EASA
European Aviation Safety Agency

Final Report EASA_REP_RESEA_2015_3

Research Project:

Startle Effect Management

An agency of the European Union

Disclaimer

This study has been carried out for the European Aviation Safety Agency by an external organization and expresses the opinion of the organization undertaking the study. It is provided for information purposes only and the views expressed in the study have not been adopted, endorsed or in any way approved by the European Aviation Safety Agency. Consequently it should not be relied upon as a statement, as any form of warranty, representation, undertaking, contractual, or other commitment binding in law upon the European Aviation Safety Agency.

Ownership of all copyright and other intellectual property rights in this material including any documentation, data and technical information, remains vested to the European Aviation Safety Agency. All logo, copyrights, trademarks, and registered trademarks that may be contained within are the property of their respective owners.

Reproduction of this study, in whole or in part, is permitted under the condition that the full body of this Disclaimer remains clearly and visibly affixed at all times with such reproduced part.

Dedicated to innovation in aerospace

NLR-CR-2018-242 |

Startle Effect Management

CUSTOMER: European Aviation Safety Agency

NLR – Netherlands Aerospace Centre

Netherlands Aerospace Centre

NLR is a leading international research centre for aerospace. Bolstered by its multidisciplinary expertise and unrivalled research facilities, NLR provides innovative and integral solutions for the complex challenges in the aerospace sector.

NLR's activities span the full spectrum of Research Development Test & Evaluation (RDT & E). Given NLR's specialist knowledge and facilities, companies turn to NLR for validation, verification, qualification, simulation and evaluation. NLR thereby bridges the gap between research and practical applications, while working for both government and industry at home and abroad.
NLR stands for practical and innovative solutions, technical expertise and a long-term design vision. This allows NLR's cutting edge technology to find its way into successful aerospace programs of OEMs, including Airbus, Embraer and Pilatus. NLR contributes to (military) programs, such as ESA's IXV re-entry vehicle, the F-35, the Apache helicopter, and European programs, including SESAR and Clean Sky 2.
Founded in 1919, and employing some 650 people, NLR achieved a turnover of 71 million euros in 2016, of which three-quarters derived from contract research, and the remaining from government funds.

For more information visit: **www.nlr.nl**

Dedicated to innovation in aerospace

EXECUTIVE SUMMARY

UNCLASSIFIED

Startle Effect Management

Problem area

Startle and surprise effects can influence pilot performance in many detrimental ways. At the very least, these effects serve as a distraction which can disrupt normal operation and erode safety margins. On a more critical level, they can lead to inappropriate intuitive actions or hasty decision making. Well learned procedures and skills can be discarded and are substituted by the first thing that comes to mind. This study is concerned with the impact of startle and surprise effects in aviation and how to mitigate these effects by pilot training. Commercial airline training has always included training for emergencies and abnormal situations, whether at ab-initio training, or during a pilot's career in type conversion and recurrent training. To some extent a level of "unexpected events" may be included in training programs, which may lead to some of the effects associated with startle and surprise. Is it possible to develop a technique to mitigate the effect of startle, and to design a training session for this technique?

REPORT NUMBER
NLR-CR-2018-242

AUTHOR(S)
J.N. Field
E.J. Boland
J.M. van Rooij
J.F.W. Mohrmann
J.W. Smeltink

REPORT CLASSIFICATION
UNCLASSIFIED

DATE

KNOWLEDGE AREA(S)

DESCRIPTOR(S)
Surprise
Training
Startle

Description of work

As part of this study into Startle Effect Management for EASA, NLR and KLM developed a simple technique that could be applied operationally in an unexpected situation, or when there was a startle or surprise experienced by one or both crewmembers. The associated training session was developed within the boundaries of a typical airline recurrent training session using the time slot and facilities that would be available to an airline. In order to evaluate the startle effect management technique steps, and the training session that was developed, an experiment was set up using KLM line pilots and KLM flight simulation training devices to put forty pilots through the training session and evaluate the results. The participants were volunteers and included both short- and long-haul pilots, experienced and less experienced pilots, and instructors. The experiment was set up to be able to evaluate both the classroom and simulator sessions using test scenarios in the simulator. The participants' performance was evaluated using the KLM SHAPE behavioural marker system. This report describes the experiment and the results of the evaluation of the startle effect management technique.

Results and Application

The results of the experiment demonstrate that the startle effect management technique developed by the project team can be applied effectively by line-pilots in the simulated operational environment. By applying the technique, the pilots were more effective on key behavioural indicators such as "information collection" after an unexpected event. The experiment also demonstrated that it was possible to train pilots in the technique within the limited time of a recurrent training slot, although it was identified that more training repetition would be required to increase the effectiveness over time. The combination of both a classroom session, and coaching-style training on the simulator was more effective than classroom training alone. Additionally, the participants rated the training session highly and scored their self-efficacy highly after the training, which has positive implications for the eventual roll-out and application of the technique within the operational environment.

NLR
Anthony Fokkerweg 2
1059 CM Amsterdam
p) +31 88 511 3113 f) +31 88 511 3210
e) info@nlr.nl i) www.nlr.nl

Dedicated to innovation in aerospace

NLR-CR-2018-242 |

Startle Effect Management

CUSTOMER: European Aviation Safety Agency

AUTHOR(S):

J.N. Field	NLR
E.J. Boland	NLR
J.M. van Rooij	NLR
J.F.W. Mohrmann	NLR
J.W. Smeltink	NLR

NLR - Netherlands Aerospace Centre

No part of this report may be reproduced and/or disclosed, in any form or by any means without the prior written permission of the owner.

CUSTOMER	**European Aviation Safety Agency**
CONTRACT NUMBER	**EASA.2015.C22**
OWNER	**European Aviation Safety Agency**
DIVISION NLR	**Aerospace Operations**
DISTRIBUTION	**Limited**
CLASSIFICATION OF TITLE	**UNCLASSIFIED**

APPROVED BY :		
AUTHOR	**REVIEWER**	**MANAGING DEPARTMENT**
J.N. Field		
DATE	DATE	DATE

Summary

Problem area

Startle and surprise effects can influence pilot performance in many detrimental ways. At the very least, these effects serve as a distraction which can disrupt normal operation and erode safety margins. On a more critical level, they can lead to inappropriate intuitive actions or hasty decision making. Well learned procedures and skills can be discarded and are substituted by the first thing that comes to mind. This study is concerned with the impact of startle and surprise effects in aviation and how to mitigate these effects by pilot training. Commercial airline training has always included training for emergencies and abnormal situations, whether at ab-initio training, or during a pilot's career in type conversion and recurrent training. To some extent a level of "unexpected events" may be included in training programs, which may lead to some of the effects associated with startle and surprise. Is it possible to develop a technique to mitigate the effect of startle, and to design a training session for this technique?

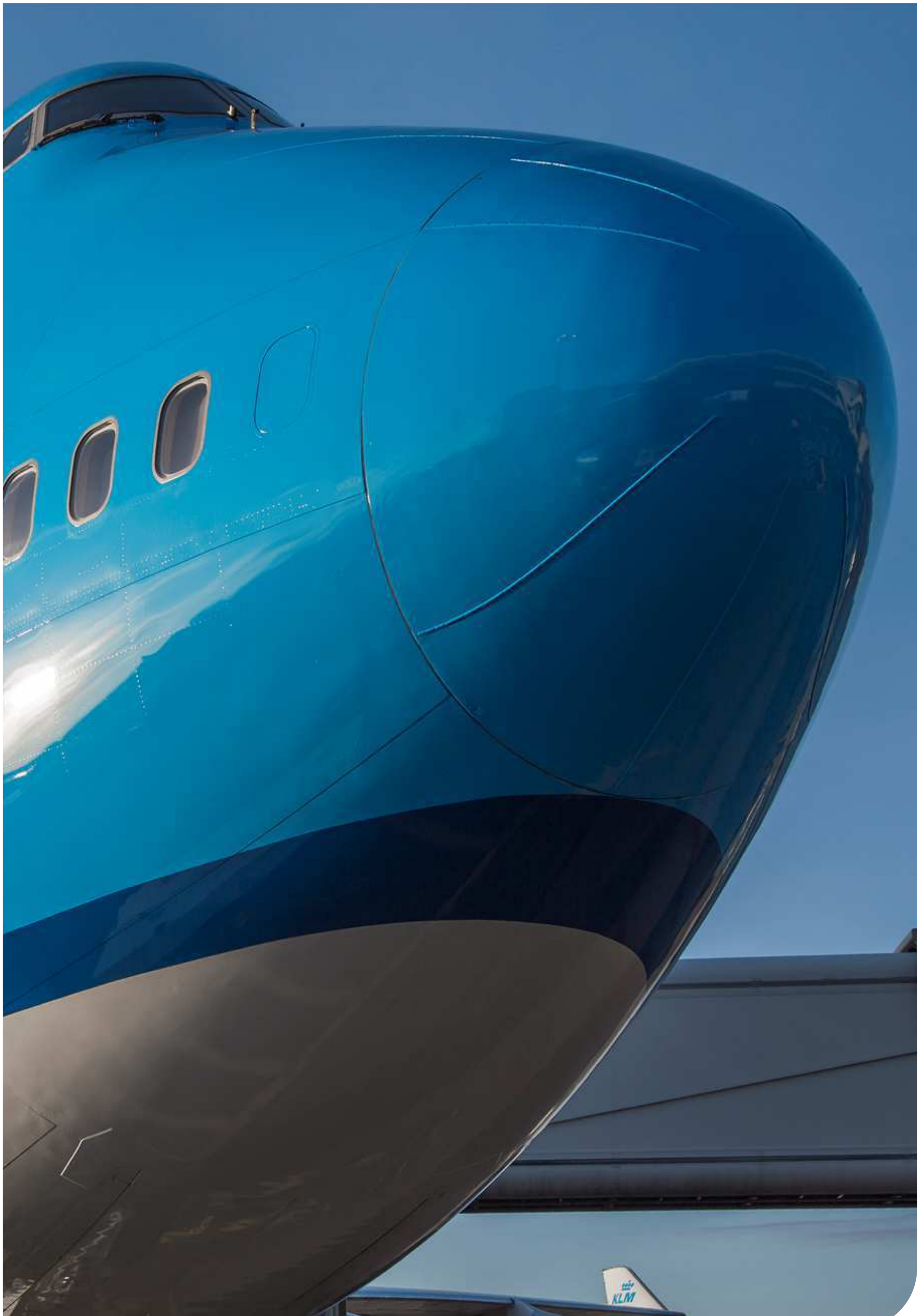
Description of work

As part of this study into Startle Effect Management for EASA, NLR and KLM developed a simple technique that could be applied operationally in an unexpected situation, or when there was a startle or surprise experienced by one or both crewmembers. The associated training session was developed within the boundaries of a typical airline recurrent training session using the time slot and facilities that would be available to an airline. In order to evaluate the startle effect management technique steps, and the training session that was developed, an experiment was set up using KLM line pilots and KLM flight simulation training devices to put forty pilots through the training session and evaluate the results. The participants were volunteers and included both short- and long-haul pilots, experienced and less experienced pilots, and instructors. The experiment was set up to be able to evaluate both the classroom and simulator sessions using test scenarios in the simulator. The participants' performance was evaluated using the KLM SHAPE behavioural marker system. This report describes the experiment and the results of the evaluation of the startle effect management technique.

Results and Application

The results of the experiment demonstrate that the startle effect management technique developed by the project team can be applied effectively by line-pilots in the simulated operational environment. By applying the technique, the pilots were more effective on key behavioural indicators such as "information collection" after an unexpected event. The experiment also demonstrated that it was possible to train pilots in the technique within the limited time of a recurrent training slot, although it was identified that more training repetition would be required to increase the effectiveness over time. The combination of both a classroom session, and coaching-style training on the simulator was more effective than classroom training alone. Additionally, the participants rated the training session highly and scored their self-efficacy highly after the training, which has positive implications for the eventual roll-out and application of the technique within the operational environment.

The technique and training was developed with a broad spectrum of operational line-pilots for transport category commercial aircraft. For the purposes of the experiment it was further tailored to the KLM operational environment, and the intention is that they should be tailored to the airline's own operational culture and environment. By tailoring the exact way of presenting the startle effect management technique to the airline's culture and operation, the acceptance of the technique is expected to be higher and as a result it will be more effective in the operation. KLM have taken the technique that was developed and further refined it for their operation for introduction into their recurrent training and operation.


KLM

Contents

Abbreviations

ACRONYM	DESCRIPTION
A/THR	Autothrottle
AIDS	Accident and Incident Data System
AMSL	Above Mean Sea Level
AMT	Aircraft Maintenance Technician
AP	Autopilot
APS	Aviation Performance Solutions
ASAP	Aviation Safety Action Program
ASRS	Aviation Safety Reporting System
ATC	Air Traffic Control
ATCA	Airline Training Center Arizona
ATM	Air Traffic Management
CAT	Commercial Air Transport
CBA	Cost Benefit Analysis
CBT	Computer Based Training
CFR	Code of Federal Regulations
CLT	Charlotte Douglas International Airport
CMC	Crew Management Course
CNS	Communication Navigation Surveillance
COCO	Co-co-pilot
CRM	Crew Resource Management
CT	Critical Thinking
CVR	Cockpit Voice Recorder
DFDR	Digital Flight Data Recorder
DNW	German-Dutch Wind Tunnels
DNW	German-Dutch Wind Tunnels
EASA	European Aviation Safety Agency
EBT	Evidence Based Training
ECAM	Electronic Centralised Aircraft Monitoring
ECCAIRS	European Co-ordination Centre for Accidents and Incidents Reporting System database
EGPWS	Enhanced Ground Proximity Warning System
EGT	Exhaust Gas Temperature
ESD	Event Sequence Diagram
EST	Eastern Standard Time
ETTO	Efficiency Thoroughness Trade Off
FAA	Federal Aviation Authority (USA)

FAF	Final Approach Fix
FAR	Federal Aviation Regulation
FMS	Flight Management System
FO	First Officer [Co-Pilot]
FTE	Full Time Equivalent
GPWS	Ground Proximity Warning System
HMI	Human Machine Interface
IAS	Indicated Air Speed
ICAO	International Civil Aviation Organization
IFR	Instrument Flight Rules
IHT	In House Training
ILS	Instrument Landing System
IMC	Instrument Meteorological Conditions
ISMS	Integrated Safety Management System
KIAS	Knots Indicated Air Speed
KLM	Koninklijke Luchtvaart Maatschappij – Royal Dutch Airlines
KLM FA	KLM Flight Academy
KTS	Knots [Airspeed]
LGA	LaGuardia Airport
LNAV	Lateral Navigation
LOC-I	Loss Of Control In-flight
LOE	Line Oriented Evaluation
LOFT	Line Oriented Flight Training
LOSA	Line Oriented Safety Audit
MCP	Mode Control Panel
MSL	Mean Sea Level
NASA	National Aeronautics and Space Administration
ND	Navigation Display
NLR	Netherlands Aerospace Centre
NPV	Net Present Value
NTSB	National Transportation Safety Board [USA]
PBT	Performance Based Training
PF	Pilot Flying
PFD	Primary Flight Display
PIC	Pilot In Command
PM	Pilot Monitoring
QRH	Quick Reference Handbook
RA	Resolution Advisory
RNAV	Area Navigation

ROC	Relax – Observe - Confirm
ROI	Return on Investment
S&S	Startle & Surprise
SA	Situation Awareness
SD	Standard Deviation
SDL	Self-Directed Learning
SESAR	Single European Sky ATM Research
SHAPE	Self, Human, Aircraft, Procedures, Environment [KLM behavioural marker scheme]
SPDBRK	Speedbrake
TC	Type Conversion
TCAS	Traffic Collision Alerting System
TEM	Threat and Error Management
TOGA	Take Off Go Around
TRE	Type Rating Examiner
TRI	Type Rating Instructor
UK CAA	United Kingdom Civil Aviation Authority
UPRT	Upset Prevention and Recovery Training
URP	Unload, Roll, Power (technique)
VMC	Visual Meteorological Conditions
VMO	Maximum Operating Speed
VNAV	Vertical Navigation

This page is intentionally left blank.

