting, the identification of interactions between the crew members, as well as with the aircraft and systems are included in the analysis. Within CSE, both people and technical systems are considered as elements collaborating as a JCS, which enables an analysis of how the humans and systems function together. CSE methods analyse the behaviour of this JCS to describe the patterns and characteristics of observable behaviour (Hollnagel & Woods, 2005). In the experiments described here we consider both pilots – Pilot Flying (PF) and Pilot Monitoring (PM) – along with the aircraft automation and systems, as the JCS.

Central for the ability to control a process and successfully adapt is sensemaking. The concept of sensemaking targets both the retrospective and anticipatory aspects of making sense, that is, sensemaking aims to frame both the processes of how we make sense of events after they have occurred and, simultaneously, how we anticipate future events (Klein, Snowden & Pin, 2010; Klein, Wiggins, & Dominguez, 2010; Weick, Sutcliffe, & Obstfeld, 2005). Sensemaking is not only described as an individual process, but can also be described and investigated as a team coordinating their efforts of gathering data and distributing the inferences. Team sensemaking has been defined as "the process by which a team manages and coordinates its efforts to explain the current situation and to anticipate future situation, typically under uncertain or ambiguous conditions" (Klein, Wiggins, & Dominguez, 2010, p 304).

At the core of the CSE perspective is the relationship between joint cognitive systems and their environment, which can be illustrated by both the ECOM and the Contextual Control Model (COCOM) sensemaking and control loop (Hollnagel & Woods, 2005; Rankin, Woltjer, Field & Woods, 2013). COCOM and ECOM models have been applied in the analysis of human-machine systems – in aviation and beyond (e.g., Feigh, 2010; Kontogiannis & Malakis, 2011). Rankin, Woltjer, Field & Woods (2013) described how COCOM can be applied within the crew-aircraft JCS context.

The Contextual Control Model (COCOM) (Hollnagel & Woods, 2005) is a cyclical model showing the relation of human perception and action and is at the core of a CSE perspective. The "sensemaking and control" loop demonstrated in Figure 1 is an adaptation of COCOM to the crew-aircraft context. The "sensemaking and control" loop demonstrates the cyclical process of how the current Understanding of the situation leads to Actions on the Process to be controlled (light blue). Actions together with External events and Disturbances produce Events in the process, and Feedback. Events/feedback modifies the Understanding of the situation, and the loop continues. In this view, it is the context of the situation that determines the actions and therefore the performance of people. The perception of events and feedback for pilots is mostly from the displays and other interfaces in the cockpit, feeling the movement of the airplane, and looking out of the windows.

Figure 1. The cyclical model of human action and perception (adapted from Hollnagel, 2002).

In the Extended Control Model (ECOM) described below (Figure 2), cognition is again described as control (Hollnagel & Woods, 2005). The ECOM comprises four parallel control loops, similar to the cyclical model described above (Figure 1), which makes ECOM a multi-layered model of cognition and human action. An extension of the COCOM cyclical model to the crew-automation JCS handling surprising events, as part of this project, has been presented earlier (Rankin, Woltjer, Field, & Woods, 2013). The ECOM model is the basis for analysis in this work and is therefore described in more detail below. The operationalization of the ECOM has been presented earlier (Field, Rankin, Woltjer, 2015; Field et al., 2015; Woltjer, Field & Rankin, 2015).

2.1.1. Modeling control through ECOM

The ECOM (Hollnagel & Woods, 2005) is a model to describe multiple layers of performance of the joint crew-aircraft system (illustrated in Figure 2). This functional model can be used to examine the distribution of tasks and roles across the different crew members and aircraft systems. Several layers of control loops are applied to describe how anticipatory (feedforward) and reactive (feedback) control are performed simultaneously by the system. As a situation unfolds the distribution of tasks and roles may change and the focus and attention of the crew may shift, demonstrating how the crew-aircraft system adjusts to respond to an event. This includes, for example, how different levels of automation affect the team play between the pilots and automation, how overarching goals provide targets for layers below and how feedback from the lower layers provide input to revision of goals and targets.

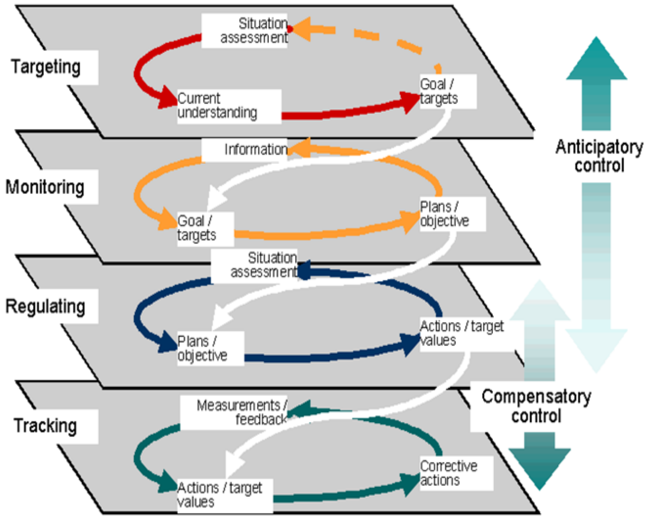

Figure 2: The Extended Control Model, ECOM (Hollnagel & Woods, 2005).

This functional account of the joint system recognizes that a systems performance takes place simultaneously on multiple layers of control (Figure 2). The four layers of interacting control loops in the ECOM thus describe how a JCS set targets (e.g. “heading for destination”), monitors (e.g. “monitor flight path”), regulates (e.g. “reduce speed ahead”) and track performance (e.g. “adjust speed”). The goals set at the targeting layer provides inputs and targets for the monitoring, regulating and tracking layers and reversely the tracking layer provide input to revision of goals and targets. The simultaneous anticipatory (feedforward) and reactive (feedback) control is shaped by the current conditions and system constraints.

2.1.2. Research questions

As part of the Cognitive Systems Engineering (CSE) and Sensemaking (SM) approach the following main research questions have guided the experiment and analysis work:

RQ: How can patterns in how crews search for information, manage uncertainties, prioritize and make trade-offs, assess risks, consider options and anticipate future events be identified?

RQ: How do crews search for information, manage uncertainties, prioritize and make trade-offs, assess risks, consider options and anticipate future events?

RQ: How can patterns in joint cognitive systems‘ control strategies be identified?

RQ: Which joint cognitive systems control strategy patterns can be identified in order to understand and explain performance?

2.2. Studies overview

The main studies that have been carried out as part of the sensemaking research are listed below. After each description reference reports and published papers on the topic is listed.

1. **Interview study with pilots.** As an initial step in the project 23 pilots were interviewed to gain deeper insight into how pilots cope with unexpected events. Using core concepts in sensemaking an analysis was carried out. (D1.2) (Rankin et al., 2013)
2. **Simulator experiment 1.** Flight simulation study where airline flight crews were tasked to handle unexpected situations (for more information on the scenario and experiment set-up see D2.1, D2.2). To describe the degree and kind of control that the crew-aircraft JCS displayed during the simulator session the COCOM and ECOM models have been translated to fit the operational context of the crew-aircraft joint system. Video observations were classified according the developed ECOM model, and assessed using the COCOM model and a sensemaking analysis of the de-brief results was carried out. (D1.3, D3.1 and D6.6) (Field, Rankin, Woltjer, 2015; Field et al., 2015; Woltjer, Field & Rankin, 2015)
3. **Simulator experiment 2.** Flight simulation study where airline flight crews were tasked to handle unexpected situations (For more information on the scenario and experiment set-up see D2.1, D2.2). The behavioural patterns of the crews were classified according the developed ECOM model and a sensemaking analysis of the de-brief results was carried out. (D6.6)

2.3. Methods

In cognitive systems engineering, methods that are used to analyse and describe the behaviour of joint cognitive systems focus on the characteristics of observable behaviour, or performance (Hollnagel & Woods, 2005). Various research settings may be used to enable the analyst to uncover the functions performed by joint cognitive systems and the relevant associated constraints. A fundamental part of the cognitive systems engineering approach for studying work practice is that it is studied in its context. This allows the influence of factors such as cognitive and situational demands, coordination of work processes and the influence of organizational demands to be part of the analysis (David Woods & Hollnagel, 2006).

In D1.1 and D1.3 we offer an overview of methods in CSE. Within the Man4Gen project we have applied an initial field study (interview study) to gain deeper insight into the problem area and two preceding simulated task experiments (two flight simulator studies). Data collection methods included audio-recoded interviews, video observations, and joint- and individual de-brief data. The methods used for the two simulator experiments have not been used at all previously in aviation research and have been developed as part of the project. The methodologies were developed as part of the process in analysing the results of experiment 1. This required several different preparatory steps such as defining the methods framework, transcription and tagging. For experiment 2 the methods where adapted to the context of the new scenario.

2.3.1. Interviews

As an initial step in the project 23 pilots were interviewed to gain deeper insight into the problem area. The experience of the pilots ranges from low-experience First Officers, to experienced Captains, Flight Instructors, Training and Safety Managers. The planned time frame of the interviews was 45-60 minutes, though in many cases the interviewees were willing to have a longer discussion.

The interviewees were asked to relate an example of a situation where they, or their crew member, were surprised or confused. Based on the responses given by the interviewees follow-up questions were asked regarding the following topics: confusion and problem solving, automation and system knowledge, manual operations, training, procedures and communication. No attempt was made to ask each respondent the same questions, or to keep questions in the same order. A method such as this one creates the potential for memory alterations and the results should therefore not be viewed as precise accounts of the incidents, but as a source to investigate challenges and form hypotheses. Further, no attempts were made to draw objective or quantifiable inferences as examples varied in type and depth, but rather to identify the constraints and cognitive demands in each context described. The transcribed data was tagged according to the focus areas noted above and subsequently tagged using theoretical concepts of frames in sensemaking (Rankin, Field, Woltjer, 2013). Further information can be found in D1.2.

2.3.2. De-brief

The debriefings were applied after the simulator experiments and were carried out in two stages: (a) individual debriefings (PF and PM separately), and (b) a joint debriefing (PF and PM together). The individual debriefings (PF and PM) were carried out directly after the flight. This debriefing took 5-10 minutes. In the joint debriefing the crew members together watched a replay of the video recording of the flown scenario. The de-brief facilitator and the crew members could stop the video recording at any time to discuss the crew’s performance. This video debriefing took approximately 30-45 minutes.

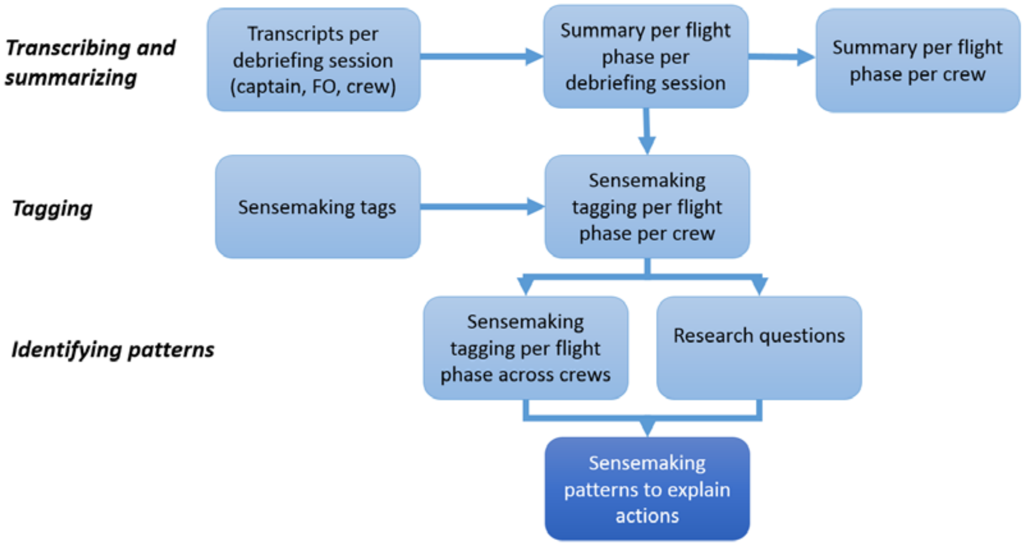

Figure 3. De-brief data analysis process

As illustrated in Figure 2 at first the data from the audio recordings from the three debriefings was transcribed and summarized along the main events and topics of interest in the scenario. In a next step summaries from the two individual (PF and PM) and the joint debriefing per crew were compiled into one summary. This also included a comment if there were any differences in answers between the de-brief sessions. Next the data was tagged per major flight event or flight phase, according to theoretical concepts in CSE/sensemaking theory. After the data was tagged the analysis was structured so that comparisons between crew members and between crews could be carried out. Given the large amount of data acquired research topics of particular interest were chosen as a means to search for patterns. These topics were based on the theoretical concepts of sensemaking.

2.3.3. COCOM/ECOM

For the description of the degree and kind of control that the crew-aircraft JCS displayed during the simulator experiments, the COCOM and ECOM models were translated to an operational context for classification of the observations from the experiment. The ECOM was used to classify the behavioural patterns (see also Field, Rankin, & Woltjer, 2014), and the COCOM was applied to assess the degree of control. The process of operationalizing and applying the COCOM/ECOM is illustrated in Figure 3.

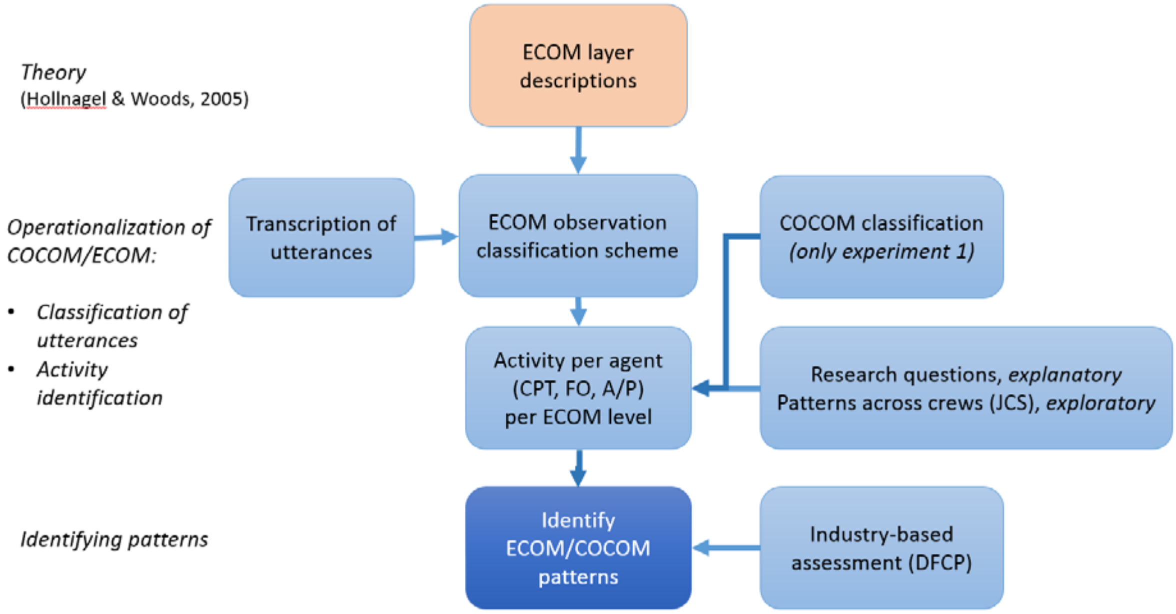

Figure 4. Analysis process for video observations using the COCOM/ECOM

To operationalise the ECOM, the four layers – Goal Setting (originally called Targeting), Planning (originally called Monitoring), Regulating and Tracking - were applied to the experimental data and context of the crew-aircraft JCS, as described in Figure 1. Assigning the observations to the different layers was done through an iterative process of classifying the observations based on the theoretical descriptions of the ECOM model (Hollnagel & Woods, 2005). Dataset from two crews were used to develop the initial classification scheme using three independent raters. This classification scheme was documented and extended, iteratively reaching full inter-rater consensus while being applied to the remaining datasets. The patterns of activities were classified according to the four ECOM layers for a selection of flight phases/segments and crews in each respective scenario. Table 1 shows the high level descriptions of each ECOM layer. The COCOM classification scheme describes the degree of control that the crew-aircraft JCS has in a specific time period of performance. A classification of the control mode (strategic, tactical, opportunistic, scrambled) per flight phase was made using the literature definitions proposed by Hollnagel & Woods (2005), based on three parameters: (i) subjectively available time; (ii) evaluation of outcome, and (iii) selection of action. These were assessed by two of the project researchers that also performed the ECOM classification.