1 Introduction

While commercial aviation experiences currently a high safety record, it has been observed that pilots may react inappropriately to unexpected in-flight situations leading-up to the destabilisation of the aircraft. Startle and Surprise reactions played a key role in a significant number of Loss-of-Control In-flight events (LOC-I) as well as in other types of accidents. The self-control, confidence, leadership and Crew Resource Management (CRM) skills demonstrated during events such as Qantas flight QF32 suggest that there are effective methods to manage surprise and startle effect, individually and as a crew, which is a pre-requisite for successfully managing the situation at hand.

Following the recommendations of accident investigations the EU regulatory framework for Commercial Air Transport (CAT) now covers the need to address the effect of surprise and startle within training programmes, namely as part of Crew Resource Management (CRM), Evidence Based Training (EBT) and Upset Prevention and Recovery training (UPRT). However, there is limited knowledge on how to effectively train CAT Pilots to manage startle and surprise effect, as well as their consequences on emotional and cognitive processes.

The EASA funded research project aims to produce a set of comprehensive and practical guidance materials in order to support the development of training programmes for Commercial-Air-Transport (CAT) pilots, to establish a quantitative framework to measure the impact of such reaction and the recovery process during training sessions (e.g. reaction time, recovery of situational awareness) as well as to measure the effectiveness of the proposed training scheme.

Chapter 2 describes the theoretical background of Startle and Surprise;

Chapter 3 identifies examples of the manifestation of startle and surprise reactions from reported flight safety events;

Chapter 4 provides an analysis of existing training provisions and methods;

Chapter 5 describes the training technique developed to counteract the negative consequences of the Startle and Surprise responses;

Chapter 6 provides the results from the simulator experiments at the KLM training centre using a B737-800 simulator and a B747-400 simulator. The experiments mainly took place in May and June of 2016 with voluntary crews from KLM;

Chapter 7 provides estimates for costs incurred for training against estimated benefits obtained in offsetting operational risks.

2 Literature Analysis

2.1 Structure

While commercial aviation experiences currently a high safety record, it has been observed that pilots may react inappropriately to unexpected in-flight situations leading-up to the destabilisation of the aircraft. Startle reaction played a key role in a significant number of Loss-of-Control In-flight events (LOC-I) as well as in other types of accidents. The self-control, confidence, leadership and CRM skills demonstrated during events such as Qantas flight QF32 suggest that there are effective methods to manage surprise and startle effect, individually and as a crew, which is a pre-requisite for successfully managing the situation at hand.
This EASA funded project aims at developing a set of comprehensive and practical guidance materials to support the development of training programs for CAT Pilots addressing surprise and startle effect, as well as to establish a quantitative framework to measure the impact of such reaction and the recovery process during training sessions (e.g. reaction time, recovery of situational awareness) and to measure effectiveness of the proposed training scheme. In order to provide a scientific basis for this training scheme a literature review has been performed.

This review of current scientific literature on acute stress, surprise and startle is used to establish the key aspects that play a role in preparing pilots to handle unexpected events. The focus of the review, and the meta-analysis of relevant literature, is intended to identify the emotional and cognitive aspects that require mitigation in the form of training or guidance materials for commercial pilots.
In order to identify the competencies that are particularly relevant to the management of the effects of startle and surprise, the literature review will cover the scientific background of the emotional and cognitive effects of startle and surprise. These general effects will be transposed to the commercial air transport pilot competencies as defined by the ICAO framework for Evidence-Based Training (EBT). The potential impact of an unexpected event on the flight crew can be summarised from the scientific literature of startle and surprise effects, and thereby the effect on the competencies and sub-elements thereof.
Startle and Surprise involve a multitude of psychological processes such as attention, situation assessment and awareness, self-efficacy, stress, fear, problem solving and decision making. Even without the addition of the aviation context, including crew cooperation, these processes require a book each to be described in total. This review describes all of the above processes in an aviation context and is not intended to be complete. The main goal this review serves is to provide a foundation for the following, operationally focused, work packages: current CAT training review, accident &incident analysis, training design, training experiment and a cost-benefit analysis.

2.2 Startle and Surprise

2.2.1 The Theory of Startle and Surprise

By definition, unexpected events and/or intense stimuli always cause surprise and/or startle. However, not all unexpected events lead to large physiological and emotional reactions. An Air Traffic Controller asking for a slightly different airspeed than expected on approach is a surprise but, hopefully, will not lead to large physiological and emotional reactions. The focus of this project is startle and surprise reactions that are large enough to have an impact on performance and can negatively influence safety. Reactions may be large due to their high level of unexpectedness or because they take place in a safety critical and, at least perceived, time critical environment.
It is clear that startle and surprise play important roles in many aircraft incidents and accidents (see the accident and incident analysis in (NLR-CR-2016-620). However, definitions of startle and surprise vary and are even used interchangeably within the domain of aviation. This review starts with providing insight into these differences, mainly based on a comprehensive review by Rivera, Talone, Boesser, Jentsch, & Yeh, 2014.

2.2.2 Startle

The startle reflex is the first response to a sudden, intense stimulus. It triggers an involuntary physiological reflex, such as blinking of the eyes, an increased heart rate and an increased tension of the muscles. The latter are necessary to prepare the body for the fight-flight response (Koch, 1999). The startle response is accompanied by an emotional component which for a large part influences how a person responds to the unexpected event (Lang, Bradley, & Cuthbert, 1990).
The duration of the startle reflex, as with most reflexes, is very short and depends on the severity of the reflex. A mild reflex lasts less than one second and a high-intensity response can last up to 1.5 seconds. Startle reflexes are more severe during very low or very high arousal levels. In addition to the involuntary physiological reflexes, startle inhibits the muscular activity, thus a startled person stops doing what he was doing (Koch, 1999). The disruption can last from 100ms to 3 seconds for simple tasks and up to 10 seconds for more complex motor tasks (Rivera et al, 2014).
On the flight deck the disruption caused by the startle reflex can have detrimental effects, particularly when the startle is elicited when the pilot is performing flight essential tasks. A pilot can lose part of the situational awareness, due to distraction which might cause cognitive tunneling. And pilots might be interrupted in a difficult cognitive process, such as making a decision (Rivera, et al, 2014).

2.2.3 Anatomy of the startle

When a person perceives a startling stimulus, the brain processes this in two ways (LeDoux, 1997); the ‘quick and dirty’ and the neocortex pathways. In the quick and dirty pathway, the perceived stimulus (see orange arrow in Figure 2-1) is first crudely analysed by a brain structure called the thalamus. This information is almost immediately sent to the amygdala (top white arrow) which will trigger a first response to the possible danger (bottom white arrows towards the body). In the neocortex pathway, the thalamus simultaneously sends a signal to the neocortex where it is further analysed (grey arrows in Figure 2-2. This path is slower than the 'quick and dirty' path directly from the thalamus to the amygdala, but processes the information deeper. The result from this analysis can be either that the threat is dangerous and that immediate action is indeed necessary, or that there is a false alarm and no actions are required.

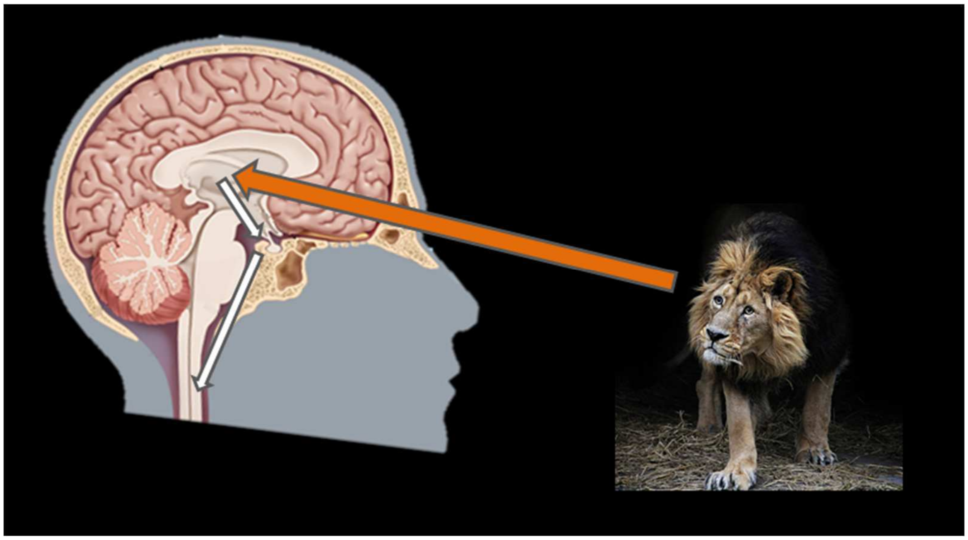


Figure 2-1 'Quick and dirty' pathway

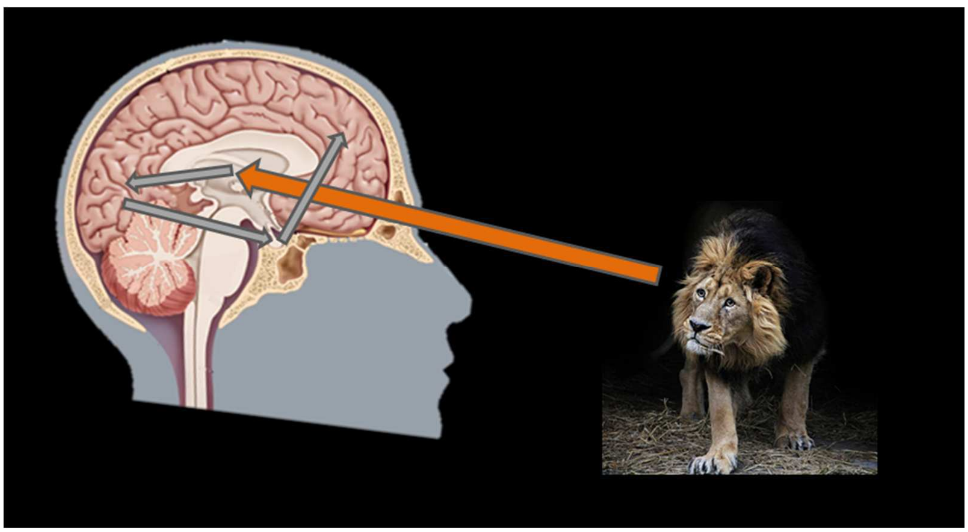


Figure 2-2 Neocortex pathway

From an evolutionary, survival perspective it makes sense to have an over-cautious, 'better safe than sorry' system in place when dealing with unexpectedness; an automatic and fast threat attribution to a Startle or Surprise. Afterwards, a slower and deliberate analysis of the situation takes place, possibly leading to a 'false threat' assessment. By then, all the physiological (increased heart rate, muscle tension and breathing, adrenaline secretion, etc.) and psychological (fear, possibly leading to anxiety and uncertainty) responses are taking place. If enough information is available to make this assessment rapidly, these responses fade away. However, in an unclear or ambiguous situation these high levels of physiological and psychological stress can persist. We assume this is what the flight crew of AF447 experienced, possibly leading to non-deliberate muscle activity (applying back pressure without being aware) and decreased cognitive capacity for situation assessment (not realizing the aircraft was in a stalled state). The slower and deliberate analysis of the situation sometimes takes place hours after the unexpected event. This happens when the fight or flight response is strong and creates a sense of urgency to take action, perceived time pressure. This action-mode inhibits slow and deliberate analysis.

2.2.4 Surprise

The psychology of surprise is about how people respond to unexpected events (Wickens, 2001). Surprise results from a disparity between a person's expectations and what is actually perceived (Horstmann, 2006). This implies that surprise can be elicited by the presence, but also by the absence of stimuli (e.g. Rivera et al. (2014); Bürki-Cohen (2010)). This contrasts with startle, because startle is always triggered by a sudden highly intensive stimulus and cannot be triggered by the absence of a stimulus.

The effects of surprise are in part comparable to those of startle. Physiological responses to surprise include increased heart rate and blood pressure, cognitive responses include confusion and loss of situational awareness, and may involve the inability to remember the current operating procedures (Rivera et al. (2014)). Even though startle and surprise often occur together, the startle reflex can be triggered without the notion of surprise. For example under anticipated circumstances when a person is told that a loud noise will be audible and when, this person will usually still have a startle reflex resulting from the loud noise (Ekman, Friesen, & Simons, 1985).

The duration of the surprise response is typically longer than that of the startle reflex. The discrepancy between expected and actual circumstances requires the person experiencing the surprise to reevaluate the situation to continue with the task. Larger discrepancies usually require more time for reevaluation than smaller discrepancies. Furthermore, the surprise also takes more time when the discrepancy requires an update of the expectations of the person experiencing the surprise (Horstmann, 2006).

The physiological response to surprise causes the attentional system to become more focused and impairs the working memory (Martin et al, 2012). The focus can help in evaluating the situation, especially when this is a dangerous situation in which you have to make choices quickly (Sapolsky, 1994). However, people tend to focus on the most salient information, which may not be the most important information at that moment (Rivera et al, 2014). Also, the combination of focused attention with the impaired working memory can cause problems for the person experiencing the surprise regarding his main tasks.

2.2.5 Startle versus Surprise

In contrast to startle, which always occurs as a response to the presence of a sudden, high-intensity stimulus, surprise can be elicited by an unexpected stimulus or by the unexpected absence of a stimulus. For a startle on the flight deck to occur according this definition a loud noise, a flash of light or a substantial displacement of the aircraft, or pilot seat is required.