ECOM layer	***Activities***
Goal setting (Targeting)	Setting and resetting high-level goals of what may happen in the future: • Anticipatory situation assessment • Considering need to reset goals and target • Prioritizing between goals • Anticipating risks/consequences of actions
Planning (Monitoring)	Plan monitoring and planning actions, where is the A/C going to be and how it is going to get there: • (Re-) plan and prioritize tasks • (Re-) plan trajectory • Monitoring and information push/pull situation, conditions, systems, for planning purposes (future-oriented) • Identify/decide process phase
Regulating	Carrying out the tasks as part of the plan that has been put together in the monitoring layers: • Allocation of Tracking tasks to operators/systems • Setting values according to plan • Monitoring if values are according to plan (here and now - includes environmental conditions, systems) • Planning of Tracking tasks
Tracking	Actions that have a direct impact on the A/C control surfaces, direct effect on the A/C behavior: • Operate controls • Monitor controls • Monitor sensor values

Table 1. High level description of crew activities

2.4. Results

2.4.1. Introduction

In this chapter main results from each study are presented. More information from each study can be found at: (1) interview study (D.1.2), (2) Experiment 1 (D.3.1, D6.6) and Experiment 2 (D6.6).

There were different experimental scenarios in the two experiments. Both scenarios were developed by operational experts within the consortium including representatives from aircraft manufacturers, operators and training organisations. The main events in the experimental scenarios of the two experiments are listed below.

Experiment 1:

- Increase in wind destabilising the approach (in combination with bad visibility this forced all crews into a go-around)
- Autopilot failure preventing the autopilot from following the commanded heading
- Birdstrike causing a failure of engine 1, and engines 3 and 4 to start surging and stalling

Experiment 2:

- Oil temperature of engine 2 causes an ECAM advisory (blinking oil temperature indication), which leads to the warning 'ENG2 OIL HI TEMP'
- The airplane is hit by a lightning strike, which leads to a disengaged autothrust and the ECAM warning 'THRUST LOCK' and 'ENG 1 THR LEVER FAULT'.
- Bad weather approaching Amsterdam airport

It should be noted that there are differences in what areas were in focus for the ECOM analysis in experiment 1 and 2. Although the overarching RQ are the same from the CSE perspective, once applied to the context of the scenario different patterns will emerge, further guiding the analysis. Given this, the results of the two experiments cannot and are not meant to be compared. The main areas of interest identified for the two experiments these are listed below.

Experiment 1:

- Manual revision
- Engine management
- Trajectory management

Experiment 2:

- ECOM activity distribution
- Long term/short term planning
- Weather checking

2.4.2. Main results

RQ: How do crews search for information, manage uncertainties, prioritize and make trade-offs, assess risks, consider options and anticipate future events?

Interview study

Results from the interview study demonstrate a number of challenges that influence the sensemaking abilities in unexpected situations. Factors that increase the potential for joint system communication and coordination breakdown identified are:

- conflicting, ambiguous or inconsistent data,
- rapid transitions,
- detection of non-events (lack of feedback),
- fixation error and oversimplifications,
- multiple tasks and multiple goals being performed concurrently.

The cases described by the interviewees exemplify how the interplay of many factors, e.g., context, timing, knowledge and expectations shape the cognitive demands that arise in surprise situations. Common to all situations is the relationship between increasing cognitive demands and escalating situations. Coping with a continuously changing environment is, in sensemaking terms, the continuous effort and interplay of retrospective (after-the-fact) and prospective (anticipatory) processes of assessing risks and considering options and future events. The current frame (i.e., current understanding or mental model of the world) of the pilots is continuously being revised as data streams in from the environment and expectations are made, including plans for actions, outcomes of actions and potential hazards or anomalies from the intended plan. An unexpected event will thus put the crews at a disadvantage as the prospective processes for example, mental simulation and preparing responses are in some way compromised. As was demonstrated in the interview cases the challenge of re-framing and identifying appropriate responses may be very different.

The results further show that both automation (e.g. sensor failures) and operational factors (e.g. ATC communication) are common sources of surprise. A surprise or an unexpected situation is however not necessarily a significant threat, although confusion resulting from a surprise may be. It was also mentioned that it is not always necessary to fully understand the problem to cope with the situation successfully. For example, a common strategy mentioned to deal with confusion was “if confused about the automation, take over and fly manually”. Other strategies to cope with surprise and confusion mentioned were to “sit on your hands”, i.e., evaluate the situation before acting and to “stay ahead of the aircraft” to minimise surprise.

The examples and strategies outlined by the interviewees demonstrated potentially conflicting coping mechanisms to deal with unexpected situations in modern airliners today. An example is deciding when to disengage automated systems and take over manual control. As mentioned, manual control should be resumed when confused about what the automation is doing. However, airline operators and manufacturers recommend using automation as much as possible and many pilots also identified that ”automation could make confusing situations safer if you know how to make the automation do what you want it to do”. Further, several respondents mentioned that manual flying takes effort (particularly if not well trained), and that this may degrade other abilities important in difficult situations, such as communication.

Varying views on the required level of system knowledge also highlights the challenge faced in modern airliners today. On the one hand a deeper understanding for systems and their interconnectedness may be useful to deal with surprises, but on the other hand ”pilots sometimes put too much effort into identifying what is wrong with the system instead of flying the aircraft”. Similarly, procedures were seen as one of the safest ways to get out of confusing situations. However, it is important not to follow procedures blindly and sometimes it is necessary to deviate from them, while it is not always obvious when to do so. The interviewees did not think that training today sufficiently provide challenging situations that can help prepare crew for surprise (e.g. situations with no clear procedures or multiple inter-system failures).

Experiment 1

In experiment 1, the analysis of debrief data highlighted a number of issues implying the need for support for crews to search for information, manage uncertainties, prioritise and make trade-offs between goals at various levels of granularity, assess risks, consider options and anticipate future events. Combined with the contextual control strategy analysis we can conclude that some crews do this well by aligning priorities and risk assessments through an interaction between discussion considering options, risks, and expectations, and action acting upon these assessments, followed by monitoring of results of these actions against high-level goals. The extent and content of the assessments made by the individual crews varies considerably, from crews that agree they should have taken more time to assess the best runway once engines were stabilized, to crews who feel there was no other option but to land on the nearest runway. Improving the crew-automation system's ability (e.g. through training, procedures, display design) to improve the understanding of technical issues and assessing their potential consequences therefore seems to be needed. Also, considerable differences were found between the two crew member's assessment of the situation, what variables are considered, what risks, goals, and priorities are identified, and what actions should be prioritised. The overall ability to assess whether immediate action is needed, how much time is available to assess and problem-solve, based on an identification and appreciation of risks and goals, seems to be a subject of considerable variability and therefore deserves attention while this project generates its recommendations. Variations and misunderstandings in the use of autopilot and engine management imply that this ability to appreciate risks and time constraints and possibilities goes hand-in-hand with the technical understanding of the complex systems that are difficult to understand in pressed situations, making these joint crew-automation issues rather than issues of individual pilots.

Experiment 2

The de-briefing in experiment 2 similarly highlights that crews felt uncertain regarding a number of issues, mostly relating to landing options and engine management. Landing options appear largely to be based on which crews collected all the relevant information at an early stage. The crews that land in Brussels did so because of bad weather in Amsterdam. Some crews also mention that Brussels is a large airport with long runways. Of the crews that land in Amsterdam few of the crews consider any other option, and base their decision on the fact that it is their home base. Interestingly, most of the crews that landed in Amsterdam felt surprised by the unfavourable weather upon approaching the airport. They mention that the weather changed faster than they had expected, and they were surprised by the direction of the wind. A majority of the crews felt uncertain about the performance of the engines, due to the risk of engine 2 becoming inoperable. All the crews set engine 2 to idle (although 2 crews later shut it down). Several crews also expressed concerns regarding the controllability and uncertainties regarding the cause of the thrust lever lock. The results in combination with the ECOM analysis show that uncertainties appear to have more impact on crews that also "get stuck" on short-term plans and spend less time gather information and planning ahead. These crews further experienced time pressure towards the end of the flight, while crews with a more elaborate "picture" of the options do not, suggesting that the crews anticipatory abilities are important during the initial re-framing process.

RQ: Which joint cognitive systems control strategy patterns can be identified in order to understand and explain performance?

Experiment 1

In experiment 1 it was found the strategies applied by the crews that performed better with respect to the industry performance assessment of their decisions and actions included interactions between the ECOM layers of Targeting, Planning, Regulating and Tracking. Most activities are triggered as part of procedures or checklists at the planning layer and then subsequently discussed between the crew at the regulating layer; decisions for actions are made, and finally implemented at the regulating and/or tracking layers. If, on the basis of feedback and evaluation, minor adjustments need to be made by the crew to the execution of the plan; this is then done at the regulating layer. If the trajectory needs to be changed to reach the same goal of the flight, these “flight plan” changes are discussed and decided at the planning layer, while higher level goals and prioritization (e.g. runway, alternate airport) is done at the targeting layer. Thus, if there is a regular and frequent interaction between the activities at the various layers of control, performance tends to be of a better quality. For these complex events crews are required to act simultaneously at multiple layers, determining strategies for multiple activities.

Further, the ECOM layers analysis showed that crews with less desirable performance tend to have difficulties in the follow-through and follow-up in the interactions between the layers. For example, if planning decisions and observations are not lifted to the targeting layer when necessary, important considerations regarding choice of runway, and consideration of alternate, and other trade-offs and prioritization of goals may be disregarded. This in turn may lead to lower-layer activities that could be better adjusted to the circumstances if they would be evaluated and reoriented by higher-layer activities, but instead continue to execute plans that are not well-adjusted to circumstances.

Additionally, we have identified that in assessing the situation of the aircraft after an unexpected situation, the flight crew did not all opt for the same procedure – indeed there were often differences between the assessment of the PF and the PM. Once crews identified a procedure to carry out, there was still variation in the way that the procedure was applied and how it affected the outcome of the flight from the perspective of the industry performance ratings of safety. Both the COCOM/ECOM and debriefing analyses’ results support the argument that flight crew had difficulties when facing the unexpected situations in the experiment.

Experiment 2

Similar to experiment 1, in experiment 2, The ECOM layer analysis shows a big variety in crew's performance regarding the time it took to complete the scenario, the number and type of activities and the order in which different steps are carried out. The main differences identified between the crews in experiment 2 are related to the crew's decision making process, and more specifically their re-framing strategies and abilities to make long-term plans within a shorter time frame following surprise. Four main strategy patterns for how the crews re-frame the situation and reach their decision on where to land were identified.

In an initial step to identify the control strategies used by the crews in their decision on where to land the first five minutes after decision to cancel flight and the last five minutes before decision of where to land were analysed. The motivation behind this time frame for the analysis was to capture the performance when crews are in a similar position in their flight and thus comparable (e.g., all crews had just experienced a lightning strike and had to re-frame, or each crew was making the final decision on where to land). It was found that crews that made the decision to land in Amsterdam spend a considerably larger portion of their time on actions than the Brussels crews (Amsterdam = 48.1%, Brussels = 23.5%) and the crews that made a decision to land in Brussels spent a considerably larger portion on of their time on planning (percentages). Further, it was noted that there was a difference in the anticipatory and compensatory planning activities carried out by the crews. In the next phase of the analysis the planning layer (long and short term planning) was thus analysed in more detail (see Table 2 for definitions).

Type of planning	Definition	Example
Short term	Planning concerned with activities that are situated in the present or the near future.	Requesting vectors towards landmark, planning continued descent.
Long term	Planning concerned with activities that are situated in a, given the context, more distant future.	Deciding at which airport to land, requesting weather conditions at an airport.

Table 2: Short term/long term planning

The crews were grouped after their anticipation-related industry assessment score (DFCP) (see D5.2 and D6.7 for more information), resulting in two strategy patterns for the low scoring group and two patterns for the high scoring group (Figure 5). Four different strategy patterns were identifies as part of this analysis (Figure 6). The differences in the patterns were greatest between the crews that landed at

Brussels and the crews that landed in Amsterdam. The two sets of strategies used by the Brussels crews (top two strategies in Figure, crews 101, 106 & 102, 108) show a similar pattern of early on gathering information and making assessment that affect the flight long-term. Some differences can be found between the crews in how the interactions of short and long term play out. The crews that land at Amsterdam (bottom two strategies, crews 103, 104 & 107, 109, 110) show two different patterns; one is the crews that “get stuck” on short-term planning (e.g., managing the engines, descent path and near-by weather). Discussion of landing options is only performed once the crews are close to Amsterdam, at which point the unfavourable weather conditions create an additional surprise. The second patterns of Amsterdam crew strategies include the group that make a rapid decision on limited information (e.g. only information regarding weather in Amsterdam) without consider other options.

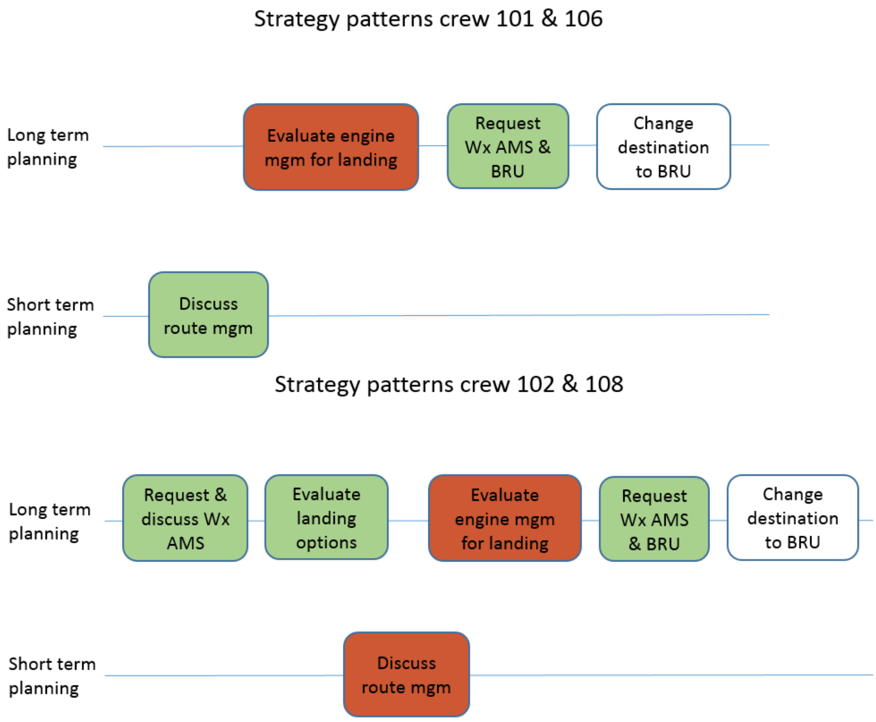

Figure 5: Strategy patterns crew 101, 102, 106 & 108

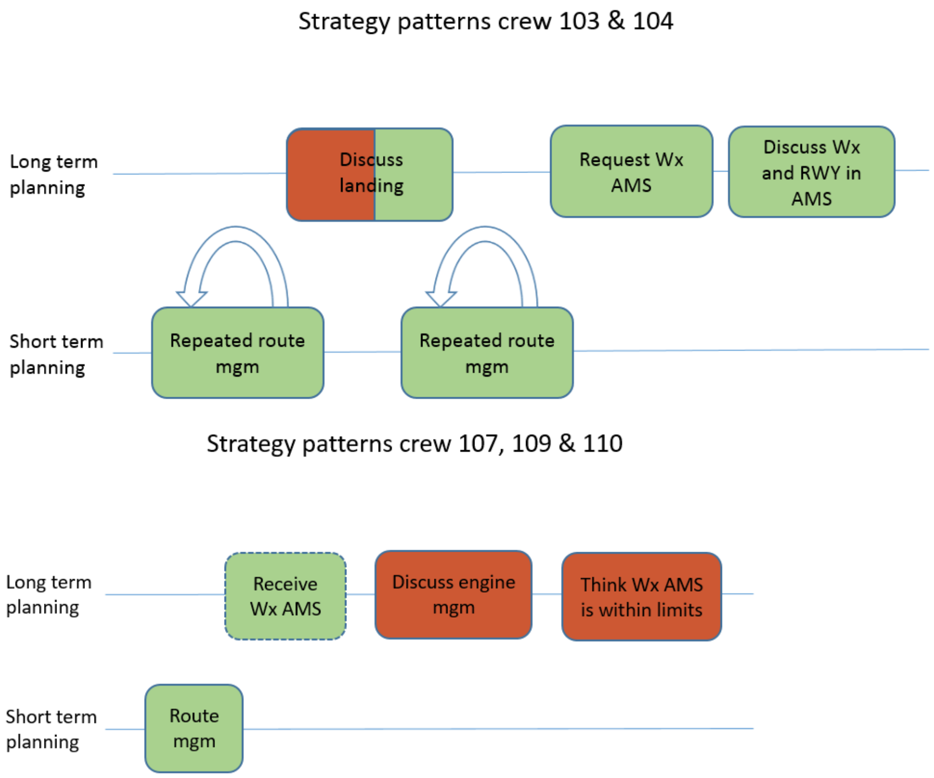

Figure 6: Strategy patterns crew 103, 107, 18, 109 & 110

Inter-group strategy patterns appear to be more similar than cross-group strategy patterns. The strategy patterns of the higher scoring group took considerable less time to gather information, evaluate options and make a decision. Lower scoring crews had a tendency of “getting stuck” with short-term planning and lack information that could potentially generate more options. The ability to quickly re-frame and get back up to the higher layers and see the “bigger picture” at an early stage following unexpected appears to be a determining factor for crews that get a higher score on the anticipatory items of the industry assessment. The results from experiment 1 showed that interactions between the ECOM layers is critical for successful operations, that is, to not get stuck either at higher layers (not taking action) and lower layers (unable to re-form plans and consider options ahead). This pattern has been identified also in experiment 2, where “getting stuck” on short-term planning resulted in less informed decisions and generating additional surprises.

2.5. Methodological and Operational considerations

2.5.1. Methodological considerations

The term situation awareness is commonly referred to and discussed as a major problem in aviation safety today, in both industry and academia. In this project a sensemaking framework has been applied, offering a different set of methodologies to investigate on how crews detect, assess and take action in surprise situations. When examining re-framing challenges, from a sensemaking perspective problems commonly classified as “loss of SA” and “limitations in working memory” (which suggests that the problem is the inattentiveness of the pilot) are described as pilots working hard to create a coherent frame, identify mismatches, elaborate their frame and take appropriate actions in a timely manner. The sensemaking perspective recognizes the importance of context and expectations and how this guides how pilots notice and make inferences about the changing situation does not suggest that pilots lose awareness following unexpected events, rather, it is about attending to what appears to be most important given their current understanding. The sensemaking research agenda focuses on better understanding how we construct, brake and elaborate frames and how this understanding can be mapped to the interplay between humans and automation.

RQ: How can patterns in how crews search for information, manage uncertainties, prioritize and make trade-offs, assess risks, consider options and anticipate future events be identified?

The interview and de-briefing methods used to analysis work and time involved is the standard for any qualitative content analysis, which can be seen as a rather laborious process when transcribing, tagging, and analyzing are considered. However, the interview and debriefing methods used may be considered to be the only tool to obtain rich qualitative data about the thought processes and actions of the crew during an experiment, as reflection during the session would distort the process observed and thus discussions on events during the session necessarily need to be performed directly afterwards. Moreover, since asking about risk assessments, options considered, the reasons for decisions, differences between crew members in these answers, etc., require a considerable depth of discussion, quantitative or questionnaire data collection methods are considered rather limited for this purpose from a CSE/sensemaking standpoint.

The interview and de-briefings data was transcribed tagged per major topic (e.g., questions in interview, major flight phase in experiment) and according to theoretical concepts in CSE/sensemaking theory (e.g, trade-offs & priorities, mismatches between frames, uncertainties). After the data was tagged the analysis was structured so that comparisons between datasets could be made. Given the large amount of data acquired research topics of particular interest were chosen as a means to search for patterns. These topics were based on the theoretical concepts of sensemaking.

Research questions for the interview study were more abstract with the aim of identifying challenges in surprise situations and generating hypotheses for experimental testing. The results generated a broad scope of patterns of challenges (e.g., conflicting, ambiguous or inconsistent data and multiple tasks). As the dataset is more coherent for the experimental setting research questions are more specific. Further, a number of data sources may be used to investigate the questions (de-briefing, video-analysis, and simulator data). However, for identification of patterns on the crews thought processes the de-briefing data offered the most in-depth insights. Patterns were identified regarding, for example, mismatches between pilots, how the crews cope with the uncertainties created in the scenario, which options were considered and their ability to anticipate risks. The findings can be mapped to patterns identified in the video analysis of the crews.

RQ: How can patterns in joint cognitive systems 'control strategies be identified?

The COCOM and ECOM classification schemes were operationalized for the cockpit environment of the NLR B747 experiment. This operationalization should be seen as a result of the study as this is (to our knowledge) the first application of COCOM/ECOM to a cockpit environment. These models, according to the CSE view of data analysis, need an application and anchoring into an operational field of practice in order to be applied to analyse observational data. This is why considerable knowledge of operations and aircraft systems is necessary to produce a useful analysis (as is true of any such experimental analysis that goes to this level of depth). Although this development was laborious and rather difficult in the beginning, the classification scheme soon developed and gained full consensus. After this consensus was established, the observational data was quick to categorize and follows the standard for qualitative content analysis. Again, as with the debriefing data, the analysis work is necessary to obtain an in-depth understanding of crew actions that is anchored in an operational understanding of the sessions as well as an academic understanding of CSE contextual control concepts.

The use of ELAN as an analysis tool in the second experiment was considered useful as it allowed the transcription and categorization of the content into different ECOM layers in the same program. The subsequent coding of the activities of the ECOM into abstract categories with different colours developed during the second experiment were also considered valuable as they allowed for a deeper analysis of the content of each layer, resulting in the strategy pattern analysis. One of the greatest challenges of applying the ECOM for this type of analysis was the great diversity in the performance of the crews. Even though all crews are faced with the same scenario and operate under the same conditions the length, process and final outcome of their flight varied considerably. For this reason the ECOM analysis has been continually developed and defined to guide the analysis in the direction of the research questions. As exemplified in this research, it was found that re-classifying the activities at the planning and goal setting layer as either long term or short term planning helped produce more meaningful results than the original classification.

2.5.2. Operational recommendations

- Improve the understanding of technical issues and assessing their potential consequences

The extent and content of the assessments made by the individual crews in the de-brief session varies considerably, from crews that agree they should have taken more time to assess the best runway once engines were stabilized, to crews who feel there was no other option but to land on the nearest runway. Improving the crew-automation system's ability (e.g. through training, procedures, display design) to improve the understanding of technical issues and assessing their potential consequences therefore seems to be needed.

- Improve crew abilities to detects mismatches, differing views and compare frames

As mentioned above, considerable differences were found between the two crew member's assessment of the situation, what variables are considered, what risks, goals, and priorities are identified, and what actions should be prioritised. The overall ability to assess whether immediate action is needed, how much time is available to assess and problem-solve, based on an identification and appreciation of risks and goals, seems to be a subject of considerable variability and therefore deserves further attention (e.g. through training, procedures, display design).

- Improve abilities to rapidly re-frame and take action following surprise situations

The ability to quickly re-frame and get back up to the higher levels of control and see the "bigger picture" at an early stage following unexpected events appears to be a determining factor in crew performance. The ECOM analysis results show examples of crews both getting stuck in the immediate problems and the inability to take action in a timely manner. The development of crew's abilities and strategies to rapidly re-frame and take action following surprises should be addressed in training programs and procedure development.

- Increase the amount of unexpected events in training scenarios

A variety of crew strategies to cope with the unexpected events have been identified, including more and less successful adaptations to the dynamic circumstances. However, although the analysis suggests that variety of strategies can work well for a situation, different types of strategies are needed depending on the situations and crew. Pilots therefore need be exposed to different types of situations to gain confidence and experience in applying different strategy types, and increase their general abilities to reflect on their actions.

3 COGNITIVE SCIENCE (SITUATION[AL] AWARENESS)

3.1. Introduction & theoretical background

Man4Gen's problem definition was investigated with an interdisciplinary, multilevel and multivariate *cognitive science* approach specifically addressing the concepts and competencies of situation(al) awareness (SA) and social situation(al) awareness (SSA) in a socio-technical system. In this way this research stream addressed basic and applied research, exploring neuronal networks as well as crew behaviour and performance associated with SA (see D1.1, **D1.4**, D2.1, D2.2, D2.3, D2.4, **D2.5**, D2.6, **D3.1**, D4.3, D5.1, D5.2, D5.4, **D6.5**, and the corresponding technical notes).

For more information and details on the approach, its research questions, the methods used, adapted and developed, the approach's results as well as recommendations and conclusions specifically see D1.4, D2.5, D3.1, and **D6.5**.

The working definition for SA that was used within Man4Gen resulted from secondary (literature) and primary (expert interviews) research and is corresponding to an operational definition of SA published by the ICAO (International Civil Aviation Organization, 2013). The Man4Gen's SA working definition namely is:

> ***"SA in an aircraft cockpit includes the recurrent and continuous perceiving, comprehending and projecting of the state of the aircraft and its systems, the aircraft's position and its environment, time and fuel states, possible threats to the safety of the aircraft, the people and their states involved in the operation including passengers as well as developing what-if scenarios for contingencies."***

Besides the competency of single SA we also introduced and explored the concept of social SA. Social aspects such as shared and team SA are important within a socio-technical system, which are also embedded in the definition of SA given above. Pertinent information also comes from the evidence-based-training (EBT) report (International Civil Aviation Organization, 2013) that includes behavioural indicators referring to competencies addressing social aspects such as communication, situation awareness, and leadership and teamwork. Endsley and Jones (2001) define shared SA as the overlapping of single SA due to the interdependency of team members. It is not necessary that all of the information from single SA is shared, but it is mandatory that details are communicated that meet the single SA requirements of other team members. Strongly related to shared SA is the concept of team SA, which relates to the requirements of the information that operators need in order to perform their task, such as what information is there, what information is needed, who has what information and what information has to be shared within the system and the operators, i.e., information management, to perform reliable decisions and actions. Consequently, shared and team SA require existing single agent SA as well as developing a common understanding of the colleagues' SA including their states and needs. In other words, empathy and more specifically, empathic accuracy are also needed. As such empathy – especially its cognitive processes such as perspective taking – represents a basic requirement for social SA and shared mental models (MM). Empathy describes the isomorphic sharing and, ultimately, understanding the different states of another person, but with full awareness that the source of the shared feelings is the other person (Singer & Lamm, 2009; Stepniczka, Tomova, Niedermeier, Peschl, & Lamm, 2015; Tomova, von Dawans, Heinrichs, Silani, & Lamm, 2014). Empathy is a basic cornerstone of successful human interaction and facilitates an understanding of the cognitive and emotional states of others. Empathic accuracy, on the other hand, is the degree of accuracy in being aware of the colleague's mental and emotional states.

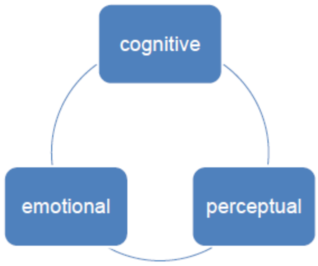

Figure 7: Types of empathy.

As mentioned before, a shared, team or distributed understanding, i.e. social SA, is supported by the pilot competencies of single SA, communication, teamwork and leadership. These competencies are strongly related to the information management processes taking place within a crew-aircraft system and its environment. Information management in aviation is supported by tools and processes to assist the pilots in proactively disseminating information between each other and with the aircraft. The tools and standardized processes for sharing information and crew-aircraft interaction comprise call-outs, checklists, standardized procedures and documents providing guidelines such as the Quick Reference Handbook (QRH) and Flight Crew Operating Manual (FCOM). These different kinds of information sources are shared via verbal and/or non-verbal communication and can be synchronous with input-actions to the aircraft or output/feedback information from the aircraft. Communication can take place between PF and PM, between crew and ATC, between crew and ground staff, and between cockpit and cabin crew. Within the framework of the "staying ahead of the aircraft" concept presented in the Automation Issue Analysis review (see D1.2) communication between pilots is an important determinant for staying ahead of the aircraft and being resilient to possible threats. Communication considerably contributes to the feedback pilots receive, which helps them build understanding and expectations, in other words SA and social SA (see D1.4, TN-1.4.1, TN-1.4.2). Thus, disturbed communication during stressful and challenging situations might represent a serious threat to a pilot´s information management and understanding of a situation (i.e., sensemaking) and consequently, to control within the socio-technical system. Because of its high relevance, especially for social SA, communication is one of the topics within Crew Resource Management (CRM) and also was one of our focuses within the SA research (see also D1.4, TN1.4.1, TN1.4.2, D3.1, TN-3.1.2, TN-6.5.1).

3.1.1. Research questions

The following is an overview of the cognitive science research questions and hypotheses investigated in the Man4Gen project:

Basic Research (fMRI; see D2.5, D3.1, D6.5)

1	**RQ:**	What are the relevant brain networks of SA? (M4G-Mon)
2	**RQ:**	Which brain networks are involved in decision making during ambiguous situations? (M4G-DM)
3	**RQ:**	Are the same brain networks activated during an abstract mental rotation task and an instrument interpretation task? (M4G-MentRot)

Applied Research (behavioural research simulator environment; see D3.1, D6.5, also D2.3, D2.4, D5.2, D5.4, D6.5)

4	**RQ:**	Is there a correlation between task load and situation(al) awareness?
5	**RQ:**	Is there a relation between task load/stress and heart rate during a stress and workload provoking scenario as measured by subjective task load/stress ratings and objective heart rate data?
6	**RQ:**	How stressed and how much in control did the pilots rate themselves during the scenario?
7	**H:**	Social SA with a focus on empathic accuracy in highly trained aircraft crew is decreased for both pilots during stressful events.
8	**RQ:**	How should a dialogue annotation scheme be structured to investigate pilot behaviour in the Man4Gen context?
9	**RQ:**	How are gestures used in crew communication?
10	**RQ:**	How does crew communication differ in various flight phases, when situations change from routine to stressful events?
11	**RQ:**	What kind of communication and behaviour are predictors for a better performance rating (DFCP)?
12	**RQ:**	Is there a difference in communication between baseline crews and crews who applied the Man4Gen procedure concept during different stages of the scenario?
13	**RQ:**	How do higher rated crews differ from lower rated crews when being assessed by tagging SA/SSA relevant categories?
14	SQ:	What are the characteristics of pilots in terms of the types of strategies they apply in their everyday lives?
15	SQ:	What are the every-day empathic characteristics of the pilots?
16	SQ:	What is the chronic stress level of aircraft pilots?

Table 3: List of Man4Gen research questions (RQ), hypothesis (H), and side questions (SQ).

3.1.2. Studies overview

The main studies that have been carried out as part of the SA research are the following:

1. WP1: Literature review and interview study with aviation experts targeting the definition of SA, its measurement and related relevant topics. Primary study on empathy and empathic accuracy (SESA). (D1.4)
2. WP2/3: First set of simulator experiments (WP2; at DLR and NLR). Flight simulator studies in which airline flight crews were asked to handle unexpected situations (for more information on the scenario and experiment set-up see D2.1 and D2.2). SA and related concepts were investigated with established and innovative methods in the field of aviation and the results were linked to the expert ratings (DFCP). (D1.4, D3.1 and D6.5)
3. WP2/3: fMRI experiment (WP2). fMRI paradigms were developed to assess the neuronal correlates associated with SA, decision making, and mental rotation. (D2.5, D3.1)
4. WP5/6: Second set of simulator experiments (WP5; at DLR and NLR). Flight simulator studies in which airline flight crews were asked to handle unexpected situations in order to evaluate the impact of newly-developed procedure and procedure-display concepts (for more information on the scenario and experiment set-up see D5.1, D5,2, D5.4). As in WP2, SA and its related concepts were assessed and the results were linked to the expert ratings (DFCP). (D6.5)

3.2. Methods

Given the results from literature research, the Man4Gen expert interviews we carried out regarding a working definition of SA and its measurement, and considering the Man4Gen problem definition, testing environments, and research questions, we decided upon and combined several methods to investigate pilots' behaviour in a socio-technical system. These included methods that are already well established in the aviation environment (e.g., NASA TLX, SART), methods that are known in principle but have not yet been applied often in an aviation context (e.g., fMRI), and innovative approaches that were adjusted or/and developed specifically suiting the Man4Gen problem definition and experimental requirements (e.g., social SA scale, innovative use of communication data). Research was performed in two different environments, at an fMRI facility (Medical University of Vienna) and at two different research flight simulators (at the NLR and DLR).

We evaluated data from the following measures:

- fMRI[1] to investigate SA, decision making, and mental rotation,
- heart rate to assess pilots' stress

[1] In cooperation with the Medical University of Vienna who was task lead for the fMRI study (cf. D2.5).