Unexpected events can cause startle, they can cause surprise or they can cause startle and surprise combined. Only startle (e.g. you expected the balloon to explode and it does) is very rare in an aviation context and usually only creates very short term physiological reactions reaction. Startle combined with surprise (e.g. a lightning strike) is more common and only surprise (Expectations ≠ Reality) is the most common. The focus of the current project is on these last two, in a flight deck situation: startle combined with surprise and surprise only.

2.3 Factors/features of surprising events in flight operations

Surprising and unexpected events in airline operations are not rare or unusual and seldom result in negative events according to Kochan, Breiter and Jentsch (2005). Everyday events that occur during the flight can be surprising for the pilots. In their study, they researched which factors are involved when an unexpected event can negatively influence the outcome of the flight. They compounded a database of incidents and accident reports from different sources. These were all read to determine which factors, from a list of 35 factors that are thought to be involved in surprising events, were described in the accident or incident report. Table 2-1 displays the factors that were most frequently involved (n = 638).

Table 2-1 Frequency and percentages of most frequently occurring factors (Kochan, Breiter, & Jentsch, Surprise and unexpectedness in flying: Factors and features, 2005)

Factor	Frequency	Percent
Aircraft Position	420	65.8
Air Traffic Control	326	51.1
Other Crewmember Actions	270	42.3
Aircraft State	202	31.7
System Status	123	19.3
Automation	95	14.9
Inflight Turbulence	74	11.6
Low Visibility	64	10.0
Delays	62	9.7
Airport Construction	60	9.4
Other Aircraft – En-route	60	9.4

Furthermore, the researchers determined to what extent the unexpected event contributed to a negative outcome of the flight. The event was rated inconsequential when there was no effect on the outcome, consequential when there was an effect and when the event made the outcome worse, it was rated “exacerbated”. Table 2-2 displays the results of the events in comparison with the flight outcome.

Table 2-2 Event outcome vs flight outcome (Kochan, Breiter, & Jentsch, Surprise and unexpectedness in flying: Factors and features, 2005)

Result of the event	Normal	Event	Incident	Accident
Inconsequential	8.2	10.5	21.5	9.2
Consequential	21.3	34	18.5	18.4
Exacerbated	70.5	55.5	60	72.4

Over half the events exacerbated the outcome of the flight, but also 70.5% exacerbated the outcome in a normal flight. This means that surprise worsened the situation at hand, but that the flight continued normally.

This study revealed that there is a relationship between the involvement of an unexpected event and the severity of the outcome of the flight. In 72.4% of the accidents reviewed for this study, the involvement of surprise or unexpectedness did exacerbate the situation. On the other hand, the surprising or unexpected event was found to be inconsequential in only 9.2% of the accidents. As can be seen from Table 2-2 that in all 'outcome of flight' categories the surprising event was more likely to exacerbate the situation than not.

Therefore, regardless of the ultimate outcome of the flight, surprise very often has a worsening effect on the situation. From a systems approach point of view this is an interesting finding. Its seems Surprise can have a negative impact on the dynamic pilot – aircraft interactions, leading to a situation in which autonomic surprise reactions lead to new unexpected events, causing even more surprise. Kochan, Breiter and Jentsch (2005) conclude that this suggests that surprise may be occurring more frequently in normal flight and having relatively little to no consequences for the outcome of the flight. Therefore, most surprises may not be reported.

Even though it was a thorough and maybe even brave attempt at quantifying the causes and effects of Surprise, the sobering final conclusion of the above mentioned study from 2005 was that: “Perhaps the most important finding from this study is that potentially any factor or combination of factors can create a surprising or unexpected event that leads to an unwanted outcome.” If almost everything, at almost any time can cause a surprise, it is not sensible to limit surprise-training to specific flight events. This will not inoculate pilots against the new unexpected situation. An important aspect of training for unexpected events is its applicability beyond the specific event trained.

The intensity of the felt surprise increases with the degree of unexpectedness and the perceived, safety or personal (loss of face) impact (see 2.3.3 and 2.3.4). Because the top five surprising factors in Table 2-1 occur quite frequently, it could be argued that most of the events related to these factors are surprising but not completely unexpected. These events are referred to as “unexpected” versus “truly surprising” by Wickens (2009). Taleb (2007) has called them gray swans and black swans respectively. Black swans potentially cause the most severe surprise effects but are by definition rare. Grey swans have less severe effects but occur more frequently. Therefore both types of events have a potential safety impact and deserve attention during training.

2.3.1 Adaptive Expertise

Kochan (2005) has researched how domain expertise and judgment abilities influence how people deal with surprises. Domain expertise is defined as the knowledge that a person has about the subject at hand. Judgment ability is defined as the skill a person has in using his knowledge to predict the consequences of an event. The conclusions from this research indicated that “high levels of domain expertise did not significantly improve or deter one’s overall ability to respond and handle the unexpected aerodynamic events presented in an Inflight Simulator Learjet. Most surprising, were the significant influences of high levels of good judgment on dealing with unexpected events, even when domain expertise was low.”

If judgment skills are really domain non-specific, then people with good judgment skills, who may not yet be experts in their field, may have distinct advantages when dealing with uncertain, ill-defined, and unexpected situations. Furthermore, those who are already domain experts can benefit from enhanced judgment skills in unexpected situations by thwarting cognitive rigidity and the overuse of intuitive decision strategies.

2.3.2 Flight Deck and Systems design

It would be possible that some flight deck systems play a role in creating more Surprise or Startle. For example, an unexpected stick shaker activation has often been described stressful and/or startling. These and other auditory warning systems have been designed with a focus on providing a highly salient cue because pilots are required to act immediately or it is important that they are aware of a certain state. Whether these cues can trigger Startle or Surprise effects if presented during low or high arousal has not been investigated. Another design aspect that can have an impact on the occurrence of Startle and Surprise is system opacity. Highly automated systems which provide no or limited feedback and transfer control to the pilot unexpectedly can possibly cause surprise.

2.4 Individual differences

2.4.1 Hyperstartlers

Some people suffer more from a startling stimulus than others and the reactions of people are different when confronted with the same startle but under differing circumstances (Martin, Murray, Bates, 2012). For example, when people are tired, when their levels of arousal are heightened or when the startling stimulus was not anticipated their response is more intense (Martin & Murray, 2013). There are even people who fall in the category of “hyperstartlers”. These people show more severe reactions to a startling stimulus and the latency of the startle is longer (Simons, 1997). Ideally, pilots are less easily startled than other persons, but an experiment performed by Martin, Murray and Bates (2012) shows a significant variation in responses when pilots are confronted with a startling event. This suggests that even though some pilots are less easily startled, others respond normally and others are even more easily startled than others.

2.4.2 Personnel selection

It is not very likely that hyperstartlers aspire to a flying career and if they do that they will successfully pass though selection and training processes. Current psychological pilot selection usually involves tests focused at stress resistance. The goal of the current EASA Startle & Surprise research is focused on training solutions. It seems likely this is the area in which the biggest gains are available, decreasing the negative effects of Startle and Surprise in flight operations. However, some psychometric tests might be useful in identifying individuals who might benefit from a specific or extra training intervention. Whether this is more effective compared to instructors being educated to identify potential performance problems due to Startle and Surprise effects, can only be determined with additional research efforts.

2.4.3 Fear-potentiated Startle or Surprise

When a person perceives a startling stimulus, he first appraises the threat level of the stimulus. If the stimulus is perceived as being threatening or dangerous the response can elicit a full stress response, known as a “fear-potentiated” startle (Martin & Murray, 2013). However, if the stimulus is perceived as being non-threatening, the body quickly resumes homeostasis.

When a person has negative emotional memories related to a stimulus, this person is more likely to classify this stimulus as threatening compared to a person who has positive memories related to this stimulus. For example, when a pilot has quickly mastered upset recovery training and feels confident in his ability to deal with potential stalls, he may appraise a stall warning alert as non-threatening which therefore does not prompt a significant startle reflex and therefore does not interrupt his cognitive abilities. However, when a pilot has had bad experiences with upset recovery training his cognitive processing may be disrupted when he experiences a sudden stall warning alert, triggering a fear-potentiated startle/surprise. This pilot will need extra training to relieve him of his negative association. Instructors play an important role here. They should be aware of the possible effects of a negative experience during simulator training.

2.5 Learning from Surprises

Surprise is regarded by many theorists as an emotion that serves important adaptive functions, including learning. Little research has been conducted on the question of what people learn from surprises. One of the few studies, performed by Baruch Fischoff (1975) provides evidence that human beings have a strong bias in hindsight toward judging unexpected events as less surprising than they actually were. The knowledge that they gain after the unexpected occurrences of an event leads them to underestimate what they have to learn from the surprise. It could be argued that if this holds true for the general public, professionals and experts, such as pilots, are even more vulnerable to this bias. For a professional, well trained pilot working in a proceduralised, safety focused environment it might feel as a loss of face to admit, even to oneself, that a real surprising event has occurred. The chances that this event will be reflected upon are small and with that the chances that learning will take place. Even though being surprised is usually not caused by making a mistake or a lack of knowledge, to the professional it might be perceived as such. Only in an organisation where learning from mistakes is actively promoted this could be prevented.
CRM training plays an important role here. Loss of face and the related aspect that influences this emotion, the authority gradient, are subjects that should be addressed. Pilots should be prepared for the idea that it (Startle and Surprise) will happen to them, both as an individual and as a team, and what it will feel like and what the possible effects are. As mentioned before, Startle and Surprise reactions and emotions should be addressed as normal human reactions to abnormal events.

2.6 Startle and Surprise effects on the ICAO Competencies

In the above sections many certain and possible effects on human behaviour, cognition and performance have been described. Besides the immediate effect of distraction, most of the other effects are related to stress. The first possible effect from startle and surprise is attentional tunneling causing a distraction. If attention is not required anywhere else this has no operational effect. If however, attention is required for the execution of a safety critical task this obviously has operational and safety effects. All ICAO competencies (Table 2-3) could be affected by distraction, from the application of procedures to manual control, from technical to non-technical.
The second possible effect from startle and surprise is an increased stress level. A significant amount of research has shown undesirable effects of stress on various facets of cognition. Narrowed attention, decreased search behaviour, longer reaction time to peripheral cues, decreased vigilance, degraded problem solving,

performance rigidity and degraded working memory function are just some of the cognitive impairments noted under the effect of stress according to Martin et al. (2015).
Besides this wide variety of cognitive impairments, increased stress levels and perceived time pressure also potentially affect all competencies. However, because mostly the higher order cognitive functions are impaired, the competencies that would be effected the most would be Situation Awareness and Problem Solving and Decision Making. These are also the competencies that have the greatest effect on operational safety. Leadership and Teamwork, Workload Management and Communication could be involved in recovering from Surprise and Startle effects. The ICAO competencies have been integrated into SHAPE 2.0, the KLM behavioural marker scheme that will be used to assess the training intervention taking place in this project.

Table 2-3 ICAO Competencies

COMPETENCY	COMPETENCY DESCRIPTION
Application of Procedures	Identifies and applies procedures in accordance with published operating instructions and applicable regulations, using the appropriate knowledge.
Communication	Demonstrates effective oral, non-verbal and written communications, in normal and non-normal situations.
Aircraft Flight Path Management, automation	Controls the aircraft flight path through automation, including appropriate use of flight management system(s) and guidance.
Aircraft Flight Path Management, manual control	Controls the aircraft flight path through manual flight, including appropriate use of flight management system(s) and flight guidance systems.
Leadership and Teamwork	Demonstrates effective leadership and team working.
Problem Solving and Decision Making	Accurately identifies risks and resolves problems. Uses the appropriate decision-making processes.
Situation Awareness	Perceives and comprehends all of the relevant information available and anticipates what could happen that may affect the operation.
Workload Management	Manages available resources efficiently to prioritize and perform tasks in a timely manner under all circumstances.

2.7 Training Interventions

Startling events such as a lightning strike are not uncommon but compared to surprising events (Expectations ≠ Reality) they occur less frequently. Accident and incident data also does not provide evidence that startling events cause have more serious consequences. The main focus of the training interventions should therefore be on surprising events, less on startling events. However, as mentioned before, the effects of surprise are in part comparable to those of startle. The training interventions aimed at recovery from surprising events will also target the longer term (more than 3 seconds) physiological and psychological effects of startle. A training solution focused at inhibiting involuntary physiological startle reflexes, such as eye blinking and muscles contractions, is unfeasible due to their autonomic nature. If these reflexes are recognised and corrected timely, the risk they pose to flight safety is limited. This should therefore also be a focus point for a training intervention

2.7.1 Training Paradox

The training of Startle and Surprise Effect Management involves a paradox. Because these psychophysiological constructs (startle and surprise) involve a multitude of other physiological and psychological processes it is difficult to pinpoint an area/competencies for training intervention(s). And yet, all these separate processes are potential areas for training intervention(s) and as can be seen in Figure 3-3, this provides many opportunities to do so. For example, interventions could target the Script Discrepancy Check, trying to positively influence the appraisal of an unexpected event; threat or not? The State of Surprise could be targeted, limiting the cognitive, affective and behavioural effects. Another area for an intervention could be Self-Efficacy, increasing this could limit the chances or effects of a fear-potentiated startle/surprise. Kochan (2005) developed this process model of surprise based on research by Meyer, Reisenzein, and Schützwohl (1997) that could be used to pinpoint where and at which level training interventions can be most effective and feasible.