- pre- and post-trial questionnaires to investigate pilots' personality traits, workload, and SA,
- expert and de-brief interviews to target at pilots' SA, decisions and actions in the given context of the scenario,
- and performance and behavioural measures through researcher observation. The latter were explored from video recordings, focused on communication and noticeable SA issues, and were linked to the expert observations.

Due to challenges and limitations in the WP2 and WP5 experiments, the testing conditions were not the same for all participants so we only evaluated data from those trials in which data comparability was ensured. A more detailed description of each of the methods and its results is given in D6.5 (and D1.4, D2.5, D3.1), and a brief summary on each method in the chapters below.

3.2.1. fMRI

The versatility of the magnetic resonance imaging (MRI) method allows for detailed in vivo assessments of metabolic processes and (patho-) physiological functions. One of these methods is functional magnetic resonance imaging (fMRI) (cf. D2.5), which can be used to investigate the spatio-temporal properties of neural activation. In essence, fMRI is based on the analysis of time courses derived from a series of MRI images that are weighted for the blood-oxygen-level-dependent contrast (BOLD signal), a measure that reflects blood oxygenation and thus local energy demand in living neural tissue. It has become an apt method to investigate neural mechanisms of small deep-brain structures, such as the amygdala or the orbitofrontal cortex. Thus, fMRI can be considered an indispensible tool for pure basic research but also for basic research in an applied context in cognitive science, neuroscience, psychology, and life sciences, as well as neuroergonomics. Neuroergonomics is characterized by investigating the brain and behaviour under realistic work conditions (Parasuraman & Rizzo, 2008).

Three different realistic aviation paradigms applicable for an fMRI environment were developed and analysed by the Medical University of Vienna in cooperation with the University of Vienna, investigating the neuronal correlates on SA, decision making, and mental rotation. A detailed description of the method and the three paradigms can be found in D2.5 and D6.5.

3.2.2. Heart Rate

Heart rate (HR) is particularly useful for operational assessments because it is relatively inexpensive to measure compared to other psychophysiological measurements (e.g., EEG, fMRI). It is influenced by the interactions of the sympathetic and parasympathetic nervous system and is a very popular physiological variable to monitor the state of a human operator (e.g. Bonner & Wilson, 2002). Heart rate and heart rate variability measures in aviation thus are applied to assess pilots' arousal at a certain point in time or over a certain period of time. For more details concerning the method and its result within Man4Gen see D1.4, D6.5.

3.2.3. Questionnaires

Aircraft pilots are working in a demanding environment and as such certain abilities can support them in managing their daily routines. We were interested if individual personality traits referring to coping strategies, interpersonal reaction and chronic stress have an impact on pilots' performance during a challenging simulator scenario facing ambiguous situations. This is why we implemented the Proactive Coping Inventory (PCI), Personal Reactivity Index (IRI) and a subscale of Chronic Stress (TICS-SSCS) in Limesurvey[2].

The **Proactive Coping Inventory** (PCI) (Greenglass, 1999) is a multidimensional research instrument which collects data concerning cognitive and emotional coping from seven scales, namely: proactive coping, reflective coping, preventive coping, avoidance coping, instrumental support seeking, emotional support seeking, and strategic planning (see D6.5).

The **Interpersonal Reactivity Index** (IRI) (Davis et. al, 1980) is a 28-item self-report measure of empathy. It includes four seven-item subscales, each tapping into a different aspect of empathy. The subscales that were used were the following: (1) perspective- taking, which measures the tendency to spontaneously adopt the point of view of others, (2) empathic concern, which assesses "other-oriented" feelings of sympathy and concern for unfortunate others, (3) personal distress, which assesses "self-oriented" feelings of personal anxiety and unease in tense interpersonal settings. The subscale "fantasy" was not applied because it was not reasonably necessary for the Man4Gen investigations. The IRI has been used across many populations, but not in aviation context yet (see D6.5).

The **Screening Scale of Trier Inventory for Assessment of Chronic Stress** (TICS-SSCS) (Schulz & Schlotz, 1999) is a subscale of the TICS questionnaire and screens for chronic stress including five different aspects of chronic stress, namely work overload, social overload, general overload, and lack of social recognition and worries. The chronicity of stress is then measured by the frequency of stress events perceived retrospectively in the areas mentioned previously (see D6.5).

Additionally, considering the results from literature research and expert interviews and the Man4Gen problem definition and research environments, it was decided to collect data using the following well established questionnaires[3]: Nasa TLXraw and 10D-SART.

NASA TLX raw to assess task-/work-load (Hart, 2006; Hart & Staveland, 1988): The NASA TLX in its raw version was used to collect subjective ratings of perceived task and mental workload in the WP2 simulator, the fMRI, and the WP5 simulator experiments. The NASA TLX is a multi-dimensional rating scale of six

[2] https://www.limesurvey.org/en/
[3] A more detailed description of these methods and their use can be found in D1.4, D2.3, D2.4, D2.5, D3.1, D6.5, and in the technical notes referring to the SA research in WP2 and WP5.

workload-related factors that are combined to derive a sensitive and reliable estimate of workload. The scale that was used had two extreme poles ranging from zero (low) to ten (high) (see D6.5).

Situation Awareness Rating Technique (SART) to assess pilots' SA (Taylor, 1990): The ten dimension SART scale (10D-SART) was used to collect ratings regarding instability of situation, variability of situation, complexity of situation, arousal, spare mental capacity, concentration, division of attention, information quantity, information quality and familiarity in the WP2 and WP5 simulator experiments. The scale of the questionnaire was anchored by word descriptors at each end and ranged from zero (low) to ten (high) (see D6.5).

3.2.4. Social SA visual analogue scale (SSA-VAS)

As mentioned earlier social SA, i.e., shared and crew SA, are important factors for a functioning socio-technical system. In this context empathic accuracy[4], the competency to accurately being able to tune into others in terms of knowledge, thoughts and emotions (Ickes, 1993), is a powerful concept. In line with Man4Gen's problem definition, it was crucial for the pilots to sense their colleague's stress resulting from unexpected and challenging situations as well as their colleague's control level in terms of making adequate decisions and controlling the aircraft in the WP2 and WP5 experiments. However, a measurement instrument for empathic accuracy specifically addressing these issues did not exist in the aviation context. Consequently, a questionnaire (social SA-visual analogue scale [SSA-VAS]) was designed addressing the assessment of the own levels of stress and control as well as the colleague's levels of stress and control. For more details see D6.5.

The questionnaire thus comprises the following four questions and was planned to be handed out to the pilots together with the NASA TLX raw and the 10D-SART questionnaires right after the simulation experiment:

1) How much were you in control of the situation?
2) How much was your colleague in control of the situation?
3) How stressed did you feel in the situation?
4) How stressed did your colleague feel in the situation?

This way the SSA-VAS assesses different concepts at the same time: (1) subjective stress ratings that can be correlated with the NASA TLX, HR data and 10D-SART and thus helps validate the overall results, (2) subjective control ratings (cf. ibid.), and after a formula for each scale is calculated (3) empathic accuracy for stress and control. Consequently, the latter was measured by comparing the control and stress ratings during certain periods of the scenarios that each pilot gave to themselves versus the ratings their colleague gave them, which enabled us to have a direct measure of empathic accuracy of both pilots. This method represents an adapted version of the empathic accuracy paradigm by Ickes (1993).

[4] For a more detailed description of the concept see TN-1.4.2, D3.1, TN-3.1.3.

The equations for empathic accuracy for each pilot, i.e., PF and PM, are the following:

$$EA_C = 10 - |PF_R\text{Control-Other} - PM_R\text{Control-Self}|$$

Equation 1: Calculation of empathic accuracy (EA) for control. Example for calculation of EAc for PF.

$$EA_S = 10 - |PF_R\text{Stress-Other} - PM_R\text{Stress-Self}|$$

Equation 2: Calculation of empathic accuracy (EA) for stress. Example for calculation of EAs for PF.

3.2.5. De-Briefing

Inspired by the de-briefing technique in the operational environment, we decided to perform de-briefings after each experimental session in the WP2 and WP5 experiments. From the SA approach the de-briefings followed three main aims: (1) giving participating pilots the opportunity to share their experience concerning the overall experiment, (2) supporting the research evaluation concerning pilot/crew performance by capturing a set of complementary information to be triangulated with other types of measures including the instructor observation data, the individual questionnaires, and video-observations (e.g. communication), and (3) collecting the opinion of the expert observer right after the simulation to check back whether to focus on specific issues, next to those that were previously defined, during the crew-debriefing interview.

The de-briefing contained four different parts: (1) individual de-briefing including the most important questions, (2) self-rating questionnaires (these included the NASA TLXraw, 10D-SART, and SSA-VAS as described in section 5.2.3), (3) crew-de-briefing with video playback, and (4) a simulation questionnaire. The de-briefing in WP2 additionally included an interview with the expert observer right after the simulation, capturing fresh impressions from the crew performance, which also served as a guideline for the crew de-briefing (see TN-2.1.7, D5.1). In WP5 the de-briefing was additionally expanded by a procedure (developed by NLR) and display design questionnaire and contained questions regarding the usability of the newly developed display (see D6.3, D6.5) and procedure philosophy (see D6.1). For more details see D6.5.

3.2.6. Communication Frameworks

One of the focuses of Man4Gen was on communication, which was researched based on knowledge and methods from previous communications investigations, which were adapted to the Man4Gen environment. Doing so included (1) researching existing dialogue schemes and their applicability for the Man4Gen problem definition and adapting an existent dialogue structure coding scheme to the research environment (aircraft cockpit and scenario) in the WP2 and WP5 experiments, (2) evaluating the communication data by itself and in relation to the data generated by other measures (e.g., pilots' self-ratings, expert performance ratings [DFCP]) (see D3.1, D5.2, D5.4, D6.5), (3) and developing recommendations based on the results. For more details see D6.5.

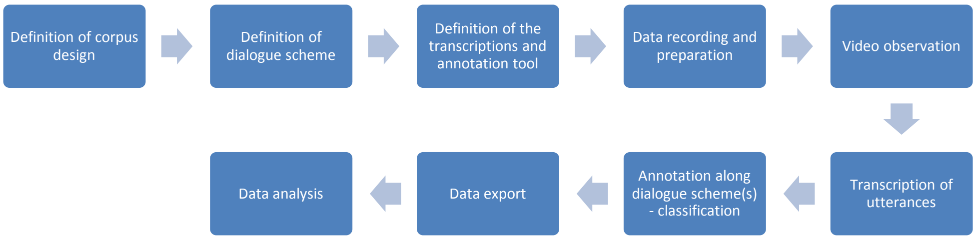

Figure 8: Process of communication data analysis.

The Man4Gen corpus is a large multimodal corpus, comprised of a collection of task-oriented crew interactions (18 in WP2 and 14 in WP5). The communication in both settings, WP2 and WP5, was elicited by scenarios matching the Man4Gen problem definition criteria. These contained challenging situations that developed over time and contained ambiguous moments that required adequate decisions to perform a safe flight simulation (see D2.2, D5.1). The focus of the analysis was on the interaction between the flight crew. Consequently, the corpus setting was contextually specific and thus useful to study interpersonal aspects of crew communication under specific conditions.[5]

The basis of the Man4Gen annotation scheme was a map task corpus including diverse phenomena which was tested and found to be very suitable and reliable when used to research dialogue structure (Carletta et al., 1997). The Man4Gen annotation scheme expands this corpus with action moves relevant for investigations of crews in challenging scenarios, allowing the communication data to be embedded in SA research and contrasted with the expert ratings on crew performance (cf. D3.1, D6.5).

Using the verbal dialogue annotation scheme the communication of the crews was structured into three different levels: The basic level, level one of the annotation scheme, classified an utterance into four possible basic dialogue structure forms, namely initiation, response, preparation, and incomplete communication. These basic tags go back to the simplest form of an information processing system where a stimulus, i.e., initiation, complete or incomplete, is processed and a response is given. The level-one tags were further broken down into more detail in level two, classifying initiation utterances into command, statement, or question and response utterances into communication, information, or question. The tags preparation and incomplete communication were not further subdivided. However, the most important division of communication classifications are the action moves. These are level three in the dialogue annotation

[5] Information concerning data recording can be found in D2.3, D2.4, D3.1, D5.2, D5.4.

scheme and allow for the most substantial analysis. Taking into account the cockpit environment and crew tasks within the scenario, initiating moves, response moves and a preparation move were defined. For the WP2 experiments, the various move classifications were split into acknowledge, align, check, clarify, excuse, explain, explain-check, info, instruct, instruct-check, ready, reject, and request actions, listed in alphabetical order (see Figure 10).

The WP5 annotation scheme was a further refinement of the WP2 dialogue framework and included the following classification moves to address the procedure and display-procedure concept (listed in alphabetical order): acknowledge, acknowledge-check*, align, check, clarify, command, complex, excuse, explain, explain-check, instruct, instruct-check, mental model sharing*, non-specific*, question-check*, reject, request, ready, and share*[6] (see Figure 11).

Concerning non-verbal communication our focus was on functional gesture coding, which is “module III” of the original coding structure (cf. NEUROGES by Lausberg & Sloetjes, 2009). However, we adapted the framework to the non-verbal gestures that are relevant in a cockpit environment, i.e., excluding gesture types that were not relevant while adding gesture-types that were not included in the original system but were of interest to us. Consequently, the non-verbal annotation scheme we used included the following moves (see Figure 9):

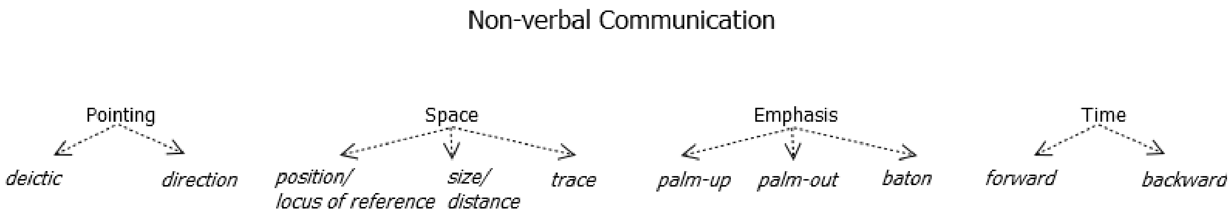

Figure 9: Non-verbal communication annotation scheme for transcriptions.

[6] *marks an annotation action that specifically was used in WP5 addressing the new procedure, and display-procedure concepts

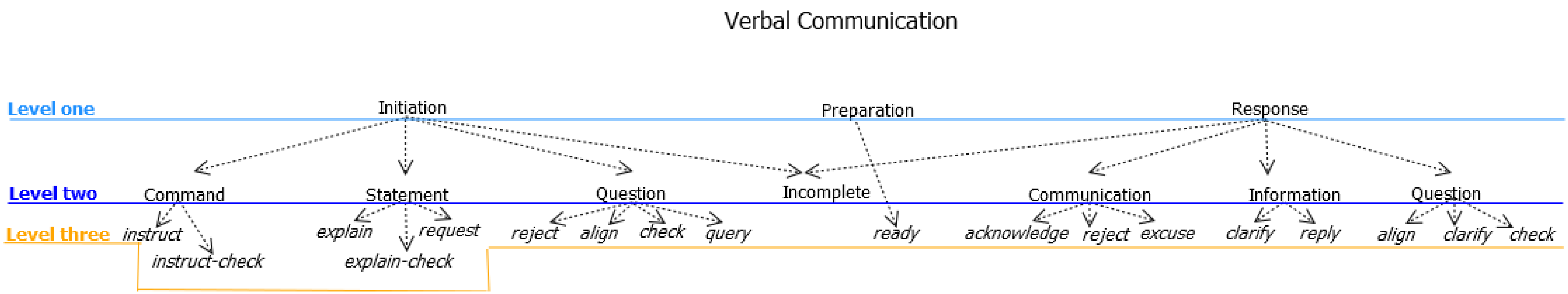

Figure 10: Verbal communication annotation scheme for transcriptions of Man4Gen WP2 utterances.