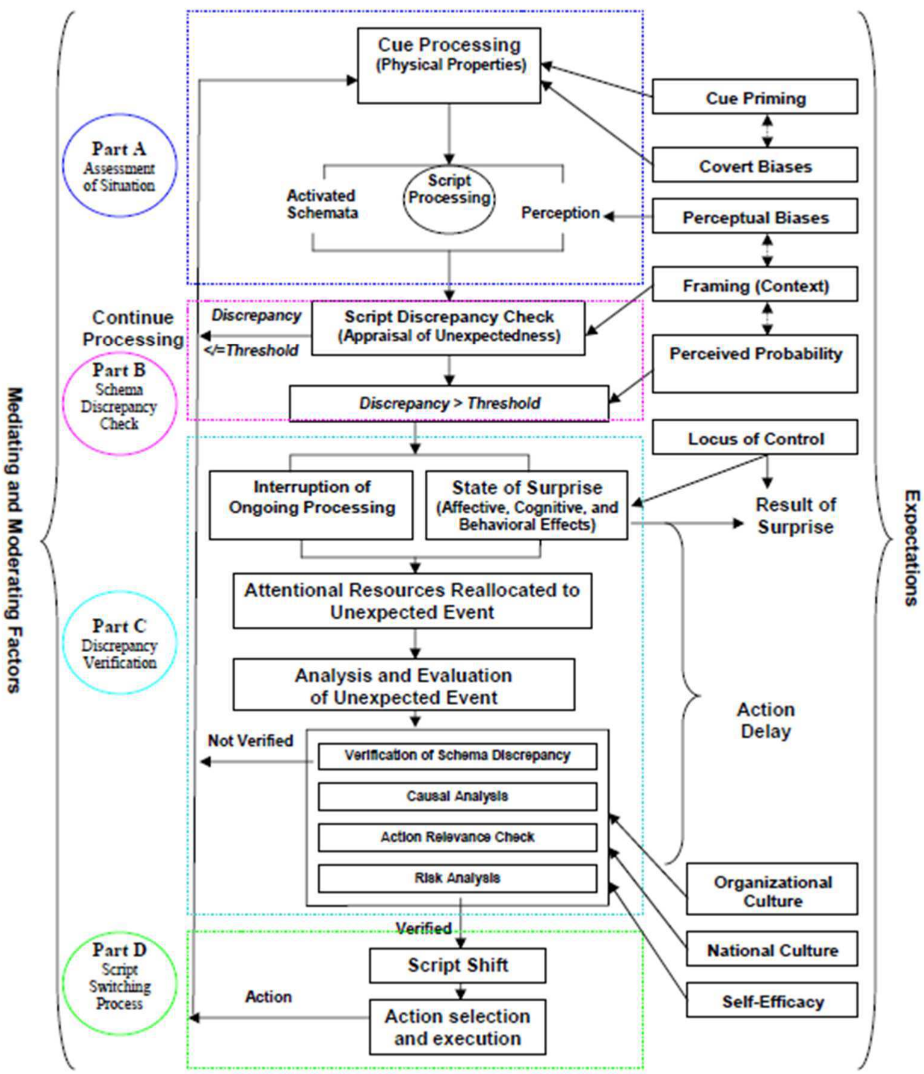


Figure 3-3 Process model of surprise

2.7.2 Training Adaptive Expertise

As was mentioned in 2.2.1, Kochan (2005) has researched how domain expertise and judgment abilities influence how people deal with surprises. People with good judgment skills, may have distinct advantages when dealing with uncertain, ill-defined, and unexpected situations. Furthermore, those who are already domain experts can benefit from enhanced judgment skills in unexpected situations by thwarting cognitive rigidity and the overuse of intuitive decision strategies. This has implications for both initial training and recurrent training, targeting both inexperienced (low domain expertise) and experienced (high domain expertise) pilots.

Training interventions might therefore focus on developing pilots into adaptive experts, i.e. more emphasis on the development of judgment skills compared to the development of domain knowledge. One trait differentiating adaptive experts from routine experts (those with expertise in one area or lacking in judgment) is the ability of the former to perform well outside of their comfort zone. They are able to deal with ambiguity and understand how their current beliefs and assumptions may affect their understanding of a situation. One way of doing this is by means of Metacognitive training: the ability to monitor one's current level of understanding and decide when it is and when it is not adequate. In other words, it is the awareness of one’s knowledge and this is a skill which can be used to control and manipulate cognitive processes (Metcalfe & Shimamura, 1994).
The long-held belief that judgment skills develop "naturally" with an increase in domain expertise does not appear to be supported in research, or more importantly, in accident and incident data (Kochan, 2004). While there is believed to be a moderate correlation of domain expertise and judgment with time, the data from this study suggest that this does not occur automatically for all people. This difference might be explained by implicit learning. Implicit or informal learning mostly takes place at the individual level. It has been found that two environmental factors, lack of time and proximity to colleagues, inhibited pilots from informal learning (Corns, 2014). The same research also identified five personal characteristics to motivate pilot’s informal learning activities: initiative, self-efficacy, love of learning, interest in profession and professionalism. The environmental factors and some of the personal characteristics might be potential candidates for longer term ‘training interventions’.

2.7.3 Key Training Objectives

Two focus points of training interventions will be ‘managing distraction’ and ‘managing stress in an unclear/ambiguous situation’. Kochan (2004) reported: “the final interesting point from the review of actual accidents and incidents was the behavioural responses which arose from the unexpected events. On a broad scale, regardless of the nature of the event or outcome severity, pilots either (a) focused on the unexpected situation, addressed the condition, and returned to pre-event duties or (b) focused on the unexpected situation and fixated on an aspect, without returning to the ongoing activities in a timely manner.” Training pilots to manage distraction and stress in an unclear/ambiguous situation seems an effective way to ascertain most pilots will fall into category a. If this is not done, even pilots with good judgment skills might not be able to use these because of their stressed state of mind and perceived time pressure leading to an action mode instead of a thinking mode. A technique that is already widely used in aviation, Threat and Error Management (TEM), could be used to create flight crew awareness about the threats, i.e. distraction, and the errors, fixation without returning to ongoing activities, of startle and surprise.

2.7.4 Individual differences and Team aspects

It is clear from the literature (see 2.3.1) that individual differences are substantial when it comes to reacting to and dealing with the effects of startle and surprise. Personal characteristics, previous experiences and self-efficacy all play a role. Ideally, instructors are able to determine which characteristics require attention during training.

The training intervention should not only improve individual performance but also crew performance. At least a part of this intervention will be focused on the crew as a team. Even if the individual performance of both crew members is degraded, working together effectively will increase overall performance. Individual differences are a potential threat. If one crew member is cognitively impaired and perceives time pressure without the other one being aware, this could seriously affect cooperation and shared situational awareness. However, individual differences could also be used as a mitigating instrument. If a particular event startles or surprises only one crew member, the other one could coach his colleague in managing the effects. It cannot be expected that a majority of pilots develop the necessary knowledge and skills automatically and/or informally.

2.7.5 Training objectives and scenario elements

In previous chapters some possible training objectives have already been mentioned: learning from mistakes, including both grey and black swans and preventing or mitigating previous negative training experiences. Some other are provided here but this is not an exhaustive or conclusive list in this project phase.

Kochan (2005) described some implications for training: "First, judgment skills, although put forth to be domain non-specific, are not necessarily domain independent." As the results of this study indicated, there is a significant interaction between domain expertise and the judgment skills. Therefore, attempts to train judgment skills might best be situated in the context of the specific domain. The topics and methods found to be successful in improving judgment skills are:

- Define, explain, measure, and discuss the concepts of metacognition, adaptive expertise, cognitive flexibility, and decision making strategies
- Train through helping to build and enhance mental models of equipment, environments, teams, and situations
- Present conceptual models of situations and tie declarative, procedural, and strategic knowledge together using judgment skills
- Integrate the above concepts into scenarios used for teaching at every opportunity
- Obtain, distribute, and reinforce examples of the use of good judgment skills in domain specific situations
- Train and evaluate using scenario-based examples and focus on the adequacy of the interaction of domain specific skills and judgment skills

Martin (2015) listed a number of specific simulator scenarios whereby the expectations of pilots are deviated from, in such a way that it creates an element of surprise or startle.

These are some examples:

Unexpected Stall Warning

- Introduce a sudden tailwind that is sufficient to induce a stall warning 30 sec after leveling at cruise altitude.
- Introduce a sudden tailwind that is sufficient to induce a stall warning at cruising altitude during a briefing or while busy with some other task.

Cargo Fire Warning with a Simultaneous Loud Bang

- Introduce a cargo fire warning (bell) with simultaneous loud bang on sim panel at 40 ft above decision altitude on a hand-flown approach requiring missed approach.
- Introduce a cargo fire warning (bell) with simultaneous loud bang on sim panel just prior to turn onto ILS (hand flown).

Sudden Tailwind at Rotate

- Introduce a 20 kt tailwind at rotate speed.

Unexpected Ground Proximity Warning System (GPWS) Warning

- In instrument meteorological conditions (IMC) introduce a GPWS warning during a briefing or while busy with some other task.

Some of the other scenarios provided in Martin's paper focused on multiple malfunctions. While this is a possibility, we consider this might put too much focus on the technical aspects of the scenario instead of the non-technical aspects, which are more important for training; managing stress in unexpected and ambiguous situations.

2.7.6 Training media

Research into the selection of training media is not within the scope of this project. However, training media can have a significant impact on training effectiveness and efficiency. Due to historical and regulatory reasons the Full Motion Simulator might seem a logical, even unavoidable, medium for Startle and Surprise Effect Management Training. In this project we will use the simulator but only as one of the possible training media that might have added value to training. For now, a high fidelity, domain specific, training device seems the most likely to deliver transfer of training. Depending on the knowledge, skills and/or attitudes that need to be developed to better deal with startle and surprise, a training medium choice could be made in the future.

2.8 Conclusions from the Literature Analysis

Even though there is a clear difference between de factors causing startle and surprise, the effects are comparable. Physiological responses to surprise include increased heart rate and blood pressure, cognitive responses include confusion and loss of situational awareness, and may involve the inability to remember the current operating procedures. On the flight deck the disruption caused by the startle reflex can have detrimental effects, particularly when the startle is elicited when the pilot is performing flight essential tasks. A pilot can lose part of the situational awareness, lose track of where he was in going through a checklist or be interrupted in a difficult cognitive process, such as making a decision.

However, not all startles and surprises have strong effects. Surprising and unexpected events in airline operations are not rare or unusual and seldom result in negative events. Everyday events that occur during the flight can be surprising for the pilots. The focus of this project is startle and surprise reactions that are large enough to have an impact on performance and can negatively influence safety. Reactions may be large due to their high level of unexpectedness or because they take place in a safety critical and, at least perceived, time critical environment.

The primary goal of this project is the development of Guidance Material for training. Because it seems that the competencies required mitigating the effects of startle and surprise are non-technical, probably some of the guidance material should be targeted at the instructors. Because they are domain experts and not so much judgment experts and/or psychologists, some guidance on how to provide this training might be necessary. Project deliverables could also be used in other areas such as recruitment, training design and CRM development. Another idea might be to include startle and surprise into Threat and Error Management. Because this is a widely used tool in aviation, this link could provide an anchor for startle and surprise management in the operational flow.

3 Startle and Surprise in Accidents and Incidents

It must be noted that, as Kochan, Breiter and Jentsch (2005) indicate, there is not a limited set of contexts which are startle or surprise prone – the phenomena are context-independent. Furthermore, startle or surprise events cannot be reduced to the onset of startle or surprise, but must take into account the crew (mis)management of the situation that follows. Crew experience, competence and other performance shaping factors also affect the outcome of the event. Therefore it must be underscored that the magnitude of the event (e.g. accidents vs. incidents vs. reportable event) does not differentiate between more or less valuable examples of startle or surprise. Moreover, to understand these two phenomena requires valuing their acute and personal nature, which can often more accurately be understood from elaborate and personal pilot reports rather than the limited CVR and FDR recordings available in high profile crew-fatal accidents. Finally, this inherent limitation in investigating most high-profile accidents is not aided by the fact that official accident and incident reporting techniques do not yet have clear, specific classifications of startle and surprise events making the search for valuable case studies a diffuse exercise of inferring these phenomena from other pre- or post-events. This is in turn conflicting with the non-contextual nature of the events.

The research group has made an attempt as expanding on the study from Kochan, Breiter and Jentsch (2005)[1] by accessing the European Co-ordination Centre for Accidents and Incidents Reporting System database (ECCAIRS) for accidents and incidents involving regular commercial operations and charter operations (FAA Part 121/125 aircraft) in the period of 2005-2016, and searching for factor descriptors possibly related to startle and surprise (e.g. “unexpected”, “unknown”, “misjudged”, “not observed”, “sudden”) These searches returned 931 accidents and incidents from the ECCAIRS database. However, the nature of these inferred-type searches returns many reports irrelevant to this study. In addition, save a few examples such as Air Asiana 214, Turkish Airlines 1951 and US Airways 1549, many reports cannot sufficiently describe the onset and progress of startle or surprise as these pilot recollections are often not available to investigators.

For this reason the research has turned to using NASA’s ASRS self-reporting database. This database hosts many pilot reports with varying degrees of seriousness, but as they are done by live pilots and often include a thorough description of their own recollection of the event, these reports provide much more insight into the acute and personal nature of startle and surprise. The outcomes of the events are not always particularly grave (e.g. go-around followed by an uneventful landing), but much in the spirit of the Safety II philosophy (Hollnagel, 2014)[2], focusing on examples of positive recovery, as they may be as valuable (if not more valuable) for developing training interventions. The NASA ASRS search is similar to the ECCAIRS search (Part 121 & 125, 2005-2016), where two independent searches were done for the search terms “startle OR startled” and “surprise OR surprised”. These searches delivered much more salient examples as the terms “startle” and “surprise” are familiar to pilots (whereas they are often not employed in official accident and incident reports). This report will address the two phenomena of startle and surprise separately. An important consideration about the ASRS database is that they are unverified pilot accounts of a situation, and are subject to bias and subjectivity. For this

---

[1] Kochan, Janeen, E. G. Breiter, and F. Jentsch. "Surprise and unexpectedness in flying: factors and features." *Proceedings of the 13th International Symposium on Aviation Psychology*. 2005.

[2] Hollnagel, E. (2014). Safety-I and Safety-II: The Past and Future of Safety Management. Farnham, UK: Ashgate.

reason the conclusions from these reports are not based solely on single instances, but rather extract conclusions from a broader set of reports, in an attempt to identify a general trend.

While the ASRS database provides a broad sweep across startle and surprise situations in the cockpit, this report will also illustrate the manifestation of startle and surprise by a more in depth analysis of several notable cases in point: Garuda Indonesia Flight 200 (2007), US Airways 1549 (2009), Loganair 6780 (2014), West Air Sweden 294 (2016) and one other case. These five cases provide an example of both inappropriate and superior startle & surprise effect management. Although there may be more cases exhibiting startle and surprise effects, the live pilot accounts of these events combined with the in-depth event reports provide useful insights on what actually transpired in these cockpits. In contrast to the more subjective accounts in the ASRS database, these official reports have been more rigorously composed according to proper investigative procedures, and can therefore be subject to more intense individual assessment.

With these investigations, this report aims to analyse startle and surprise effects in the cockpit in two dimensions; illustrating both the details of how these effects have manifested them in the cockpit, as well as providing a broader view of the causes and resultants of these startle and surprise effects throughout the sector.

3.1 Startle in the cockpit: Cases from ASRS

As introduced, the ASRS reports are a good source of personal recounts of the startle effect. The research has found 15 clear examples of startle-related cases, and analysed each case. The analysis identifies the onset of startle (high-intensity stimulus), the effect on pilot performance (startle-response) and what the pilot self-reflected on the experience. Although not all pilots describe their take-away from the event, those that do, provide interesting suggestions for possible solutions to manage the startle effect. Table 3 provides an overview of the three analysis steps of each of the 15 selected reports. Appendix A contains more case details in the form of synopses and important excerpts.

Reviewing these 15 cases, they seem to illustrate a variety of possible startle causes and effects. Observing the causes, these range from wake turbulence and vibration to verbal callouts by a warning system. Some situations are more common such as a late go around during landing, and some are rarer such as a weather balloon directly in the flight path.