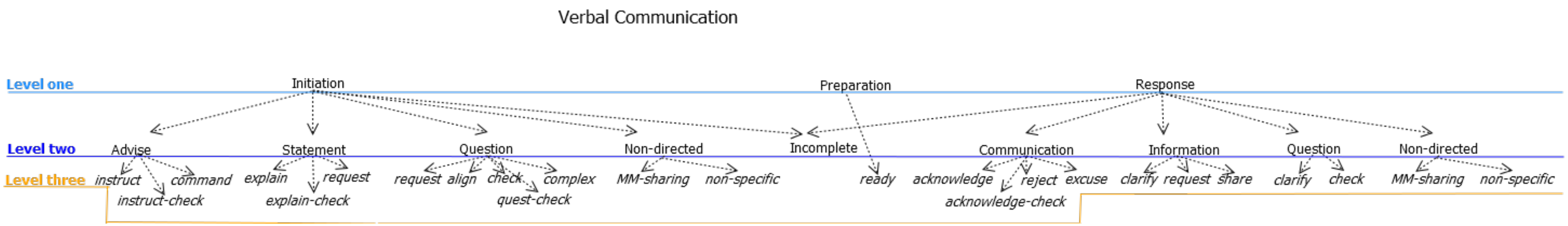

Figure 11: Verbal communication annotation scheme for transcriptions of Man4Gen WP5 utterances.

3.2.7. SA Annotation Framework

Next to the communication dialogue schemes an annotation scheme was developed and applied to the communication data for a selection of WP5 crews to capture relevant aspects related to SA and social SA. For more details see D6.5.

An annotation scheme for SA relevant concepts was developed based on conclusions from Man4Gen investigations into the formation, maintenance, and implications of SA. Verbal and nonverbal activities related to SA in the opinion of the raters were categorized into five main categories: perception, process of understanding, understanding, anticipation and decision-making (see Figure 12).[7] The last two categories were subdivided further. Anticipation was split into one category relating to the descriptive proposition of future states and another category relating to a sensemaking proposition. The decision category was subdivided into four types of decision-making: creative, choice making, rule-based, and intuitive. Furthermore, the first four categories (i.e., all except decision making) were aggregated into an overall SA score (Sum SA).

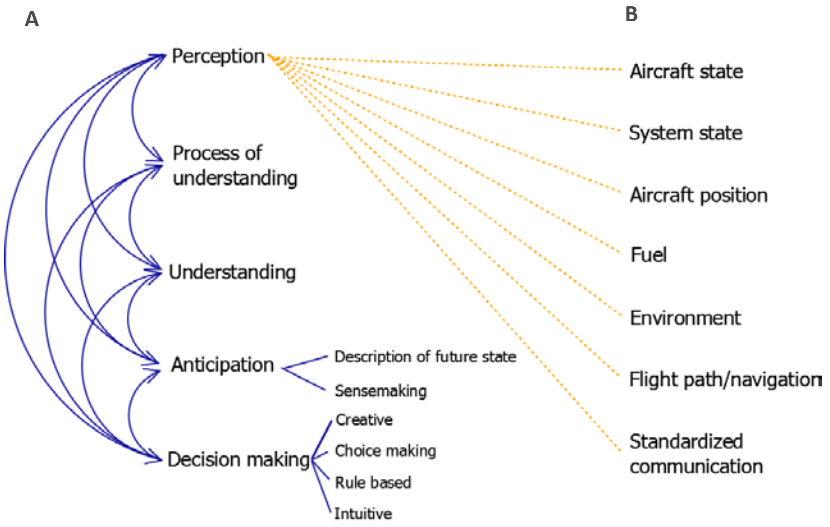

Figure 22: Annotation scheme including SA (A) and content categories (B). The dotted orange lines indicate that all items of the SA group can be related to the items of the content group. Double-headed arrows indicate the interdependency between SA category items.

[7] The concept of „action“ (manual input to and output from the system) is not included into the framework for Man4Gen purposes, as actions can be tracked more accurately, adequately and economically by SIM log data than by annotation from video analysis. However, action data from SIM log data could be added smoothly to the analysis.

Overall, four crews were evaluated in the exploratory analyses: the two crews with the highest DFCP ratings (crews 106 and 202) and the two crews with the lowest DFCP ratings (crews 204 and 109) (see D6.5 for more details).

3.3. Main Results

An excerpt of the results can be found below. For more details please see **D6.5.**

RQ: What are the relevant brain networks of SA? (fMRI)

Although the Man4Gen fMRI tasks were very complex, the paradigms delivered highly consistent group results and the following brain areas can be seen as central hubs within a complex network that is mandatory for having SA: For monitoring, i.e., perception and understanding -related activation was found in brain regions responsible for vision (visual cortex, bilateral geniculate, ventral and dorsal visual stream), attention (thalamus), somatosensory cortex (S1, frontal insula), and motor areas (M1, [pre-] SMA). The latter might be associated with preparation of the imminent button press that always succeeded a monitoring period (e.g. match/mismatch and okay/intervention/manual). Naturally, motor planning, execution, control, and suppression can also be expected to be fundamental processes during real -world flight operations. Furthermore, task activation in the somatosensory cortex, which is central in perception of the body, may be elicited by actual perceptions and mental imagery. It might be plausible that pilots also use their body sensory system to envision and simulate on-going aircraft behaviour, even when lying inside an MR scanner and being physically disconnected from the environment of a freely moving aircraft. However, further studies might be needed to explore the influence of these embodied cognitive processes that support monitoring and subsequent decision making. Additionally, a task network in the prefrontal cortex (PFC) was found. More specifically, activation was found in the dorsolateral PFC, typically associated with working memory performance and executive functions. Furthermore, stronger blood circulation was found in the lateral orbitofrontal cortices, which are central for complex decision making and information integration. Additionally, the task activated Broca's area, which is responsible for language synthesis and syntactic thinking. Anticipation, however, activated brain areas different to monitoring. An increased activation in the ventral striatum and putamen was found, which suggests that anticipation in the context of this study is an active process, in which ventral striatum and putamen play an essential role in the prediction of future events. Activation in these regions has been associated with the estimation of present and future reward, and is relevant for context dependent future action planning. Importantly, these brain areas are typically associated with planning and preparation of future actions and decision.

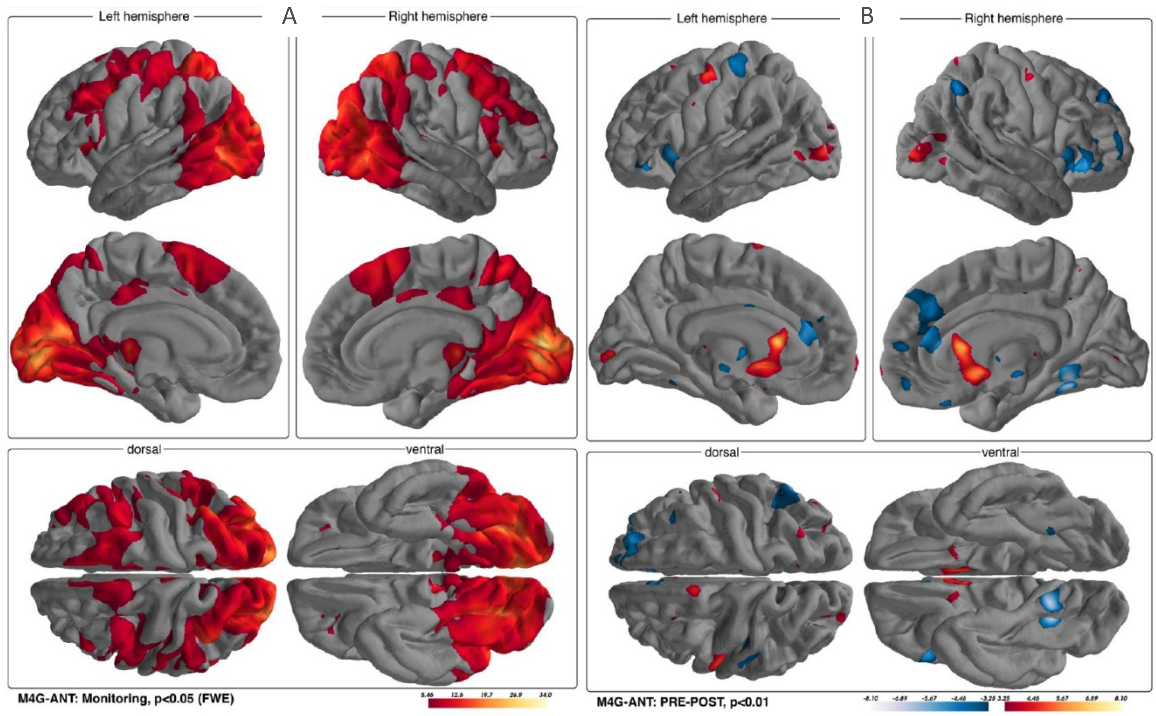

Figure 33: A) Brain activation during Monitoring (PRE and POST). For the monitoring contrast (in PRE and POST), the significance threshold was set to $p<0.05$ (FWE corrected); B) Comparison of positive (red) and negative effects (blue) during Anticipation, the significance threshold was set to $p<0.01$. For more details see D2.5.

RQ: Which brain networks are involved in decision making during ambiguous situations? (fMRI)

Besides investigating the concept of SA we also had a closer look at decision making which is interlinked with SA (cf. D2.5). The effects of interest were on the one hand monitoring with anticipation and on the other hand decision making with four possible conditions, i.e., (1) OK, when participants indicated that the situation does not require any form of intervention, (2) intervention, when participants reported that an intervention within the autopilot guided flight pattern was required, (3) manual, when participants reported that a situation was imminent that would require manual operation of the aircraft, and (4) missed, when participants did not respond on time. The results indicated that an increase in an operational demand such as during decisions requiring actions in terms of intervention and manual control, lead to an increase in activation of prefrontal regions (e.g., VLPFC, DLPFC, Broca's area). Additionally, we found a deactivation in missed trials in the left ventrolateral prefrontal cortex, Broca's area, and anterior cingulate cortex. This suggests that these areas, which are central components of the decision making network observed herein, were not in a receptive state (Sadaghiani and Kleinschmidt, 2013) or did not receive appropriate neuronal input from other brain regions.

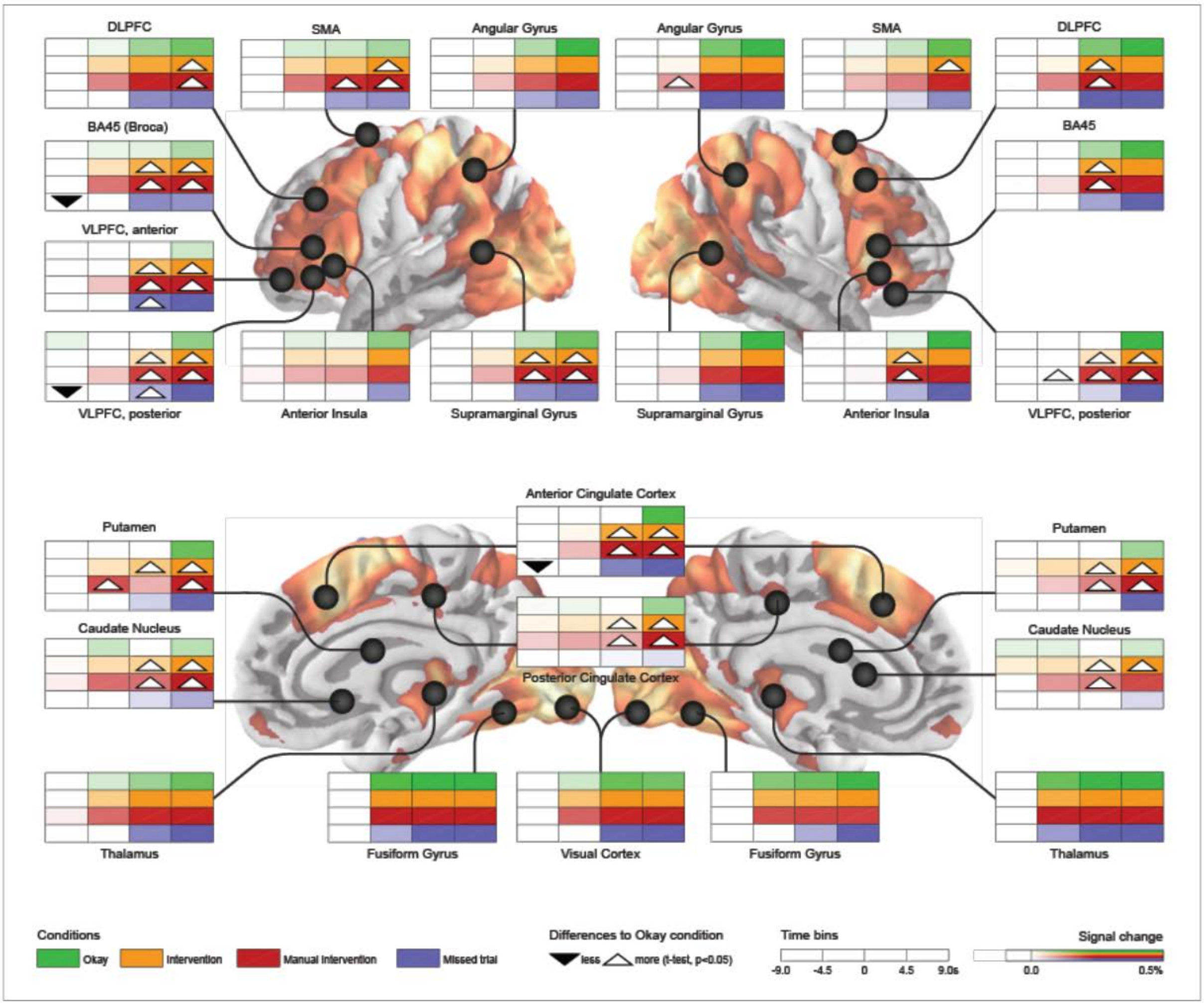

Figure 44: Statistical parametric map of decision making conditions and summary of regions of interest (ROI) analyses. Local activation maxima of whole-brain analysis (p<0.05 FWE, shown in background) were used for hypothesis-free data-driven post-hoc ROI analysis (group maximum center, r=8mm). Time courses were compared over four time bins (button press set to t=0) using two-sided t-tests (OKAY vs. MANUAL, INTERVENTION or MISSED; p<0.05 Bonferroni-corrected). For more details see D2.5.

Are the same brain networks activated during an abstract mental rotation task and an instrument interpretation task? (fMRI)

Brain activation, related to both task conditions, was found in visual areas (occipital lobe, fusiform gyrus), superior parietal lobe, dorsolateral PFC, thalamus, putamen, somatosensory cortex, and motor cortex. Despite this substantial overlap, the contrast between the two conditions revealed task-specific differences. Mental rotation elicited significantly stronger activation in bilateral pre- and supplementary motor areas, and superior and inferior parietal lobe. A substantial overlap in brain activation related to instrument interpretation and mental rotation was found which implies that similar brain networks are required for abstract mental rotation paradigms and an airplane attitude judgement task. While there is a substantial overlap of activation across the task conditions, we found that there are also significant activation differences between instrument interpretation and non-aviation based mental rotation. Taken together, our results indicate that abstract mental rotation might not be a reliable predictor for actual real-life spatio-motor skills.