Irrespective of the cause, the effects of these stimuli also vary greatly. Some pilots report following the startle with immediate and well-tuned actions (e.g. 1291766, 1230172, and 704645) others will blunter actions (e.g. 1268324, 1142116, and 795767) and some report complete cognitive freeze effects (e.g. 1268324, 1004144). Several pilots report that they would like to be better trained and prepared in future, to expect such (untimely) stimuli (e.g. 1268324, 1168197). These reflections shed light on the idea that the startle response is, similar to surprise, also dependent on (sometimes hard-wired) situational expectations (e.g. 1142116).

Looking toward solutions in managing startle, there are several suggestions from these cases. Some pilots exhibit a level of metacognition by noting their awareness of a delay in action (e.g. 1219870), illustrating this by being in the “green” or “yellow” zone (some basic indication of human performance capacity). Sometimes they then reverted to swapping flying and monitoring tasks as they felt they were “getting behind the aircraft”. These examples contrast to other cases which illustrate a frozen first officer requiring the captain to take control (e.g. 1268324). It cannot be concluded that a pilot’s startle reaction or level of metacognition directly results in safe recovery instead of an accident, but there do seem to be different methods in which crew members jointly address a startle situation, possibly improving flight crew resilience against any effect that the high-intensity stimulus might have on either pilot. Such a collaborative reaction is described more succinctly in the analysis of US Airways Flight 1549.

Table 3: Overview of 15 ASRS reports describing the startle effect

ASRS No.	Onset	Effect	Reflection
1291766	False Runway Overrun Prevention System warning (verbal callout “runway too short, brake max”)	Immediately adjusted auto brakes (per call out)	Still no understanding of the reason for the warning
1268324	Wake turbulence during touchdown, gear contacted ground violently	Immediate Go-around action (“firewall throttles”), go-around was “intense”, no TOGA mode activated	Too unprepared for very late go-around, desire for automatic pilot GA response also in worst case scenario
1268324	Unstable approach and “UNSTABLE” callout	Student frozen, inaction. Late TOGA action, no mode callouts	Train more unplanned and unscripted go-arounds, prepare automatic response
1230172	“Too low flaps” alert	Quickly adjusted speed for less flaps	Go-around would have been proper reaction, poor flaps awareness (to be improved)
1268057	“Too low flaps” warning	Immediately scanned flaps/gear/brakes	Too fixated on other tasks, missed flaps awareness
1227427	Green laser flash in the cockpit	Whole crew distracted, vision affected	None
1219870	Two terrain warnings despite visual terrain contact	Level off, took time to recover from startle	Combined with fatigue the startle had a clear and prolonged disruption on task performance
1168197	Wake turbulence during landing, intense movement	Immediate throttle up, no TOGA button (flow disruption)	Better training in late go-arounds, preparedness
1157235	Impact with large birds, engine vibrations	No immediate action (no memory items)	None
1142116	Wake turbulence, speed increase	TOGA buttons could not be found, manual override after verbal command. Rough steering, behind aircraft	Early auto-throttle disconnect (simplify), too much overreaction
1045570	Terrain warning despite visual contact with terrain	AP disconnect, but no manual action (realized margins were OK)	None
1017650	TCAS warning	Incorrectly responded (descend vs. climb), blank for several seconds	TCAS dial prone to misinterpretation, poor SA prior to TCAS

ASRS No.	Onset	Effect	Reflection
1004144	Sudden physical object (weather balloon) in front of aircraft	Physical jerk reaction and scream, induced startle in other pilot, descent mismanaged	None
795767	Master caution (smoke) during rotation	Momentary pause of rotation, reduced power	Surprised at own actions, currently poor guidance on momentary indications & messages
704645	Loud runway friction noise (assumed gear/tire failure)	Aborted take-off	Later discussion pointed to the rough runway surface conditions

3.2 Surprise in the cockpit: Cases from ASRS

Similar to analysing startle, the ASRS database has also been sourced to investigate the onset of the surprise effect in the cockpit. The research has found 13 clear examples of surprise-related cases, and analysed each case. The analysis identifies the onset of surprise (cognitive mismatch), the effect on pilot performance, and what the pilot self-reflected on the experience. Similar to the startle analysis, pilot reflections provide interesting suggestions for possible solutions to manage the surprise effect. Table 4 provides an overview of the three analysis steps of each of the 13 selected reports. Appendix B contains more case details in the form of synopses and important excerpts.

Reviewing these 13 examples of surprise in the cockpit, they seem to describe a wide variety of instances. Several cases featured automation-related surprise, in particular when the automation exhibits action or inaction incongruent with the crew's mental expectation of the functioning of automation (e.g. 1030005, 897215, and 881125). In some instances the crew managed to understand (post hoc) what occurred, in others it remained unclear (e.g. 1104105, 1174038, and 990058). When surprised, most crews exhibited a form of manual control reversion which varied in intensity. Nonetheless some reactions were too acute (e.g. 1137763, 845610) while others too late (e.g. 1030005, 1104105). As the proper reaction to such surprising events is quite dependent on the context, it cannot be generalized to a basic mandate for or against manual control reversion.

The amount of reflection provided in these cases is less than for startling events, yet the available evaluations often point to unclear functions, annunciations and limited instrumentation which otherwise may have prevented crews from not anticipating a particular automated event. Surprise is the phenomenon of experiencing a cognitive mismatch with all the human performance fallout thereof. Although cockpit design evolutions will not alone resolve the surprise problem being researched, it does direct a possible solution trajectory toward improved knowledge (trained, communicated or displayed) as a barrier against possible cognitive mismatches.

Table 4: Overview of 13 ASRS reports describing the surprise effect

ASRS No.	Onset	Effect	Reflection
1223274	Unexpected low fuel state (caused by centre tank disabled via wing fuel interface)	Current tasks (altimeter setting) were obfuscated, task saturation with other menial issues	None
1174038	Inadvertent approach mode activation caused a sudden unexpected acceleration during descent	Immediate action to revert to Selected Mode continued descent	Confusing mode switching, poorly documented, should be clarified
1137763	Suddenly clear of clouds, runway not as expected, further aggravated by an EGPWS warning	Abrupt PF action (pull back), without power = very low airspeed	Too abrupt action in GPWS escape manoeuvre
1104105	Aircraft suddenly dived and accelerated	FO attempts automation intervention, captain Reacts with manual control (pull up, disengage)	Surprised by automation, puzzled by FO automation reaction
1032254	Untimely ATC call	PF disables automation and assumes manual control, PF frozen (unresponsive)	None
1030005	Uncommanded high pitch high power turn	Reselected alt on MCP, no effect. AP disconnected, A/THR still high power, late A/THR disconnect	None
1028073	Aircraft did not descend as programmed	Disengaged A/THR and used SPDBRK	Bothered about all precautions and dutiful flying, still gross error. Limited FMA change annunciation
990058	Speed decay during glideslope capture	Go around and second approach better prepared for behaviour	Rapidly changed from "mind-numbing" standard to "something never seen before"
964358	Realized incorrect flap settings during take-off run	Delayed reaction (several seconds), continued take-off	Hard-coded habits lead to surprises in alternative configurations/situations
897215	Sudden uncommanded turn to the right during arrival	Recovered and landed	Likely preventable with glass cockpit
881125	Uncommanded right turn and pitch up, losing airspeed	Jump seat captain first to call airspeed (delay reaction)	Later discussion explained behaviour

ASRS No.	Onset	Effect	Reflection
845610	Surprised by approach situation, compounded by poor visual contact	Fell into can-do mentality	Would have been better to go-around and re-ascertain situation
973820	Small aircraft suddenly appeared during night taxi	Immediate braking, pilots distracted, boggled, missed taxiway exit, lost big picture	Better to take a minute to recover instead of pushing on

3.3 Cases in point

In addition to the ASRS incidents, there are five cases which warrant special attention in this report: Garuda Indonesia Flight 200 (2007), US Airways 1549 (2009), Loganair 6780 (2014), West Air Sweden 294 (2016) and one other case. These cases provide examples of inappropriate and of superior startle & surprise effect management. These cases in point will be examined in more detail than the ASRS reports in the previous sections.

3.3.1 Garuda Indonesia Flight 200 (2007)

Synopsis

> *"On 7 March 2007, a Boeing Company 737-497 aircraft, registered PK-GZC, was being operated by Garuda Indonesia on an instrument flight rules (IFR), scheduled passenger service, as flight number GA200 from Soekarno-Hatta Airport, Jakarta to Adi Sucipto Airport, Yogyakarta. There were two pilots, five flight attendants, and 133 passengers on board. The pilot in command (PIC) was the pilot flying, and the first officer was the support/monitoring pilot. During approach into Yogyakarta the aircraft was in an unstabilized approach condition with the speed varying between 229 and 244 knots, pitch varying between 3.5 degrees up and 3.8 degrees down, and the rate of descent reached 3,520 feet per minute. After a late touchdown at an airspeed in excess of 200 knots the aircraft overran the departure end of runway 09, to the right of the centreline at 110 knots. The aircraft hit an embankment and was destroyed by the impact forces and an intense, fuel-fed, post-impact fire. There were 119 survivors. One flight attendant and 20 passengers were fatally injured. One flight attendant and 11 passengers were seriously injured."*

This accident was in essence an unstabilised approach and landing which in turn was the result of a loss of control on behalf of both pilots. The Pilot in Command (PIC) was also the Pilot Flying (PF) in this case, and had succumbed to a cognitive freeze rendering him unable to respond to alerts from both the aircraft and the co-pilot. The first officer, in this case Pilot Monitoring (PM), repeatedly attempted to re-engage the PIC, but also failed to take control when his attempts to re-engage failed.

> *"His [PIC/PF] attention was fixated or channelized on landing the aircraft on the runway and he either did not hear, or disregarded the GPWS alerts and warnings and calls from the co-pilot [PM] to go around. The co-pilot did not follow company procedures and take control of the aircraft from the PIC when he saw that the PIC repeatedly ignored the GPWS alerts and warnings."*

Indications of startle & surprise

This case is relevant to startle and surprise effect management in two ways. The first and foremost way is related to the PIC's apparent cognitive tunnelling after he is confronted with the fact that he is flying the approach too high. It seems that the PIC continued the approach in accordance with a classic plan-continuation-error. This has been reported as an effect in several ASRS cases (both startle- as well as surprise-related), such as cases 1268324, 1032254, 1030005 & 845610. Although these cases explicitly indicate plan-continuation-error, several other cases also show how many crews react with immediate action, instead of careful consideration (provided time is available).

> *"The PIC subsequently assessed the situation by calculating the altitude and the remaining distance to the runway, and decided that the flight profile was not as he had expected. Eleven seconds after expressing concern about the wind, the PIC said 'Target enam koma enam ILS, kagak dapat dong' (the target is 6 point 6 ILS, we will not reach it). The PIC then attempted to trade off excess airspeed and lose height, but*

only succeeded in flying a flight path that was erratic in pitch, causing the airspeed and altitude to vary considerably. The PIC flew an unstabilized approach… <> …The PIC again expressed concern that the vertical flight path was not proceeding normally, when at 23:56:49 he commented 'Wah nggak beres nih' (oh there is something not right). Between 23:56:49 and 23:57:20 the aircraft was in an unstabilized approach condition with the speed varying between 229 and 244 knots, pitch varying between 3.5 degrees up and 3.8 degrees down, and the rate of descent reached 3,520 feet per minute at 23:57:20."

The second instance is the first officer failing to engage as well and assume control when the PIC clearly did not respond. Although this can be argued as a CRM shortfall, it is quite possible that the first officer was himself so surprised by the PIC's lack of response to clear alerts, that the first officer also succumbed to the same tunnelling effect.

"The first officer did not attempt to take control of the aircraft from the PIC and execute a go around, in accordance with company instructions that require taking over when an unsafe condition exists."

The report continues to detail the PIC's supposed mental state during the approach. The report concludes that the PIC suffered from a "fixation" effect, and provides several pieces of evidence to support this analysis.

"Absorption: A state of being so focused on a specific task that other tasks are disregarded.

Fixation: A state of being locked onto one task, or one view of a situation, even as evidence accumulates that attention is necessary elsewhere, or that the particular view is incorrect.

The 'tunnelling or channelizing' that can occur during stressful situations, which is an example of fixation. Note: The term 'fixation' has been chosen to describe the PIC's state of alertness, which provides a clearer idea of 'being locked onto one task', than 'absorption'. Several 'findings' support this 'tunnelling or channelized' condition, for example:

a. The PIC's attention became fixated on landing the aircraft. The concept of fixation is reinforced because he asked the co-pilot a number of times to select flaps 15 and asked if the landing checklist had been completed.

b. The PIC did not respond to the 15 GPWS alerts and warnings and the two calls from the pilot monitoring to go-around. The PIC did not change his plan to land the aircraft, although the aircraft being in unstabilized condition. The other tasks that needed his attention were either not heard or disregarded. The auditory information about other important things did not reach his conscious awareness.

c. The PIC said 'The target is 6.6 ILS, we will not reach it'. The PIC flew an unstabilized approach. He also realized the abnormal situation when he commented 'Wah, nggak beres nih!' ('Oh, there is something not right'). So, the PIC's intention to continue to land the aircraft, from an excessively high and fast approach, was a sign that his attention was channelized during a stressful time.

d. The PIC also asked several times for the co-pilot to select flaps 15. During interview he said to investigator that 'his goal was to reach the runway and to avoid severe damage'. He 'heard, but did not listen to the other voice (GPWS), and flaps 15 and speed 205 was enough to land'. The PIC experienced a heightened sense of urgency, and was motivated to escape from what he perceived to be a looming catastrophe, being too high to reach the runway (09 threshold). He fixated on an escape route, 'which seem most obvious', aiming to get the aircraft on the ground by making a steep descent. His decision was flawed, and in choosing the landing option rather than the go around, fixated on a dangerous option.

e. The PIC was probably emotionally aroused, because his conscious awareness moved from the relaxed mode 'singing' to the heightened stressfulness of the desire to reach the runway by making an excessively steep and fast, unstabilized approach."

In addition to the cognitive (and emotional) effects on the flight deck during this approach, the investigation also indicated that the crew in question was not properly trained to act on EGPWS warnings. Although this would not likely have prevented the tunnelling to occur, it may have provided an escape via recognition-primed decision making related to the aural warning, which in this accident was clearly not appreciated by the crew.

> *The Garuda Boeing 737 simulator training did not include training or proficiency checks in the vital actions and responses to be taken in the event of GPWS or EGPWS alerts and warnings, such as 'TOO LOW TERRAIN' AND 'WHOOP, WHOOP PULL UP'.*

Despite a reported lapse in training EGPWS events, cognitive tunnelling can itself also be sufficient to induce inattentional deafness, a phenomenon in which auditory signals are not consciously processed due to the pilot's mental workload being saturated. A similar phenomenon is inattentional blindness, in which visual signals are not consciously processed. Such deafness may also have contributed to a lack of action from the PIC after the EGPWS warnings.

Main conclusion

This accident case in point provides a vivid account of how a relatively "mild" cognitive mismatch (approach planning) can manifest in a surprise and a degradation of cognitive performance on the flight deck, ultimately indicated by the inability to process new information such as aural alerts and crew communication. This becomes a dangerous spiral into the proverbial "rabbit hole", disabling a functioning crew and resulting in the crash of a mechanically sound aircraft. The CRM and EGPWS practices were sub-standard and were not effective enough to break the hold that surprise had on the crew. In this situation in particular, CRM-like intervention from the first officer and the EGPWS alerts both only presented themselves near the end of the approach, during a highly progressed state of tunnelling. It is hypothesised that earlier CRM-founded interventions, focused at startle and surprise awareness, or standardised stable-approach checks might have been effective to recover the crew from an early, less-progressed state of surprise and tunnelling.

3.3.2 US Airways Flight 1549 (2009)

Synopsis

"On January 15, 2009, about 1527 eastern standard time (EST),1 US Airways flight 1549, an Airbus Industrie A320-214, N106US, experienced an almost total loss of thrust in both engines after encountering a flock of birds and was subsequently ditched on the Hudson River about 8.5 miles from LaGuardia Airport (LGA), New York City, New York. The flight was en route to Charlotte Douglas International Airport (CLT), Charlotte, North Carolina, and had departed LGA about 2 minutes before the in-flight event occurred. The 150 passengers, including a lap-held child, and 5 crewmembers evacuated the airplane via the forward and overwing exits. One flight attendant and four passengers received serious injuries, and the airplane was substantially damaged. The scheduled, domestic passenger flight was operating under the provisions of 14 Code of Federal Regulations (CFR) Part 121 on an instrument flight rules flight plan. Visual meteorological conditions prevailed at the time of the accident."

This accident is a now-famous case of flight crew response to an acute, severe deterioration of the flight situation. The crew's effective response to the situation have been thoroughly praised in the aviation world, but do provide a stark contrast to Garuda Indonesia Flight 200, in which the flight situation was nominal, with no additional challenges forced upon the crew.

Indications of startle & surprise

The NTSB investigation documentation provides in depth records of the interviews with the flight crew. These first-hand flight crew accounts of the accident provide insight into the startle and surprise effects present during the flight. First the recount of the first officer (FO) will be investigated, followed by the captain recount. The first officer recalled several large birds just before the bird strike, which allowed him to understand the cause of the sounds, although the aftermath would be more difficult to predict. In addition to a causal hint, the FO was also familiar with compressor stalls from his DC-9 experience.

"The Airbus was climbing and he saw the birds go down below the windscreen and thought they would miss the birds. Then he heard birds hit: "BOOM, BOOM, BOOM". He then heard compressor stalls and the engines quit… …He used to experience compressor stalls when they put the DC-9 in reverse so he was familiar with the sound."

The first officer noted that the first reaction came from the captain, who took control of the aircraft from the first officer. He then committed himself to the dual engine failure checklist in the QRH while the captain maintained control of the flight path. The FO noted that there was discussion before choosing the river, and agreed that both LGA and TEB airfields were no options, both pilots committed to the same plan at this point.

"The Captain ("Sully") said "my aircraft" and directed the FO to perform the dual engine failure procedures in the QRH… …There was some talk about returning to LGA but it was far away at that point they were coming down fast. TEB did not look viable to Mr. Skiles and the only other option was straight ahead down river."

From this point the first officer recalls that both crew members were working on different parts of the same plan. When asked about crew coordination, the FO noted that the captain's proficiency and his own experience contributed to the positive outcome.

"He said his total flight time was about 20,000 hours and that he may have been the most experienced first officer in history… …When asked to compare the captain's proficiency to other pilots he had flown with, he said it was extremely high ... He could not think of an area in which the captain could improve. When asked if he heard anyone ever complain about flying with the captain, he said he never read the captain's name on the "bathroom wall"."

Ironically, the FO indicated that training was less of a positive factor in this accident, despite that he was just out of training (this was his second flight on the Airbus). The training provided him with the knowledge to act acutely, yet both pilots were not prepared for a low altitude loss of complete thrust, and their experience (underlying to their effective collaboration) seemed to be the most important success factor.

"Dual engine failure was an exception to the ECAM so he went to the QRH and never read the ECAM. There were procedures in which they did not follow the ECAM. Dual engine failure follows the checklist."
"One of the training spots was dual engine failure, but it was performed at altitude with plenty of time to do procedures… …The current training philosophy was to take your time to try to assess the situation. It did not apply in the actual accident… The only training for ditching was given in ground school… … so the training did not help in the actual emergency."

The captain's account of the accident was somewhat similar, but with some subtle differences. First and foremost, the captain was able to recall many more vivid sensations at the time of the bird strike, indicating how the unexpected event engaged multiple senses.

"He noticed a large flock of large birds filling the windscreen and there was no time to react. He felt and heard birds colliding with the aircraft. He heard and felt vibration. The engines had stalls. He felt yaw. He smelled cooked bird and the smell came through the cockpit and cabin. He felt an immediate and dramatic loss of thrust simultaneously. There was some yawing as thrust reduction was not even."

The captain also explicitly stated an experience of surprise and shock.
"He was surprised at how symmetrical the loss of thrust was… …When asked if he noted the engine instruments, he said it was so shocking and he was so focused on maintaining aircraft control."

This surprise quickly led him to assume control of the aircraft and ensured that the aircraft, although having become a glider effectively, maintain sufficient airspeed and the best glide ratio to maximize time available given the altitude.

"When asked about taking over the controls, he said "my aircraft" and FO Skiles said "your aircraft". Captain Sullenberger took control of the airplane and dropped the nose to obtain green dot speed because they had slowed below it. He said he then turned on the ignition."

Immediately after these initial reactions, the two crew members took up different roles, but were both working at different ends of the same problem. Similar to the FO, the captain was very appreciative of the CRM in the cockpit.

"He was happy that it was clear that they understood each other and worked well together without the need for a lot of words… …When asked to describe the crew coordination, he said it was amazingly good considering how suddenly it occurred, how severe it was, and the little time they had. He said FO Skiles was a good pilot and it was amazing how he handled it. He said if he had not been told it was the FO's first trip after OE he would not have known."

When asked about training, the captain indicated that although this specific scenario was not trained, the more universal competencies trained were of great value in this situation. This contrasts somewhat with the more unappreciative view that the FO had on training.

> *"absolutely training has helped… …fundamental values to maintain aircraft control, manage the situation, and land as soon as the situation permits He also said the basic rules of airmanship and CRM helped."*

Main conclusion

The recounts from the pilots share several important facets. Firstly, both definitely experienced a startle from the bird strike, and had to overcome the surprise (even disbelief) of the crippling of the aircraft. The second shared recount is the immediate *collaborative* decision making in task division, and the subsequent capacity to work diligently at different parts of the same problem. Both pilots were on the same page, and there were no misunderstandings in their cooperation. Lastly, both crew members were extremely experienced, familiar with compressor stall sensations, with the FO even fresh out of training for the aircraft type. Reviewing these conclusions, it becomes clear that the crew factor, i.e. the combined level of experience in technical matters and crew coordination, can be highly valuable in addressing startling and surprising situations. In particular, the aspect of engaging each other and build shared awareness helps refocus the attention from past events toward future tasks. The resulting harmony between the crew members allowed for a much more effective and efficient response to the emergency (especially in this case of extreme time pressure).

3.3.3 Loganair Flight 6780 (2014)

Synopsis

"The Saab 2000 was inbound to land on Runway 27 at Sumburgh (Shetland Islands) when the pilots discontinued the approach because of weather to the west of the airport. As the aircraft established on a southerly heading, it was struck by lightning. When the commander made nose-up pitch inputs the aircraft did not respond as he expected. After reaching 4,000 ft AMSL the aircraft pitched to a minimum of 19° nose down and exceeded the applicable maximum operating speed (VMO) by 80 kt, with a peak descent rate of 9,500 ft/min. The aircraft started to climb after reaching a minimum height of 1,100 ft above sea level. Recorded data showed that the autopilot had remained engaged, contrary to the pilots' understanding, and the pilots' nose-up pitch inputs were countered by the autopilot pitch trim function, which made a nose-down pitch trim input in order to regain the selected altitude. The crew decided to return to Aberdeen, climbed the aircraft to about FL240, the aircraft flying between FL237 and FL245, and performed a safe landing at Aberdeen about 35 minutes later."

This case provides an interesting insight into a startle-triggered situation which severely impacted the crew's ability to maintain cognitive control over the aircraft's (admittedly complex) autopilot system. Nevertheless, the rapid action response from the PIC further accelerated the cognitive tunnelling the crew succumbed to, as it led to surprise (mismatch between expected pitch control and autopilot behaviour).

Indications of startle & surprise

Unfortunately this Saab 2000 was fitted with a 30-minute CVR which only captured the taxi-in after landing. Anecdotes from crew experienced are based on crew interviews and their recollection of the event. The interview points out that the PIC may have already been under a higher level of stress before the startle event:

"There were indications of an increase in stress evident in the commander's actions before the lightning strike. Although he was pilot flying, he made the transmission announcing the intended manoeuvre away from the localizer himself, when the normal operating procedure would have been for the co-pilot, as pilot monitoring, to do this."

"The commander was making a radio transmission to ATC about his intentions at the time, but when the lightning struck, he uttered an expletive and stopped transmitting."

These excerpts illustrate a situation in which the PIC's stressed condition was elevated further by the lightning strike, and clearly led to a startle reaction.

"The commander's actions following the lightning strike were to make manual inputs on the flying controls, which appear to have been instinctive and may have been based on his assumption that the autopilot would disconnect. However, the autopilot did not disconnect and was attempting to maintain a target altitude of 2,000 ft AMSL by trimming nose-down while the commander was making nose-up pitch inputs. The control forces felt by the commander were higher than normal because the autopilot was opposing his inputs and he may have attributed this to a flight control malfunction caused by the lightning strike. He did not recall having seen or heard any of the aural or visual mis-trim cautions which were a cue that the autopilot was still engaged. This was probably the result of cognitive tunnelling."

The AIB report addresses the concept of "Automation Surprise" and how this also played a role in the events following the lighting strike. Combined with the high stress state, the PIC likely became cognitively tunnelled and did not address the control issue appropriately. Despite the poor cognitive ergonomics of the autopilot (e.g. not disengaging when manual control is applied), it is not a difficult system to control or take control from, which further reinforced the argument of cognitive tunnelling.

> *"Automation surprise can occur if the autopilot does not behave as expected, for example if the system remains engaged when the flight crew believes it is not. Clear feedback of the system's status can help to prevent this inconsistency. Stress, which might be experienced in the moments after a lightning strike, leads to an increase in physiological arousal. This may lead to 'cognitive tunnelling', in which individuals exhibit a tendency to focus on a small number of the most salient or expected information, and only information that supports the prevailing understanding of the situation may be processed. Cognitive tunnelling not only affects perception of visual signals, it can also affect auditory processing at times of high cognitive load; this is 'inattentional deafness'. Clear and prominent status indicators can assist."*

Furthermore, in this particular case the investigation noted how the combination of cognitive tunnelling and manual flight path control can lead to pilots "instinctively" focusing on flight path control and ignoring other tasks until control has been regained. This echoes the familiar and sound reasoning of Aviate – Navigate – Communicate, but fails when manual control is the first reaction any crew has to a situation (in particular when the flight path is not abnormal).

> *"Pilots who are experiencing difficulty simply in achieving the desired flight path may accord lower priorities to non-handling tasks; at least until the flight path is under control. Their ability to seek and process information, and then analyse it and diagnose the root cause of their difficulty, may be impaired. There is uncertainty over the effectiveness of simulator training in altering the fundamental behaviour exhibited by pilots under such stress. Previous regulatory action, which required modification of other aircraft types to address similarly confusing and stressful conditions, is consistent with this conclusion."*

Main conclusion

This is an interesting case in which the severity of the accident was not defined by the cause of the startle (in this case the lightning strike) but in the sequence of events after this. In effect, after the lightning strike, the aircraft was fully functional and a simple autopilot-disengage would have been sufficient for the pilots to manoeuvre the aircraft in any way they would like. However, the effects of the startle, likely coupled with the pre-startle stress, reduced the PIC's cognitive frame of mind to make immediate manual inputs, ignoring other control modes. Of course the alternative hypothesis is that the PIC (thinking that the autopilot had disengaged due to the lightning strike) may have assumed that his manual control system was impaired, and instigated his tunnelling in that direction. Unfortunately, if the pilots had refrained from an instant manual reaction, it may have been possible that the secondary problem of fighting the autopilot would be prevented altogether, and led to a much safer flight.

3.3.4 West Air Sweden Flight 294 (2016)

Synopsis

"The flight (Canadair CRJ200) was uneventful until the start of the event, which occurred during the approach briefing in level flight at FL 330. The event started at 00:19:20 hrs during darkness without moonlight, clouds or turbulence. The lack of external visual references meant that the pilots were totally dependent on their instruments which, inter alia consisted of three independent attitude indicators.
According to recorded data and simulations a very fast increase in pitch was displayed on the left attitude indicator. The pilot in command, who was the pilot flying and seated in the left seat exclaimed a strong expression. The autopilot disconnected automatically, accompanied by a "cavalry charge" aural warning. Both elevators moved towards nose down and nose down stabilizer trim was gradually activated from the left control wheel trim switch. The aircraft started to descend, the angle of attack and G-loads became negative. Both pilots exclaimed strong expressions and the co-pilot said "come up".
About 13 seconds after the start of the event the crew was presented with two contradictory attitude indicators with red chevrons pointing in opposite directions. At the same time none of the instruments displayed any comparator caution due to the PFDs declutter function in unusual attitude.
Bank angle warnings were heard and the maximum operating speed and Mach number were exceeded 17 seconds after the start of the event, which activated the overspeed warning. The aircraft collided with the ground one minute and twenty seconds after the initial height loss. The two pilots were fatally injured and the aircraft was destroyed."

Indications of startle & surprise

Due to the fact that the pilots both perished in this accident, there are only several brief CVR recordings and investigation interpretations which provide insight into the presence of startle and/or surprise. The table below is an excerpt from the investigative report, providing insight into the CVR recordings the first 15 seconds of the event.

Time UTC	Pilot In Command	First Officer	ATC	Remarks, sounds and warnings
23.19.22	What (!)			
23.19.23				Continuous Cavalry Charge
23.19.24				Single Chime
23.19.28				Irregular sound
23.19.29				Triple Chime
23.19.29		What (!)		
23.19.30	What (!)			
23.19.30				SV: Engine oil
23.19.31				Warning: Stabilizer trim clacker
23.19.33		Come up		
23.19.33				SV: Bank angle
23.19.35	Come on, help me, help me, help me			
23.19.35		Turn right		
23.19.35				SV: Bank angle
23.19.36		What		
23.19.37				Warning: Overspeed (Clacker)

The initial “what!” exclamations clearly reflect a sense of disbelief and cognitive mismatch between the left side PFD and the expected (and actual) flight profile. Such a mismatch can be designated as a surprise event.