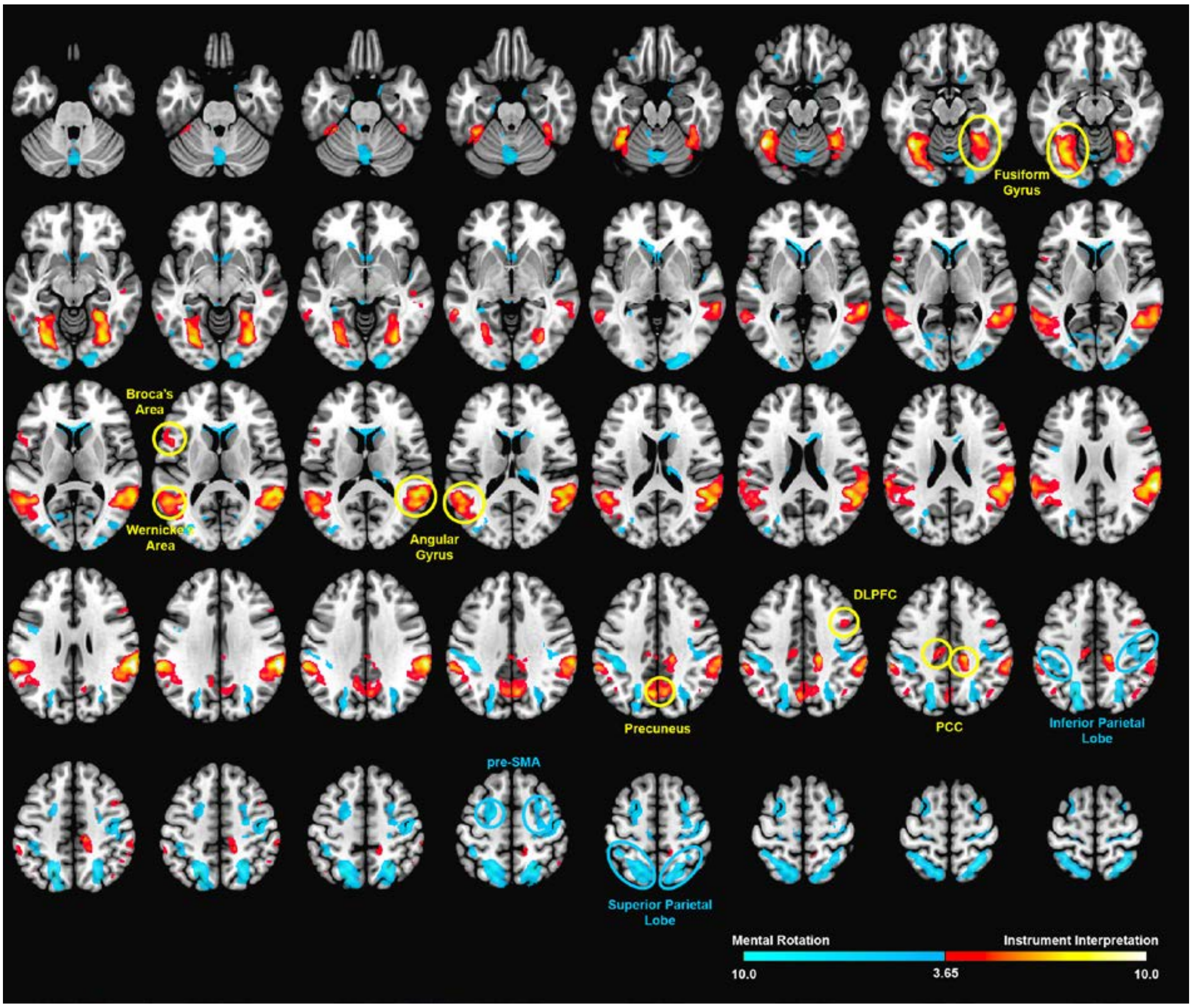

Figure 55: Statistical differences of Shepard-Metzler task (cool) and instrument interpretation task (hot). Statistical threshold for t-statistics was set to $p<0.05$ FWE cluster-level corrected. Increased activation during instrument interpretation was found in the fusiform gyri (first row), Broca's and Wernicke's area, angular gyri (third row), precuneus, DLPFC, and PCC (fourth row). The Shepard-Metzler task induced more activation in the inferior (fourth row) and superior parietal lobe and the pre-SMA (last row).

Given the present studies, SA and its intertwined concepts decision making and mental rotation, the latter especially important when controlling an aircraft manually, seem to be embodied in distributed and complex brain circuitry. SA therefore is task-oriented and highly context dependent and utilizes a diverse set of functional brain networks. Overall the fMRI method gave insights into SA investigations from a basic research perspective and allowed to develop a conceptual neuronal model of SA (see D6.5). Additionally, fMRI allows for task designs that work as experimental proxies for real world aviation skills, and as such is a useful tool for neuroergonomics.

RQ: Is there a correlation between task load and situation(al) awareness? (NASA TLXraw, 10D-SART, SSA)

Man4Gen data indicates that there is a negative correlation between SA and workload, meaning that if pilots reported their task load as being higher their level of SA was lower. The data signals that workload, stress and SA are different concepts that require topic specific measurement tools. However, the results from our analyses strengthen existing conclusions that the concepts of workload, stress and SA are related and have an impact on each other (e.g., Berggren, Prytz, Johansson, & Nahlinder, 2011; Szalma & Teo, 2012; Vidulich, 2000). With increasing workload and stress, pilots' SA can be diminished. Substantial efforts in the fields of display/system design, procedure and training philosophies have already been made and have positively influenced crew performance. However, since the Man4Gen results in WP2 and in WP5 showed that stress and workload are still main factors impacting gaining and maintaining SA, it was concluded that crews could benefit if their workload and stress levels could be lowered, if information dispersal were minimized and linked to data that is currently being processed (short-term), and if the overall picture of a situation were presented over time (long-term). Therefore, it was suggested that the Man4Gen display/system, and procedure philosophy concept should support pilots in their time and information management with the aim of lowering crews' feeling of stress and workload.

RQ: Is there a relation between task load/stress and heart rate during a stress and workload provoking scenario as measured by subjective task load/stress ratings and objective heart rate data? (NASA TLXraw, 10D-SART, SSA, HR)

The analysis of the heart rate (HR) data in WP2 showed that there was a positive correlation between mean HR (1/minute) and subjective stress ratings. An increase in stress went along with a rise in mean HR. In line with these results there was a negative correlation between mean HR and the standard deviation of the NN (normal RR) interval (SDNN), as well as root mean square of successive differences (rMSSD) data. A lower rMSSD indicates a decline of parasympathetic activity and thus lower heart rate variability (HRV). However, Man4Gen's exploratory sample did not show an association between HR data and subjective reported workload (measured using the NASA TLXraw) or subjective reported SA (measured using the 10D-SART). Although stress and workload are closely related to each other, our findings indicate that these two are different concepts that require different types of assessment[8].

[8] Questionnaire results in WP5 additionally indicated that there is a difference in long-term, chronic stress and spontaneous short-term stressors (cf. TN-6.5.1).

Real-time HR measures can be used to assess how effective aviation training is/was by showing variations in the stress levels during a scenario. Additionally, real-time HR measures could also be used to personalize the degree of difficulty, hence the level of stress, during a training session by adding or deleting events as appropriate.

RQ: How stressed and how much in control did the pilots rate themselves during the scenario? (SSA)

H: Social SA with a focus on empathic accuracy in highly trained aircraft crew is decreased for both pilots during stressful events. (SSA)

The results implied that crews were uncertain regarding their colleague's stress level during stressful phases in the scenario (WP2). Under less stressful conditions, pilots were better able to accurately assess the emotional state their colleague was in. When stress levels were higher, pilots' empathic accuracy decreased and they were less able to judge how stressful their colleagues perceived the situation to be. This is likely exacerbated by the fact that flight crews do not usually work in fixed pairings, decreasing the time individual pilots fly with one another, and the fact that surprising and challenging situations are rare due to good trainings and reliable systems.

During the most stressful flight phase according to the self-ratings, pilots had the worst empathic accuracy regarding their colleagues' perceived level of control. When pilots cannot accurately judge how much control their colleagues feel, competencies such as leadership, communication, decision making, monitoring, and SA can be negatively impacted. As a result, pilots may be less able to sufficiently guide their colleagues through challenging situations, carry out effective communication, and make adequate decisions (for more details see D6.5).

Taken together, in the present SSA-VAS analysis, we were able to confirm the hypothesis that stress, i.e., challenging and surprising situations during flight, leads to altered social cognition. We found that empathic accuracy between pilots was decreased during stressful situations in a challenging ambiguous scenario. This emphasises the conclusion that new concepts of display philosophy, procedures, and training should support crews in a way to lower their stress levels during challenging situations.

RQ: How should a dialogue annotation scheme be structured to investigate pilot behaviour in the Man4Gen context? (Communication annotation schemes; for the results see page 38; for more details see D6.5)

RQ: How are gestures used in crew communication? (Communication annotation frameworks)

Verbal communication clearly dominated in the cockpit environment; in the context of the Man4Gen problem definition, non-verbal communication, i.e., gestures, mainly supported and reinforced the verbal communication. Additionally, non-verbal communication functioned as a call for attention but it only rarely substituted verbal communication (for more details see D6.5).

RQ: How does crew communication differ in various flight phases, when situations change from routine to stressful events? (Video observation, communication annotation schemes)

In WP2 we found significant difference in the number of different basic communication items. Looking at the pairwise comparisons, the number of initiations and responses significantly differed from each other and from the other communication items, except for the number of preparation and incomplete communications. Initiation was the most frequently used communication item. Additionally, we identified that communication was different for PF and PM during the various flight phases. PF's initiation of communication significantly increased during events that were characterized as more stressful (e.g. go-around) than initial flight phases (e.g. initial approach). Furthermore, PF generally initiated communication more often than responding. During most stressful flight phases (e.g. bird strike), the PM changed his behaviour from being mainly reactive to predominantly initiating communication. These results support on-going discussions in the aviation community about the importance and role of the PM (Civil Aviation Authority, 2013).

Next to the numerical model approaches we also performed a qualitative analysis to complement the statistical analyses. The following behaviour was observed: within the group of crews belonging to the higher rated 50%, the PM was initiative but also listened and followed the instructions of the PF. In addition, PM shared their mental model approximately twice as often as in the group of crews belonging to the lower rated 50%. Moreover, PM often explained their actions verbally and verbalized their thoughts during problem solving. Such behaviour, i.e., verbalization, helps pilots improve their performance by offloading cognition and by enforcing joint performance power through interactive and collaborative cognition. This finding also was supported by the fMRI studies carried out in WP2. Finally, the PM belonging to the higher rated 50% also prioritized their actions.

Higher rated 50%	Lower rated 50%
Proactive behaviour	Reactive behaviour
Self-confident and clear language	Uncertainty in language (e.g., maybe)
Active recognition, sharing and advise on abnormal engine conditions	"passive" recognition and sharing on abnormal engine conditions
Accompanying actions with explanations	

Table 3: Types of behaviour identified in the crews belonging to the higher and lower rated 50% of the WP2 sample.

RQ: What kind of communication and behaviour are predictors for a better performance rating (DFCP)?

(Communication schemes, video observation, DFCP)

Automatic linear modeling in terms of predictive analytics was used to build a predictive model for crew performance in WP2. The results highlight that the role of the PM within the crew is very important and the PM's actions had the greatest predictive effect regarding whether crews would be rated higher or lower. Consequently, if the PM did not need to request information during stressful phases but instead was able to retrieve information autonomously, i.e., from correct application of procedures or checklists or from memory, crews tended to be rated higher; i.e., performed better. Furthermore, higher rated crews had a very active PM during stressful situations who explained a great deal and shared his/her mental model of the situation. A PF behaviour that emerged as being relevant for better crew performance ratings was an increased amount of explaining actions during go-around and during the final approach and landing. This implies that sharing one’s mental model has a positive effect on shared and team SA. Additionally, better crew performance was linked to the PF appropriately selecting the type and amount of communication. The PM in crews which received higher rating scores seemed to be better able and more ready to receive information. A lack in these abilities, i.e., the ability and amount of sharing mental models and verbalizing current and planned actions, had an impact on the different gaps in SA observed in our WP2 scenario. Also, if the PF used initiative and gave directions when required, then the crew received better ratings. All together the results went along with the operational evaluation of the WP2 crews, which was performed by the Man4Gen industry partners. Our colleagues used an aviation expert performance indicator rating system (DFCP) and also found that communication and leadership competencies were crucial for successful WP2 scenario performance. The results from both types of analysis consequently validated each other.

Various analyses were also conducted with the communication data in WP5 to (1) test for predictors of better DFCP ratings and (2) to look for possible differences between procedure crews (crews using the newly developed procedure by the NLR) and baseline crews, as well as between the first half (pre-phase; from oil temperature advisory until lightning strike) and the second half of the experimental session (post-phase; lightning strike until simulation end) (see D6.5).

In WP5, as in the previous analysis, linear modeling was used to identify predictors for the DFCP ratings from all crews (including baseline, procedure, and procedure-display crews). The results from WP5 highlight that crews received better ratings if the pilots were clear about the distribution of roles. Crews also received better ratings if there were few misunderstandings from the PM perspective and if the PM regularly validated the context of the interaction with the pilot flying. Crews received better ratings if the PF concentrated on the initiation of actions according to checklists or standardized procedures.

Additionally, it was advantageous if the PF made sure that there were no misunderstandings regarding the information communicated and performed clarifying moves.

RQ: Is there a difference in communication between baseline crews and crews who applied the Man4Gen procedure concept during different stages of the scenario? (Communication framework)

Communication data (WP5) indicated that most crews applied mandatory checklists during the post-phase (lightning strike, descent, and landing). However, the pre-phase was characterized by higher levels of standardized communication referring to the ECAM messages while the post-phase, which was the more stressful flight phase, was characterized by an increased incidence of other types of communication actions. The more challenging post-flight phase was characterized by pilots more often thinking aloud, which helped them assess the situation and gain/retain SA more easily. This result corroborates findings in WP2 (simulator and fMRI), that verbalization supports not only information sharing but also the process of understanding. However, during the post-phase there were also significantly more disagreements between the crew members, mistakes happened more often, and more non-directed verbalizations that did not deal with the current situation were performed. The differences between the more routine and more challenging and stressful flight phases were probably a result of the stressful phase being more ambiguous, thereby eliciting more discussions (e.g., planning, problem solving). Nevertheless, pilots more often made sure that the communicated information was correctly understood which indicates an emphasis on effective communication and the successful alignment of mental models. Supporting this assumption, communication data also indicates that pilots shared mental models of the situation more often in the post-phase than in the pre-phase.

Concerning the crew members of the group applying the new Man4Gen procedure concept, it can be summarized that they engaged in slightly more proactive communication, meaning that relevant information about the current situation including task planning and management was shared more often than in the group of crews who were not trained to apply the newly developed procedure. The same holds true for proactive standardized communication (explanations and questions) including callouts, checklists, SOPs, ECAM messages or information from other supporting standardized materials (e.g. FCOM and QRH). In line with results concerning proactive standardized communication, the procedure crews also demonstrated their understanding when receiving information from their colleagues. Furthermore, procedure crews more often showed that they had shared awareness of the current situation. This indicates that the Man4Gen procedure concept had an influence on the crews and triggered them to share relevant information more frequently.[9]

[9] For more details concerning the analysis see TN-6.5.1.

RQ: How do higher rated crews differ from lower rated crews when assessed by tagging SA/SSA relevant categories? (SA framework)

The two highest rated crews used relevant components (i.e., perception, process of understanding, understanding, and anticipation) for situation assessment on average 135.75 times, whereas the two lowest rated crews used them only 106.75 times. Furthermore, a noticeable difference was found regarding the ratings of the crews' understanding of a situation. While the two highest rated crews on average demonstrated 46 instances of understanding, the two lowest rated crews demonstrated only 25.75 instances. Similar results were found regarding anticipation of future states and anticipation of future strategies, hence sensemaking. While the two highest rated crews showed 10.5 and 12.75 instances of anticipation of future states and possible strategies, respectively, the two lowest rated crews showed only 7 and 4.25 instances (for more details see D6.5).

Regarding the whole scenario, descriptive analyses reveal that the two highest rated crews were rated as showing more overall SA related components, more understanding of the current situation, and slightly more anticipations of future states and possible strategies to deal with the situation in the future than the two lowest rated crews did.

This pattern was also observed in some but not all individual parts of the scenario when separating the session into different flight phases based on various events. Essentially, this pattern also occurred during the lightning strike when most scenario threats accumulated and partly appeared during the descent approach and localizer capture. Only the flight phase including the oil temperature advisory and warning was characterized by a different pattern. During these events there were almost no differences between the highest and lowest rated crews regarding their SA assessment. However, overall in this flight phase the lowest rated crews showed slightly more observable SA components and more choice making decisions. It is important to state though that this result turned up because one of the two lowest rated crews (204) was characterized by a great amount of choice making that differed from the other crews in several aspects. The decisions of this crew were not made when the mental models of the PF and PM were aligned (see TN-6.5.1). The PF of crew 204 made his decisions without thinking through and discussing the occurrences/situations with the PM. Therefore, the descriptive analyses suggest that there was a qualitative difference between the two highest and two lowest rated crews during the scenario regarding the performance of SA components they demonstrated (perception, process of understanding [sensemaking], understanding, and anticipation).

Concerning the different types of SA the following observations can be summarized (see D6.5, TN-6.5.1). The content of the crew's nonverbal and verbal actions was categorized into seven classes (cf. D1.4): aircraft (AC) state, fuel, environment (e.g. weather), system state including modes and energy awareness, aircraft position and its environment, flight path/navigation (e.g. landing possibilities), and standardized communication (callouts, checklists, procedures). In general, all crews demonstrated the most relevant SA components regarding the aircraft state and the system state. All other categories were rated somewhat lower, reflecting that the threats from the scenario were mostly concerned with malfunctions of the aircraft and aircraft modes. Discussions and perceptions regarding the fuel status were rarely present, reflecting the fact that the amount of fuel was not a threat during the scenario. Overall, the two highest rated crews demonstrated more SA components regarding the aircraft and system states and preformed more callouts. Although the two lowest rated crews showed slightly more perceptions of SA relevant information concerning the AC state, aircraft position and the flight path and navigation they seemed not to have processed this information as successfully as the two highest rated crews. The two highest rated crews on the other hand demonstrated more perceptions of SA relevant information regarding the system state and carried out more callouts regarding this topic. This seems to have helped them process information better, i.e., make better sense of the situation. These findings underline the importance of effective communication, which supports crews in their awareness of crucial aircraft and system states. This is in line with the result that the two highest rated crews were rated as showing more understanding of the aircraft state, the environment, aircraft position, system and energy management and the standardized communication performed by the colleagues. The two highest rated crews were also characterized by showing more anticipation of future states concerning the aircraft, the environment and the flight path and showing more anticipation of possible strategies to deal with the aircraft status and the flight path.