> *“The malfunction occurred when the crew was performing the approach briefing, which meant that attention was divided between two simultaneous tasks. This probably contributed to the surprise effect…*
> *…Although there are risk factors related to fatigue present, for instance working during night hours, the Swedish Accident Investigation Authority in this event regard the cognitive-emotional surprise effect that the pilots were subjected to, to outweigh the possible state of fatigue.”*

The other two attitude indicators were displaying the correct attitude (which was straight and level), so there was the possibility of the crew being aware of the mismatch. However, the PIC (assumed to be pilot-flying) had little time to properly assess the situation. There was a very brief pitch-mismatch indication after T1 (1 second after the initial pitch increased) and lasted until T4. After 3 seconds the autopilot automatically disconnected, and after 4 seconds the pitch-mismatch annunciation was removed from the PF by an HMI declutter feature. This decluttering introduced to “PITCH DOWN” red chevrons when the pitch attitude increased above 30 degrees nose up (around T4).

At this point the PIC, pilot flying, had to assume manual control within 3 seconds, and was given very strong indications from his PFD that he had to pitch down. Similar to Loganair 6780, a cognitive tunnelling focused on manual control (a basic pilot instinct) moved the crew to act with immediacy, without verifying the need for such actions. This has shown to be a dangerous combination with startle and/or surprise events. In particular, the PIC was faced with very little outside visual cues, and was more than otherwise reliant on his instruments.
Third and lastly, the initial “required” action was a strong pitch down movement:

> *“Approximately at the same time, DFDR-data indicate that both left and right elevators moved to a position that causes the aircraft to pitch down. The aircraft started to descend with vertical acceleration values momentarily reaching negative values of -1G.”*

This rapid negative-G movement can itself induce a physiological startle effect. Negative-G’s are more potent than positive-G’s due to their unnaturalness in our regular day to day experiences. If this manoeuvre introduced a follow-up startle effect, the crew may have become victim to a promulgated state of cognitive impairment, further reducing their ability to recover.

Main conclusion

This case is a recent example of how startle and surprise can be causal to a serious accident. The combination with immediate manual control magnifies the effects that startle and surprise can have on the safety of the flight, and it may be so that such reactions may require attenuation as a buffer against incorrect actions based on incorrect perceptions. A second conclusion is that the redundancy of information (in this case three independent attitude indicators), designed to facilitate cross checking of base aircraft flight parameters, was not effective in recovering the crew’s situational awareness before they took to action. Hence, a solution to surprise management does not stop with the provision of information; crews must consciously provide themselves with the opportunity to pay attention to that information before they determine their reactions.

3.3.5 Cracked Window Case

Synopsis

The crew of a Boeing 737 was engaged in a conversation, when suddenly a sound resembling a 'tick' was heard, described by the crew as similar to 'a pen falling on the floor'. Subsequently the crew observed a number of cracks in the left hand windshield. Shortly thereafter the left FWD WINDOW OVERHEAT light, together with the Master Caution light, illuminated. Due to the general appearance of the cracked window, the crew was convinced that the cracks implied an immediate threat to the safety of the flight making an immediate descent necessary. Subsequently the captain addressed the situation to ATC and requested a descent to FL200. During this VHF transmission, the selected altitude was changed to FL200 and the descent was initiated. After explaining the urgency for the descent to ATC, ATC instructed the crew to turn 30 degrees to the right, because of opposite traffic. On the new heading, the flight was cleared to descent to FL200. Shortly after the descent initiation, the crew donned their oxygen masks, established crew communication and the First Officer took over the RT. During the descent, the crew requested a further descent to FL100. ATC granted this request. After the read back of this clearance, the FO made an urgency call to ATC, by calling out "PAN, PAN" 3 times. According to the crew the QRH procedure "Window damage - Forward" had been consulted but discarded again as the flight was already descending. During the descent, the crew decided to divert. After contacting the diversion airfield, the crew cancelled the PAN-call. The diversion was otherwise uneventful. After landing, inspection by an AMT revealed that only the outer pane was damaged. No cabin pressurization problems were encountered throughout the flight.

This case provides a clear case of the effects of a startle-induced stress reaction at a moment with low workload, fine flying conditions and a fit, relaxed crew. It is in particular interesting to note the rapid escalation of this situation, which was unwarranted from a technical standpoint.

Indications of startle & surprise

Both crew members indicated during the interviews that they had been startled by the shattering of the windshield. As a result of this startle effect, both crewmembers wanted to descend and land as soon as possible. However, due to this initial startle/surprise and the consequent actions, the use of the Non-Normal checklist, as mentioned in the QRH checklist instructions, was postponed until after the initial start of the rapid descent and related actions.

In contrast with other startle triggers, this case of a cracked window has an additional contingent effect which is called *fear potentiated startle*. In essence, the cognitive recognition of a potential threat that such a busted window represents makes the initial physiological startle reaction evolve into a full stress reaction.

The perceived fear was different for the two crewmembers. During the interviews, the captain stated that he was afraid that the window would fail and collapse inwards. The FO was also anxious about the possibility of a total window failure, however in his view the window would be pushed outwards by the cabin pressure and he therefore decided to increase the IAS to counteract the force exerted on the window. Neither crewmember had expressed his specific fear to the other crewmember. Nor had the crew adopted its strategy on how to execute the rapid descent. Although the descent was initiated without mutual consult, the descent itself was coordinated with ATC and conducted in a safe manner.

Main conclusion

After inspection at the diversion airfield it turned out that the outer pane was cracked, yet the fail-safe principle of the window is such that the structural integrity of the airframe is still guaranteed with this kind of damage. According to an UK CAA date review over a same 5-year period, approximately 170 windshield failures occurred on FAR Part 25 airplanes. In all events, the structural integrity was always maintained.

Had the crew been familiar with the design of the window, it would be conceivable that such knowledge would present as a barrier against the contingent stress response, as the crack would not have been categorized as an immediate threat. The startle effect of the crack may still have occurred, but the exacerbation of the crew's stressed state would not. In this case the flight was diverted safely, yet the inherent risks of an emergency descent could have been avoided by taking more time to analyse the situation, the risk and potential options.

3.4 Conclusions from the Accident Analysis

The incidents and accidents reviewed in this report provide an insight into the causes, manifestation and outcomes of startle and surprise effects in the cockpit. Due to the nature of incomplete reporting, biases in retrieving true startle or surprise cases and various understandings of what startle and surprise are, the report cannot make any direct quantitative claim to the prevalence of startle and/or surprise effects.

However, upon reviewing the incident and accident cases in this report, it becomes clear that startle and surprise have a multitude of causes (triggers), and are equally varying in the outcome: there is no direct, unambiguous connection between the onset of startle and/or surprise effects, and the outcome of the situation. However, incidents such as West Air Sweden Flight 294 show how sinister startle and surprise can be when it manifests itself in the cockpit.
Hence, the accident analysis makes three conclusions:

1. Startle and surprise do not always lead to a serious incident or worse. However, they have exhibited the potential to do so. For this reason it poses a real threat to aviation safety.

2. Startle and surprise do not have a clear set of defined causes. The effects can manifest themselves in many different situations by means of many different triggers.

3. Startle and surprise can perpetuate subsequent startles and surprises due to effects such as cognitive tunnelling and fear-potentiated reactions. These have the potential snowball into a greater loss of cognitive and physical control of the situation.

Based on these three conclusions, it seems that appropriate startle and surprise effect management must not focus on particular causes or situations, but rather provide a universally relevant solution irrespective of the actual trigger or onset of startle or surprise effects.

Referring to the pilot reflections from the ASRS reports as well as the sub-conclusions from the five accident reports, several comments may be made concerning possible solution directions for startle and surprise management:

A. With respect to startle and surprise management, it seems that inter-personal interaction (possibly embedding within CRM behaviours) can help to overcome the cognitive toll that startle and surprise takes on the individual pilot.

B. Because startle and surprise can snowball quickly, it is imperative that resolution strategies are able to stop this snowballing as early as possible. In particular, a level of restraint in crew reaction may play a role to halting this perpetuation.

C. With respect to surprise management, it seems that a level of operational knowledge (whether trained or presented at the moment) can help prevent surprise (as less situations are unforeseen), but also provide more cognitive handles to rebuild awareness of the situation (e.g. by improved heuristic analyses, effects analysis and mental simulation).

4 Analysis of Airline Startle & Surprise Training

4.1 Training at KLM

4.1.1 Ab-initio (KLM Flight Academy)

Although supply of new pilots is in principle not restricted to one party, KLM Flight Academy (KLM FA) is preferred supplier and therefore this syllabus will be elaborated upon. Other civilian or military parties such as the Royal Dutch Air force will use different syllabi.

The following KLM FA syllabus items concerning, or related to startle and surprise have been identified:

- Theory:
  - Incident/accident cases are discussed concerning Threat and Error management (TEM), however not specifically focussed on startle and surprise.
  - Theoretical background in Human Performance Limitations course concerning the physiological and psychological effects of stress.
- Flying training:
  - Upset Recovery training sessions at Airline Training Center Arizona (ATCA) flying school (formerly Aviation Performance Solutions - APS), containing briefing time, a total of 3:30 flying time in a Grob 120 aircraft (stalls, spins, spiral dives, nose high attitudes, nose low attitudes, slipping turns, skidding turns) and debriefing time
  - Multi Crew Course: Multi crew related accident and incident cases are being discussed during briefing and TEM is addressed in debriefing. This is related to, but not specifically focussed on startle and surprise.

4.1.2 Type Conversion (KLM)

Type Conversion (TC) at KLM typically covers 3-days of Computer Based Training (CBT), followed by 14 to 16 simulator sessions (depending on previous experience) and an exam. The amount of focus on difference training between types of aircraft depends on previous type experience of the instructor.

- Each TC is divided into 3 modules:
  - <u>Basic flying skills</u> (incl. stalls, upset recovery): simulator sessions where candidates fly the entire session without Autopilot, Autothrottle and Flight Director. Focus is on pitch and thrust as the basic parameters that will control the aircraft's flight path. With these basic flying skills the recognition, avoidance and recovery of a loss of control situation/upset should be increased. The purpose of this is, first, to enhance basic flying skills, and second, to increase self-efficacy in case automated systems fail or otherwise are unavailable. During briefing and debriefing, focus

is on TEM regarding raw data flying (the case of Turkish Airlines Flight 1951 is used). Startle and Surprise is therefore handled in the basic flying skills module in a preventive way and by means of providing recovery techniques (as an element of the upset/loss of control training). Effort is made to confront crews with an actual experience of surprise.

- System & Failure Management (procedure handling & automation): simulator sessions where the focus is on normal system handling and the procedural part of failures, not on management in an operational setting. The goal of this module is to integrate the knowledge of systems in practice, internalise Standard Operating Procedures and train the handling of Non-Normal Procedures. Also, automatic flight is introduced in this module. During briefing and debriefing, focus is on automation (including automation dependency and automation surprise), situation awareness and the relationship between the two. TEM knowledge and application is expanded to automation and procedure handling. In the system & failure management module, when considering startle and surprise, the training can be considered preventive in nature, as focus is on providing knowledge and procedure handling regarding systems management. The element of actual surprise caused by problems with systems management is hardly addressed as almost all exercises are known beforehand.
- LOFT (Line Oriented Flight Training) scenarios (program not known to candidates): simulator sessions where the acquired skills and knowledge from the previous modules have to be demonstrated in an operational environment which comprises the handling of unexpected and unforeseen events. Focus is on a realistic line oriented flight in the more demanding environments of the routes flown by the aircraft type. During briefing and debriefing the focus is on CRM (Crew Resource Management): communication (with crew, passengers, company and other relevant parties), Information acquisition and building crew situational awareness, Workload Management (task division, time management, ETTO (Efficiency Thoroughness Trade Off principle). TEM knowledge and application is expanded to an operational environment including the threat of disruptions and distractions. The LOFT scenarios of the TC course put the different elements of the training into an integrated operational scenario, which also includes handling the above aspects of the handling of startle and surprise in the operational environment.

4.1.3 CRM courses (KLM)

Two types of initial CRM courses have to be completed in a pilot's career at KLM with a half-yearly refresher of CRM items in the recurrent training (see below). The first initial training is provided when coming into service and another during the command course, when First Officers approach command position:

- Crew Management Course (CMC) 1, initial CRM training, addresses TEM, automation dependency, stress management, fatigue and vigilance, communication, information acquisition and processing, situation awareness, workload management, safety culture, decision making and subjects brought up during these interactive sessions. Several theoretical models about these subjects are provided to pilots,

together with possible operational measures. Subjects are presented in a facilitating style where the participant's involvement and experiences are part of the learning process.
- Crew Management Course 2 (Command Course) addresses the subjects stated above once more, but now focus on the captain's responsibility to motivate crewmembers to display desired CRM behaviour.

The content of these courses is related to startle and surprise because unexpected events such as malfunctions and human error are dealt with. However, specific startle and surprise training is not provided.

4.1.4 Recurrent training and checking (KLM)

Every 6 months pilots receive type recurrent training, comprising of a 2 hr briefing, 3:30 hr simulator session and 0:30 hr debriefing. In the invitation letter a homework assignment is provided. Also, every 6 months a proficiency check has to be performed. Although this is not a training event and therefore does not provide a true training environment, exposure to unexpected events and instructor's feedback do have training value concerning startle and surprise. For all of these sessions though, it is conceivable that pilots share training programs and that they will not experience true surprise. This is especially the case for the proficiency checks as the tasks in these sessions are prescribed by regulations (for each take-off each pilot can suspect either an aborted take-off or an engine failure after V1). The introduction of Line Oriented Evaluations (LOE) in 2017 will provide opportunity to create surprise in a checking environment, depending on its design (more or less unique programs or programmed lesson plans).

Concerning startle and surprise a number of things are noteworthy:
- A full simulator session concerning Loss of Control and Upset Recovery was provided in 2012. All KLM pilots received training in awareness, recognition, avoidance and recovery of airplane upsets. The startle factor had an explicit role in the training. A countermeasure to the possible debilitating effect of a startle response was trained: a recovery method focused on the simple procedure: 'unload, roll, power'.
- An Upset Avoidance & Recovery skill item is provided in each Type Recurrent: for example Type Recurrent 1-2016 provided an in depth briefing of Low Speed at High Altitude (EASA SIB 2015-07) as well as a simulator exercise with an extreme wind shear at high altitude to force the airplane into a low speed situation which was unrecoverable with adding thrust alone. In this way the skills, knowledge and attitudes formed in the Loss of Control sim session or in the Type Conversion are reinforced.
- Each Type Recurrent (every 6 months) has a CRM briefing item with alternating subjects such as: automation surprise, stress & stress management, threat & error management, information acquisition and processing and situation awareness. Every 3 years all subjects are covered. In this way the training from CMC 1 and CMC 2 is reinforced.
- The LOFT scenario in each Type Recurrent comprises the element of surprise. This element can be a complex failure, a sudden weather change, a passenger situation or other challenging situations. After flying this scenario the pilots are briefed on their situation handling including CRM aspects such as workload management, decision making, leadership and communication.
- Also, in Proficiency Checks pilots are confronted with surprise, however, probably much less than in Type Recurrent training because checking scenarios contain tasks prescribed by regulation and not much

more. Surprise only comes from the specific timing of the failures. For example, during take-off an engine will fail before or after V1, during approach the visibility will be such that go-around has to be flown at some point, and windshears are introduced somewhere in the program.