Overall, the two highest crews demonstrated more understanding, anticipation of current and future states and planning of strategies concerning information and actions coupled to SA than did the two lowest rated crews. This higher incidence of demonstrated SA components may bolster the understanding and anticipation of the threats and possibilities of the scenario and thus lead to a higher DFCP rating. For a more detailed discussion and more results, please see D6.5, TN-6.5.1.

Methodologically, the results imply that verbal and non-verbal communication are not only useful for semantic and numerical communication analyses, but also that it is feasible to apply SA and social SA relevant categorisations to produce additional SA assessments that can supplement existing pilots' assessment and expert performance ratings. However, at this stage this type of analysis needs to be defined as exploratory and thus has to be investigated further in subsequent experiments.

Additionally, findings from the simulator experiments together with the results from secondary research and fMRI studies allowed us to develop a different kind of conceptual model of SA: the Man4Gen SA and social SA model (see D6.5).

3.4. Methodological and Operational considerations

3.4.1. Methodological considerations

It can be summarized that a great deal of investigation has already been performed on how to measure or assess SA, but that there are gaps concerning the assessment of the processes behind SA (perception, comprehension, and projection), concerning various issues associated with those processes, and there are gaps concerning the assessment of social SA, especially in unexpected and challenging situations. A single, best method of measurement or assessment tool for SA and its social variations most likely cannot be found. Instead it was concluded that a selection of existing methods and their combination with updated and newly developed tools would best allow the research questions and objectives within Man4Gen to be addressed.

Concerning the methods and tools applied the following can be summarized (see also D6.5):

fMRI: Although the research environment in an MR scanner is very limited in space and requires pilots to refrain from motion,, fMRI allows for task designs that measure real world aviation skills. As such it has proven to be a valid method in the field of neuroergonomics. Consequently, it is recommended to use the fMRI method when insights concerning neuronal networks associated with relevant pilot competencies are sought. Using fMRI in an aviation context is at an early stage but results indicate high future potential and promising findings, especially when combined with methods simultaneously used, such as eye-tracking.

Heart rate: HR measures within Man4Gen proved to be a measure of pilots' short-term stress levels which complement the subjective stress ratings gathered through questionnaires. With current measurement equipment, such as HR-recording wrist-watches and the accompanying software, monitoring pilots' arousal state in real-time is possible at low costs and stressful situations can be quickly identified. HR measures thus also can be used to assess if certain training is/was effective, or can be applied related to social SA issues. When carrying out more advanced analyses such as a correlation with other types of measures, additional software, different analytical tools, and a deeper knowledge of the method are however necessary.

Personality trait questionnaires: The PCI questionnaire did help identify significant characteristics in the specific study population of the WP5 experiments. In future studies, the PCI questionnaire could therefore support the grouping of pilots into crews. The questionnaire could also be adapted to assess pilots' coping strategies in more detail and thus support research regarding resilience. Comparable to the PCI, the IRI questionnaire helped to get insights into the sample's personality characteristics. In terms of social SA research in an aviation context IRI can help put together pilots with certain characteristics to specifically investigate the concept of social SA in future experiments. Concerning aviation research a questionnaire such as the TICS-SSCS scale, which assesses chronic stress, can help further determine an experimental sample and put together adequate crews.

Self-rating questionnaires: If the NASA-TLXraw is applied correctly, i.e., within 30 minutes after an experiment was completed, it delivers good results in terms of pilots' own task load experiences. The NASA TLXraw can deliver important insights into whether a scenario design successfully triggered the desired level of task load. Furthermore, collecting data with the NASA TLXraw scale is easy and inexpensive. However, subjective measures such as this self-rating questionnaire also have their limitations and should not be applied on their own. They should be combined with other objective measures when evaluating pilots SA and its related competencies. As with the NASA TLXraw, if the 10D-SART is applied in the right manner, i.e., right after the simulation or up to 30 minutes after an experiment at the latest, it delivers valuable results in terms of a pilot's subjective experience concerning his/her SA. The application of the 10D-SART is also quick and easy and this questionnaire has the same advantages and disadvantages as the NASA TLXraw. A limitation of the questionnaire is that is lacks the assessment of social SA though.

Social SA scale: The social SA visual analogue scale (SSA-VAS) was developed within Man4Gen and is designed to capture pilots' subjective estimates of stress and control for themselves and for their colleagues. By applying two different formulas empathic accuracy for each crew member can be assessed, which is an important concept in social SA. Thus, this quick and easy to apply self-rating questionnaire closes a gap in SA research addressing a social component of crew performance. As a self-rating method, the SSA-VAS has the same advantages and disadvantages as other subjective measures. The questionnaire is easy to apply and since it can be easily adapted, the SSA-VAS can be used in future aviation research investigating the social aspects of pilots' resilience.

Expert and de-brief interviews: Problem-centred semi-structured expert interviews are a reliable method to collect opinions from experienced professionals in the context of SA and social SA research; And de-briefing interviews, both single and joint crew, including video replay, support post-experimental expert and research observations validating the results and highlighting specific issues during a scenario. The method further serves the purpose of giving participating pilots the opportunity to share their experience concerning the overall experiment with the experimenters; this also ensures that only data is included into an analysis that is valid.

Man4Gen dialogue schemes (verbal & non-verbal): Communication data (verbal and non-verbal) is a rich data source in an aviation context, enables dynamic analyses to be carried out over time and serves single SA as well as social SA research. The various Man4Gen annotation schemes generated an almost perfect inter-rater accuracy and therefore are a reliable way to create a detailed and comprehensive data source for different kinds of analysis, i.e. qualitative and quantitative. By itself, the dialogue scheme data gives insights into variations in communication patterns over time. Communication data is well suited to be combined with various other kinds of measures, especially with expert performance ratings, such as the DFCP. In this combination behavioural markers and predictors for better crew performance can be identified. Annotating communication data along the Man4Gen dialogue schemes allows points in time during a scenario to be found at which performance deteriorates. The annotation schemes used in Man4Gen can also be adapted to future research questions of interest, for example to safety research dealing with the resilience of pilots. Additionally, corpus communication data analysis has great future potential to be combined with psycholinguistic measures. The corpus initiated in Man4Gen regarding "*crew scenarios including challenging and surprising events*" can be expanded in the future and can thus provide an important contribution to investigations in automatized communication analysis. Taken together, communication analysis is an important tool for pilots' performance assessment, especially regarding SA and social SA.

Man4Gen SA framework from communication: The annotation of the communication transcriptions along the SA framework enables the analysis of SA and social SA over time. The single concepts of the framework aligned to the Man4Gen SA working definition and involved in situation assessment and SA as well as social SA, are substantiated by communication semantics (e.g. future tense for anticipation). Additionally, the data captured with the framework allows for clear visualizations of single pilot's and crew situation assessment patterns and thus delivers behavioural markers for performance, especially in combination with expert performance ratings. The data can be easily combined with the procedure (see D6.1) and ECOM (see D6.6) evaluations and can therefore substantially contribute to an overall crew evaluation.

3.4.2. Operational recommendations

A summary and list of key findings from an SA and social SA perspective that support crews to be more resilient during challenging and unexpected situations (when a rapid transition from monitoring to decision making and successfully controlling an aircraft is required [cf. DFCP ratings]), is provided below.

- Improve crew abilities to carry out workload and information management and the processes of problem solving and decision-making, by prioritizing information and identifying task relevant issues.

Analysis from questionnaires (i.e., NASA TLX raw, SART, SSA-VAS) indicate a significant relationship between workload and SA. Workload and stress patterns increased during surprising and challenging flight phases and at the same time SA decreased. Thus, in Man4Gen an increase in workload and stress were linked with a decline in different types of SA. Therefore, it is assumed that if information dispersal is minimized and linked to data that is currently being processed (short-term) and that the overall picture of a situation is presented over time (long-term), the crew's workload and stress levels could be lowered.

Correlation between task load and perceived control and stress highlight that highly demanding situations lead to lower control and higher stress experiences. The latter is a finding that was also supported by the HR measures.

- Improve social awareness in pilots while at the same time building consciousness about focusing on the task at hand if necessary.

A high task load is associated with less empathic accuracy for stress during demanding situations. More precisely, we found evidence from the social SA evaluations that the PF's and PM's abilities to assess their colleague's level of stress and control were diminished during challenging and stressful flight phases (cf. WP2). The results show that stress affects the perception and understanding of others, i.e., leads to decreased empathic accuracy.

- Improve the efficiency of crew communication and cognitive offloading strategies within a socio-technical system.

The fMRI and simulator results in Man4Gen highlight the important role of verbalization, i.e., effective verbal and non-verbal communication. Both serve various purposes within a socio-technical system, which are the following: a) assisting first, second and third level SA, as well as the process of understanding and decision-making; b) supporting recall from memory; c) offloading cognition to the environment; d) forming a theory of mind and thus a mental model of a situation; and e) sharing one's mental model and consequently fostering social SA.

- Put a greater focus on the role of the PM - i.e., how the pilot monitoring should be actively engaged in challenging situations.

Qualitative observation and examination of communication data suggests that crews perform better if the pilot monitoring is able to a) take initiative but also listens and follows instructions; b) share his/her mental model regarding threats (e.g. engine status in WP2) in regular intervals; c) accompany his/her action with a verbal explanation; d) verbalize his/her thoughts during problem solving; and e) prioritize his/her actions.

- Improve leadership and teamwork by concentrating on when to use initiative and give directions, when to communicate relevant concerns and intentions, and when to tune into others for developing empathic accuracy; and put a focus on communication strategy (what, when, how) and on the application of non-verbal communication.

Quantitative analysis and statistical modeling from communication and expert rating data indicate that crews perform better if a) the PM is able to retrieve information autonomously (e.g., from memory, checklists, FCOM, QRH); b) the PM is able to share his mental model regarding plans and actions; c) the PF shows initiative and gives directions when required, thus performs adequate leadership; d) the crew uses non-verbal communication sufficiently in order to support and reinforce spoken words, to call for attention, and if necessary, to substituting verbal communication.

4 COMBINING THE RESULTS

In the Man4Gen project two research strands have been working in parallell with analyses work from different theoretical approaches, the situation(al) awareness (SA) approach and the sensemaking (SM) approach (see chapter 2, D6.6 and capter 3, D6.5 for an overview). In this chapter we discuss the overlap and differences of the results from the SA framework analysis (SA) and the ECOM analysis (SM). A comparison of two crews from the NLR experiment in WP5 (see D5.2 for more information) is used as an example to illustrate how the different analyses complement and differ from each other. The industry based assessment (DFCPs) of the highest (106) and lowest (109) scoring crews have been chosen for the example. The analysis inlcuded starts at the point of the lightning strikes until a decision is made of the final destination.

Figure 16 and Figure 18 represent the annotation patterns for the SA framework based on communication data from the WP5 experiments. The content of the crew's nonverbal and verbal actions was categorized into five main SA labels: perception (square), process of understanding (diamond), understanding (triangle), anticipation (cross), and decision (circle). The content of the five main SA framework labels was further broken down into seven context classes (cf. D1.4, D6.5): aircraft (AC) state (CPT = dark blue, FO = light blue), system state including modes and energy awareness (CPT = dark orange, FO = light orange), aircraft position/flight path/navigation (e.g., landing possibilities; CPT = dark green, FO = light green), standardized communication (callouts, checklists, procedures; CPT = dark grey, FO = light grey), environment (e.g., weather; CPT = dark violet, FO = light violet), and fuel (CPT = dark red, FO = light red).

Figure 17 and Figure 19 shows the distribution of activities across the four ECOM layers based on video data from the WP5 experiments. The activities of the crews was abstracted and categorized into 10 different categories, each with a different color: planning flight path (yellow), changing destination (purple), briefing concerned parties (turquois), assigning responsibilities (dark green), assessing threats/certainties/uncertainties (dark green), follows checklist (teal), engine management (red), aircraft configuration (light blue), monitoring aircraft status (pink and address loud bang (green). In Figure 17, in the first minutes of the time interval the focus is on the two lower layers of the ECOM, followed by an elevation of activities almost exclusively at the two higher levels. The majority of the activities are related to planning of the flight path (yellow). In figure 4, in the first minutes of the time interval is focused on the two lower layers of the ECOM, followed by engine management activity (red) on the two higher layers. The activities are subsequently spread across the layers until the last minutes when the crew spend more time on the higher layers. Throughout the interval the dominating activities are short-term engine management and flight path planning.

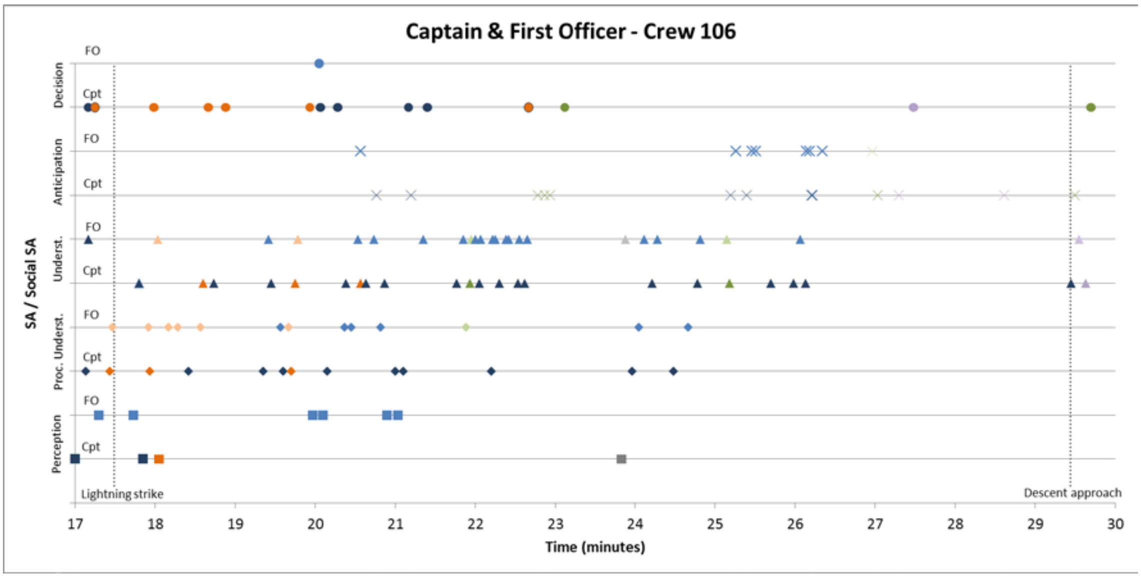

Figure 6 Annotated categories for Captain (Cpt) and First Officer (FO) of crew 106. The Cpt of crew 106 performed the role of pilot flying (PF) during the lightning strike. Annotations are colored regarding to their content: blue = aircraft state, orange = system state, green = flight path & navigation, grey = callouts, violet = weather, red = fuel. SA = Situation(al) Awareness; Proc. Underst. = Process of Understanding; Underst. = Understanding; OT = oil temperature advisory; LOC = localizer. Dashed lines indicate the critical scenario events.

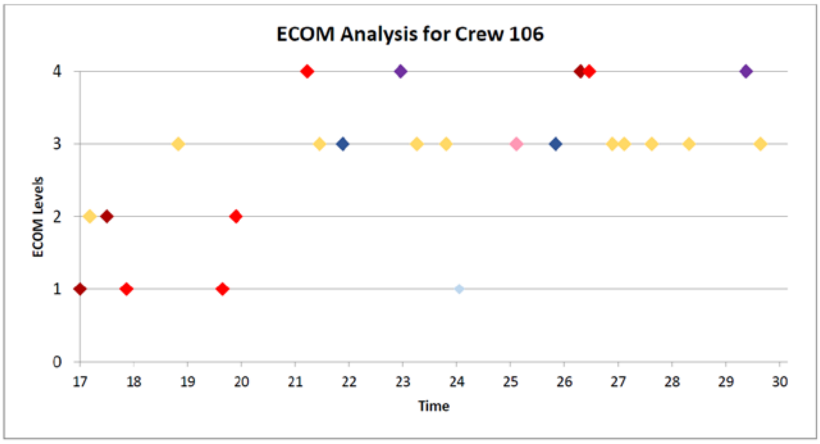

Figure 7: ECOM analysis 106. Level 1 on the Y-axis denotes the tracking layer, level 2 regulating, level 3 planning and level 4 goal setting. Activities are coloured regarding their content: planning flight path (yellow), changing destination (purple), briefing concerned parties (turquois), assigning responsibilities (dark green), assessing threats/certainties/uncertainties (dark green), follows checklist (teal), engine management (red), aircraft configuration (light blue), monitoring aircraft status (pink and address loud bang (green).

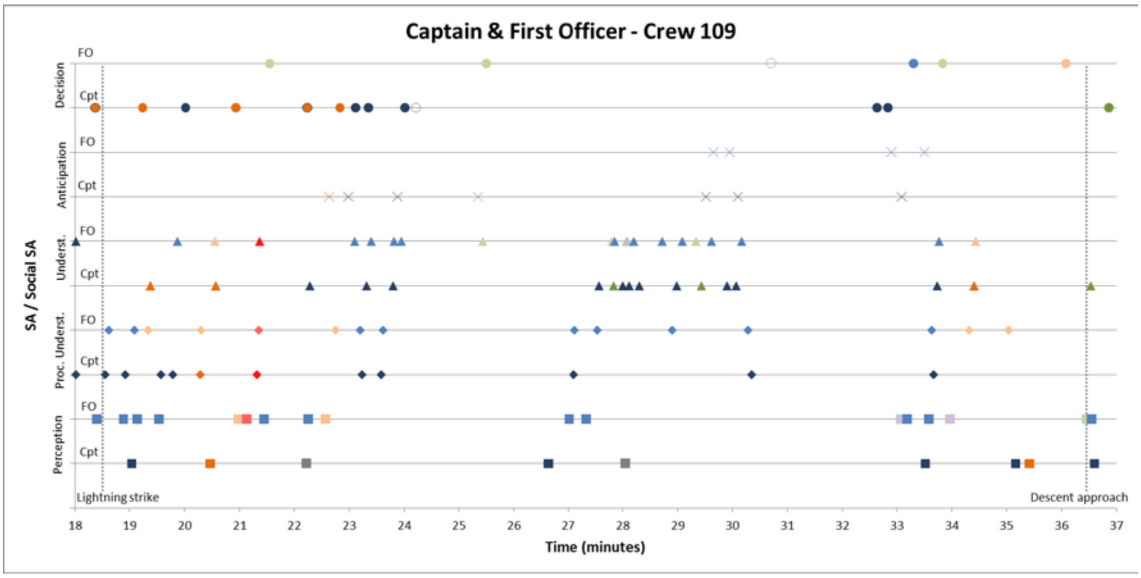

Figure 8 Annotated categories for Captain (Cpt) and First Officer (FO) of crew 109. The Cpt of crew 109 performed the role of pilot monitoring (PM) during the lightning strike. (see Figure 16 for color codes)

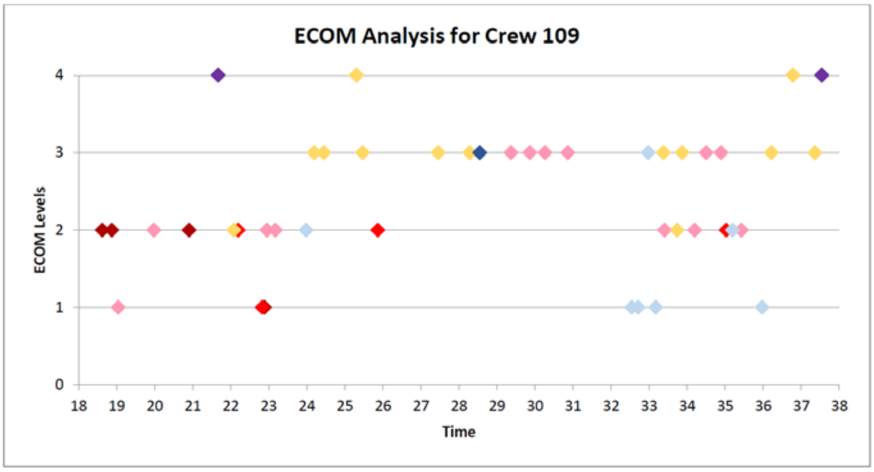

Figure 9: ECOM analysis 106. Level 1 on the Y-axis denotes the tracking layer, level 2 regulating, level 3 planning and level 4 goal setting. (see Figure 17 for color codes)

4.1. Combined results

Annotations along the SA framework indicate that the most frequently rated content of annotations refers to the *aircraft state* and *system states* in both crews. While system states were discussed most frequently right after the occurrence of the lightning strike and lasted until the first third of this phase, aircraft states were debated over the whole course of the flight phase. In both crews the CPTs were more active in anticipation and decision-making, although, the roles of the CPTs were different (CPT 106 was PF, CPT 109 was PM). Further contrasts between crews 106 and 109 are the following: crew 106 is characterized by patterns that distinguish them from crew 109 by having more closely aligned SA activities of PF and PM regarding *processes of understanding* and *understanding* during the lightning strike phase. Additionally, crew 106 shows patterns concerning more activities in higher cognitive SA categories (e.g. understanding and anticipation) whereas crew 109 presents patterns with less activity in higher cognitive SA categories and more in the lower (i.e., perception). The FO from crew 109, who is the PF, makes decisions that are not aligned with topics that were discussed, i.e., understood or anticipated. This kind of behavior/pattern never shows up in crew 106 over the period of the lightning strike.

The ECOM graphs of abstracted activities show that Crew 106 are initially active on a tracking and regulating layer, managing their engines following the lightning strike. The remaining time is mainly spent on a planning and goal-setting layer. Their focus is on evaluating engines for landing and assessing options of where to land. Crew 109 on the other hand spend considerably more time on a regulating layer, focusing on engine management and short-term route management. They request weather at the alternate airport, but never discuss the option of landing there. Following some initial short-term route planning the crews get information about the weather in Amsterdam (without requesting it). They further discuss the status of their engines and finally conclude that the weather in Amsterdam is within their limits. Overall crew 106 spend more time on the goal-setting layer in relation to planning compared to crew 109 and crew 109 spend considerable more time on the regulating layer in relation to planning compared to crew 106. The finding is complemented with the long- and short-term analysis and strategy patterns for the two crews (see section 2.3.3), demonstrating the differences in the crews ability to re-assess the situation, consider options and take action.

Similar patterns can be found in the SA framework analysis and the SM ECOM analysis. The ECOM analysis trend shows crew 106 rapidly climbing from short-term planning on the tracking and regulating layers to the higher layers of long-term planning and goal-setting. Similarly in the SA framework analysis crew 106 are initially characterized by perceptions and processes of understanding in the beginning of the flight phase whereas in the second half they are mainly characterized by understanding and anticipation. In the SA framework analysis crew 109 is characterized by perception and trying to understand what is going on. The ECOM analysis complements this view and shows that crew 109 is making short term plans and "*gets stuck*" on discussions of what is going on here and now (e.g. engine management and short-term route management).

The SA framework analysis along with the communication data further shows that the decisions made by crew 109 are not congruent with the content of SA categories (i.e., understanding and anticipation). For example, the FO makes a decision referring to the flight path and navigation but does this without the preceding processes of understanding and understandings of the topic. The same results are found in the ECOM analysis, where decisions are made without considering different options, even in situations where the relevant information had been gathered. Additionally, the de-briefing results further show that crew 106 felt that they had ample time and that they could have flown around for a long time. Crew members of crew 109 were, on the other hand less confident as they did not know for how long engine 2 would sustain after they started it.

As mentioned above the ECOM and SA frameworks partially share the same research environment and follow similar processes of analysis. All types of analyses (SA framework, communication and ECOM) are dynamic and consider time as a dimension of analysis. However, next to overlapping results they also deliver approach specific insights and thus add value to each other. In terms of the ECOM model and analysis the uniqueness is the description of activities carried out simultaneously on different time spans, including a detailed analysis on the short term and long term planning. In regard to the SA framework the specificity lies in addressing the social aspects, and thus the interaction between the single SA tagging categories between crew members.

4.2. Conclusions

Each of the two approaches applied delivers results concerning crew patterns during stressful and challenging situations. Although standing alone the two methods deliver unique insights they also overlap in various findings. This not only strengthens the interpretation drawn from each method but also validates the single approaches' results. The application of both frameworks offers possibilities for future research and analysis and can also be combined with additional measures.

5 SUMMARY AND FURTHER RESERACH

In this report the central theories, methods and main outcomes from the two research strands, situation(al) awareness and sensemaking, have been presented. Methodological considerations from each perspective have been discussed at the end of the individual chapters (2 and 3). A triangulation of methods has been carried out in the project, including the two research strands and the industry based assessment. Next to conventional assessment techniques, methods were applied that are new and innovative, especially in the field of aviation (such as fMRI, social SA tools, and ECOM). The combination of the analyses and outcomes of these techniques (fMRI, [social] SA and sensemaking) resulted in a mutual validation of the applied approaches. Therefore, we are now better able to test the interfacing between man and machine (MMI), the required levels of training, and associated factors in a dynamic test and simulation environment, which allows developing new MMI configurations. The results from SA and SM have individually been linked to the industry assessment and in Chapter 4 we have combined part of the results from the two research perspectives to demonstrate how different methods can complement and validate each other. The overall results from the WP2 and WP5 experiments show that there is a great variation in how the crews understand, prioritise, manage uncertainties and take action following the unexpected events introduced in the scenarios. Operational recommendations and the findings that support them on how to improve crew abilities are summarised in sections 2.5.2 and 3.4.2. The project has opened up a new domain in research options and techniques that can be instrumental for cockpit design, layout and training. The findings further raise several important questions regarding challenges and possibilities for pilots to maintain control in surprise situation. Suggestions for further research are summarized below.

5.1. Further research questions

- What factors (patterns) obstruct and enable detection of a mismatch crew’s mental model? This includes mismatches between crew members as well as between crew members and automated systems. For example, factors hindering or facilitating the detection of abnormalities, different mental models of the crew members, the coupling between sensors and symptoms (what is trusted, what is not?), what symptoms do experts pick up on that a novice doesn’t?
- What level of system knowledge is necessary to be able to take appropriate actions following unexpected events? What are patterns of breakdown due to oversimplifications (i.e., simplified models of the automated systems?) Such studies may include, for example, the understanding of what are the models of system knowledge today, if they are similar or very different between pilots and identification of problematic systems.

- What vulnerabilities are created as a result of managing increase of tasks and cope with trade-offs in unexpected situations? How do multiple tasks and goals shape the operational environment? The analysis in Man4Gen has identified several vulnerabilities, including difficulties in re-framing following unexpected events. Further studies could broaden the understanding of crew breakdown patterns relating to unexpected events.
- What strategy patterns facilitate re-gaining control following unexpected events? Several strategies and patterns of strategies have been identified as part of the research carried out in this project. Abilities to adapt through prioritizing tasks and, the ability to rapidly gain an overview of the situation and to have interaction on all ECOM layers are strategies that have been identified as useful as coping mechanisms. Building on this research more strategy types and patterns could be identified to increase our understanding for successful adaptation following surprise.
- What types of mental efforts are required to comprehend particular display types to allow for an evidence-based design process that minimizes the cognitive workload or potential ambiguities for the operators?
- What kinds of behaviors and strategies support individual pilots and crews in their stress and workload management? How should the socio-technical system be further improved to foster crews' resilience to possible threats and challenging events?
- What crew resource management factors significantly support shared risk anticipation and threat management, helping create an ultra-resilient socio-technical system?

6 REFERENCES

Bonner, M. A., & Wilson, G. F. (2002). Heart rate measures of flight test and evaluation. *The International Journal of Aviation Psychology*, *12*, 63–77.

Carletta, J., Isard, S., Doherty-Sneddon, G., Isard, A., Kowtko, J. C., & Anderson, A. H. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, *23*, 13–31.

Davis, (1980). A multidimensional approach to individual differences in empathy.

Endsley, M. R., & Jones, W. M. (2001). A model of inter- and intrateam situation awareness: Implications for design, training and measurement. In M. McNeese, E. Salas, & M. R. Endsley (Eds.). *New trends in cooperative activities: Understanding system dynamics in complex environments*. Santa Monica, CA: HFES. 46–67.

Feigh, K. M. (2010). Incorporating multiple patterns of activity into the design of cognitive work support systems. *Cognition, Technology & Work, 13*, 259–279.

Field, J. & Lemmers, A., (2014). Man4Gen: Manual Operation of 4th Generation Airliners. In *Proceedings of the 31st Conference of the European Association of Aviation Psychology*. Valletta, Malta: EAAP.

Field, J. Rankin, A. & Woltjer, R. (2014). Modelling Flight Crew Strategies in Unexpected Events: A Cognitive Systems Engineering Perspective. In *Proceedings of the 31st Conference of the European Association of Aviation Psychology*. Valletta, Malta: EAAP.

Field, J., Woltjer, R., Rankin, A., & Mulder, M. (2015). Experimental investigation of flight crew strategies in handling unexpected events. In *Proceedings of the 18th International Symposium on Aviation Psychology*. Dayton, OH: Wright State University.

Greenglass, E. (1999). The Proactive Coping Inventory (PCI): A Theory-Guided Multidimensional Instrument Esther Greenglass*, Ralf Schwarzer**, Dagmara Jakubiec* Lisa Fiksenbaum* & Steffen Taubert*.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*, 904–908.

Hart, S. G., & Staveland, L. E. (1988). Results of empirical and theoretical research. *Ergonomics: Skill, Displays, Controls, and Mental Workload*, *2*, 408.

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. Boca Raton, FL: CRC Press/Taylor & Francis.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality*, *61*, 587–610.

International Civil Aviation Organization. (2013). *Manual of evidence-based training.*

Klein, G., Snowden, D., & Pin, C. L. (2011). Anticipatory Thinking. In K. L. Mosier & U. M. Fischer (Eds.), *Informed by Knowledge: Expert Performance in Complex Situations* (pp. 235–246). New York, NY: Taylor & Francis Group.

Klein, G., Wiggins, S., & Dominguez, C. O. (2010). Team sensemaking. *Theoretical Issues in Ergonomics Science*, *11*, 304–320.

Kontogiannis, T., & Malakis, S. (2011). Strategies in controlling, coordinating and adapting performance in air traffic control: modelling "loss of control" events. *Cognition, Technology & Work, 15*, 153–169.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, *41*, 841–849.

Parasuraman, R., & Rizzo, M. (Eds.). (2008). *Neuroergonomics: the brain at work*. New York: Oxford University Press.

Sadaghiani S, Kleinschmidt A (2013). Functional interactions between intrinsic brain activity and behaviour. NeuroImage, 80:379-86.

Singer, T., & Lamm, C. (2009). The Social Neuroscience of Empathy. *Annals of the New York Academy of Sciences*, *1156*, 81–96.

Stepniczka, I., Tomova, L., Niedermeier, D., Peschl, F.-M., & Lamm, C. (2015). Social Situation Awareness: Empathic Accuracy in the Aircraft Cockpit. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (COGSCI 2015).*

Taylor, R.M., (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational awareness in aerospace operations (AGARD- CP-478; pp.3/1-3/17). Neuilly-Sur-Seine, France: NATO-Advisory Group for Aerospace Research and Development.

Tomova, L., von Dawans, B., Heinrichs, M., Silani, G., & Lamm, C. (2014). Is stress affecting our ability to tune into others? Evidence for gender differences in the effects of stress on self-other distinction. *Psychoneuroendocrinology*, *43*, 95–104.

Rankin, A., Woltjer, R., Field, J., & Woods, D. (2013). "Staying ahead of the aircraft" and Managing Surprise in Modern Airliners. In I. Herrera, J. M. Schraagen, J. Van der Vorm, & D. Woods (Eds.), *Proceedings of the 5th Resilience Engineering Association Symposium* (pp. 209–214). Soesterberg, NL: Resilience Engineering Association.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science*, *16*, 409–421.

Woltjer, R., Field, J., & Rankin, A. (2015). Adapting to the Unexpected in the Cockpit. In *Proceedings of the 6th Resilience Engineering Association Symposium*. Lisbon, PT: Resilience Engineering Association.

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. Boca Raton, FL: CRC Press/Taylor & Francis.