In addition to these simulator-related training sessions a 3-yearly Inflight Safety Recurrent (Joint CRM) is provided to all flight crew. In this one day classroom session effective communication between flight crew and cabin crew is one of the subjects, but also sharing points of view regarding emergency situations. Hijack and unruly passengers are recurring subjects, where actors are hired to confront crew members with the emotional effects dealing with these situations. Recovery of startle and surprise is addressed in this course.

4.1.5 Instructor training (KLM)

- Basic Instructor Training is a 3-day training course that covers the full range of didactical skills, knowledge and attitude, but also attention is given to assessing CRM and TEM (an overview of SHAPE, KLM's behavioural marker system, is provided).
- Advanced Instructor Training is a 3-day training course for instructors with more than one year instructional experience. Assessing CRM and TEM is treated more in depth with a broad discussion about SHAPE and the application in a simulator environment.
- Instructor Recurrent Training and Checking reinforces Basic Instructor Training and Advanced Instructor Training, again with attention to assessing CRM and TEM using SHAPE.

During these courses specific startle and surprise training is not provided.

4.1.6 Operational Instructions (KLM)

- In 2010 TEM was introduced in training and soon thereafter the Basic Operations Manual was adapted to embed this in normal flight procedures. Each crew briefing before departure and approach should focus on identification of possible threats, assessment of potential risks and the required measures to mitigate those threats and risks.

Organisational provisions

- Since 2010 'What if'-cases have been distributed by management to challenge crewmembers and to expand their knowledge base by discussing these cases with their colleagues. Discussion is mandatory at the yearly line check and voluntary at other moments. Also, in numerous publications discussion of these cases and self-fabricated cases have been promoted.
- The magazine 'In for Safety' covers all recent KLM-incidents and near incidents, together with in depth analysis of certain cases (also from other airlines) and stories from pilots sharing their experiences.

Effects of startle and surprise are sometimes explicitly mentioned. In this way the mental models of pilots are increased in order to better anticipate on threats and to prevent surprise.

- In 2015 the Integrated Safety Management System (ISMS) was introduced with a focus on reporting when an incident was about to happen, instead of only reporting when something did actually happen the wrong way. This approach seems to be effective as a lot more reports are coming in these days and by giving feedback it is helping crews to anticipate potentially dangerous situations and prevent surprise.
- Since 3 years the Alternative Training and Qualification Program has been developed, and is now nearing the implementation phase. The focus of this program is to improve crew performance by reinforcing the 8 core competencies (ICAO) by means of a feedback loop from simulator training, line training, inspection flights (LOSA), the safety department, and the flight technical department. To accommodate this, the behavioural marker system SHAPE is updated covering aspects such as 'Distraction Management' and 'Self-Control' (both contributing factors in startle and surprise).

4.1.7 Summary KLM

Three methods concerning startle and surprise can be discerned throughout KLM training. The first two are more general strategies and the third is more flight-skill related:

1. Crew Resource Management (CRM). In this context it is regarded as a way to prevent surprise, amongst others by stressing the importance of building crew situation awareness by effective communication, and recover from startle and surprise (by training on core competencies such as Problem Solving and Decision Making, Workload Management, Situation Awareness, Communication, Leadership & Teamwork).

2. Threat and Error Management (TEM). In this context it is regarded as a way to prevent surprise, by anticipating on possible threats, and to recover from startle and surprise, by suggesting mitigating measures in advance thereby creating a mental model which can be accessed more easily following possible startle by a threat.

3. Basic Flying Skills and Upset Prevention & Recovery training (UPRT). In this context it is regarded as a way to prevent surprise, by recognizing possible threatening airplane states, and to recover from startle and surprise, by training on the manual flying core competency thereby promoting adequate response to upsets.

4.2 Other airlines

By means of a questionnaire (Appendix A) Air France, Lufthansa, British Airways and Air New Zealand were asked to provide information about their training provisions regarding startle and surprise (March 2016). The overall picture shows a wide range of handling the subject, with two out of four airlines currently not addressing startle and surprise specific subjects in simulator training. Surprising events in simulators are, as could be expected, trained in LOFT sessions, but specific training regarding the possible reactions on startle and surprise is not offered. The other two airlines state they are already meeting new EASA requirements, but are also developing new programs.

Specifically, theoretical training in startle and surprise related subjects is provided in classroom Human Factor training programs, but still in a rudimentary form; no counter measure techniques are trained. All airlines are developing new briefings to be compliant with EASA requirements. Regarding stress management techniques such as breath control, all airlines state this is addressed in theoretical training, however these techniques are not practiced, nor are they actively promoted by instructors.

At the same time the airlines indicate that the best safety value will probably be derived from practical training. The two airlines that have already implemented startle and surprise training have done this mainly in Upset Prevention and Recovery Training. Focus is on specific events (stall, overspeed, unreliable airspeed) rather than the training of skills which can be used in any surprising event. All airlines are however struggling with the constraints of training in current Full Flight Simulators. First of all, simulators do not provide perceived potential life threatening situation, thus triggering a fight or flight response is virtually impossible. Second, crews that go to simulator training know ‘things’ will happen, thereby eliminating most of the element of surprise. Third, producing effects such as sudden high or sustained G-loading is not possible, which reduces the approximation of real life upset events where startle can be expected. Another aspect they indicate is that pilots tend to find out which events will be trained soon after a new recurrent cycle is implemented. Some airlines tackle this problem by letting the instructor decide which events they give to crews (in the Line Oriented Evaluation), but still a limited number of events can be chosen. Practical problems arise in developing a large number of suitable events.

All four airlines indicate that currently they do not address startle and surprise related subjects in instructor training. They do however see the high value of training instructors on the subject and are developing programmes. Most of those are focusing on the element of surprise in combination with UPRT training.

4.3 Preliminary Training Recommendations

Although airlines have taken training measures to prevent and recover from startle and surprise, most of this is done in general terms, with a focus on prevention in specific scenario events. A specific training concerning recovery is recommended because there are doubts about the effectiveness of airline pilot training for abnormal events (Casner et al., 2013). In this chapter we describe some a training factors that focus on controlling the emotional and behavioural effects and some instructor and organisational factors that influence the potential success of such training.

4.3.1 Ab-initio:

Startle and surprise training is considered useful at all times in a pilot's career, but the benefits can be gained most when this is presented at an early stage. Pilots receiving startle and surprise training at a later stage in their career may have to unlearn specific behaviours, while pilots in early stages of learning can be trained with less effort, integrating these skills early into their behavioural patterns. The recommendations in the following paragrpahs also apply to ab-initio training.

4.3.2 CRM courses:

As training concerning the effects of startle and surprise is currently not provided, it is recommended to provide pilots with sufficient specific knowledge such as: what is Startle, what is Surprise, what are causes of these reactions and how does it affect you? Currently, pilots are trained in the safe environments of simulators how to solve problems in numerous different scenarios. However, as far as we are aware, no training programs deal with the emotions that pilots might experience when they actually encounter such an event in real life. Explaining that those emotions are very normal human reactions to an abnormal situation might take away some of the (additional) surprise or even shame when they are experienced.

4.3.3 Type conversion, and recurrent training and checking:

As training programs still tend to practise and test abnormal events in the same way every time. For the sake of creating surprise, it is recommended to present different forms of these abnormal events (e.g. instead of presenting a decompression in the form of an explosive decompression every time, alternating with a subtle decompression or a recoverable decompression will add exposure). In this way recognition of failures will not be limited to a single mode of presentation and chances of recognition in real-life situations will increase (see 'Sources of Power', Klein, 1998)

The other aspect of training sessions is that they tend to be fixed in content and well-known by pilots for the duration of the program, thereby excluding the element of surprise. It is recommended that Type Conversion courses contain enough opportunity to train unexpected events by not publishing the full lesson plans. For other training sessions (e.g. LOE) it is recommended to give the instructor enough freedom to introduce variety.

Training specifically focused on the Pilot Monitoring is still not well developed in most airlines, while good monitoring practices can enable crews to detect deviations in an early stage, thereby reducing the chance of (automation) surprise. It is recommended that Pilot Monitoring training is developed to enhance flight crews startle and surprise prevention skills. This training should focus on Situation Awareness Level 1, perception (Endsley, 1995). This teaches pilots when to pay attention to which cues.

4.3.4 Instructor training:

To facilitate instructors in recognising, training and providing feedback on startle and surprise effects, it is recommended that they receive specific training on the subject matter. In this way not only feedback on the (negative) outcome can be provided, but also on the process of handling a surprise.

Conducive to learning in general is a strict difference between training and checking environments. In a training environment it is acceptable, even desirable, to make mistakes and to try another time; to practice more. This is especially important when startle and/or surprise is a factor. The outcome of scenarios is irrelevant when training general startle and surprise techniques. It is recommended that airlines point out the importance of this to their instructors.

4.3.5 Operational instructions:

An operational instruction regarding checking of colleagues after a startle and/or surprise can be considered in enhancing safety on the flight deck. This can be done in different ways: using a code word or term, a colleagues name or touching a colleague on the shoulder. The choices between these options can be made taking airline communication styles and (national) cultures into account.

4.3.6 Organisational provisions:

A culture promoting the exchange of experiences is useful for expanding the mental models of pilots. It is recommended that companies promote and facilitate this in various ways (e.g. promoting the mentor role of captains towards first officers, presenting and discussing cases in training and in safety publications, promoting a reporting culture including stories of incidents about to happen, etc.)

In order to get a grip on problems concerning startle and surprise a feedback system should be able to capture those, including a behavioural marker system for examiners/observers. It is recommended that companies review their reporting system regarding startle and surprise effects.

5 Development of Training Program

5.1 Overview Startle and Surprise Training

The goal of this document is to describe the 'Why' behind the choices that were made that led to the actual training design: 1,5 hours classroom and 1,5 hours simulator training. Training details are described in Chapter 3. The rationale for all different parts of the training; classroom, Unload-Roll-Power (URP) technique and simulator will be described. Except for some initial effects after a startle, the effects of Startle and Surprise are very similar. Because of this and the fact that Surprise is more prevalent in aviation, only surprise is mentioned in this text to prevent repeating the two terms over and over.

Research into literature and incident and accident data shows clearly that Surprise can have detrimental effects on cognitive abilities. These effects are partly caused by and always exacerbated by the emotions associated with Surprise. Surprise intensifies any emotion connected to the event that causes the surprise, whether positive or negative. Humans generally do not like Surprise because it creates uncertainty and is a potential threat. Well trained professionals, whether they are pilots, athletes, or military personnel, are arguably even less fond of Surprise. Admitting that you are surprised can be regarded as a loss of face, admitting that you 'do not know', or 'you were not prepared'.

Teaching pilots rational and structured decision making strategies is a sensible idea. This is being done in aviation by teaching pilots to use decision making acronyms or Basic Failure Management procedures. However, as long as humans are experiencing strong emotions such as uncertainty and fear, they are incapable of performing a complex cognitive process. The proposed and to be tested short term training intervention will therefore focus on managing these emotions and rational and structured decision making. The training intervention intends to inhibit two basic human reactions to Surprise: flight/fight behaviour and cognitive paralysis (also known as 'freeze'). Flight/fight behaviour can result in hyper vigilance. This is an enhanced state of sensory sensitivity accompanied by an exaggerated intensity of behaviours whose purpose is to detect and/or handle threats. In other words: a lot of mindless activity that is not being evaluated. Cognitive paralysis or freeze occurs when emotions overrule all cognitive processes in the face of (perceived) danger.

The training goal is to teach pilots to apply a technique that lets them manage their emotions in all surprising situations where some time is available, i.e. no immediate action is required. In most abnormal and surprising situations some time is available, with two exceptions:

- Flight path is not under control (aircraft upset, terrain and traffic warnings)
- Personal safety is at immediate risk (explosive decompression)

The surprise management technique can also be applied after these exceptional situations are under control, to manage any on-going surprise reactions. Prerequisite to the intervention training is that pilots are already trained for these situations, to a level that requires automated behaviour, including upset prevention and recovery.

5.1.1 Training Paradigms

The three paradigms that are applied in the training design are Self-Directed Learning (SDL), linking theory to trainee experiences, and Performance Based Training (PBT).

SDL is about making trainees responsible for their own learning process. For example, during the simulator training, trainees will be stimulated to indicate which part of the technique or which scenario's they would like to practice again. A self-scoring form focussing on the important parts of the URP technique will assist them in doing this. Such a self-evaluation process increases retention of the theory and transfer of the learned technique to the operational environment.

Linking theory to trainee experiences is important because it increases knowledge retention. However, for this particular training it is also used to achieve personal relevance for the trainee in relation to the URP technique. Because a simulator environment is unable to provide operational stress levels during emergencies and/or startle and surprises, trainees might not feel the need to apply the technique during simulator training. By linking operational startling and surprising experiences (including the personal physiological and emotional responses) with the URP technique this risk is reduced. This is achieved, for example, by requiring trainees to perform a homework assignment: 'write down an unexpected operational event that is remembered well, i.e. which made an impact'. During theoretical training this event and the reaction time test at the start of the classroom training will be used to provide a link between theory and personal (operational) experiences.

With PBT the overarching question to be answered is 'How well can the trainee use what he knows'? That means the focus is on using the skill in an operational context and to adjust training and/or coaching to individual needs to increase training effectiveness. This is applied during the simulator training whereby instructors assess performance on the different aspects of the technique and adjust their coaching to individual needs. This increases training effectiveness which is particularly important in short training interventions such as the surprise management training session.

5.2 Training Development Rationale

Before working out the training design in chapter 3, the rationale behind the training is outlined in this chapter. There are several possibilities for surprise prevention and recovery training but due to limited time and resources we had to make choices, focussing on the most efficient and effective ways to reach the goal: less aviation accidents and incidents caused or contributed by surprise effects. Also, we tried to link new training and techniques to existing training and techniques as much as possible, for example, UPRT (Upset Prevention and Recovery Techniques), CRM and TEM.

5.2.1 Prevention and Recovery

Early on in the project the decision was made to focus on surprise recovery instead of prevention. There are two main reasons for this. First, as discussed in the literature review (NLR-CR-2016-619), the fight-or-flight reaction to startles and surprises is evolutionary very old and therefore virtually impossible to change or influence. Second, as surprise is defined as a mismatch between expectations and reality, it is realistic to state that humans cannot have the right expectations all the time. In a complex system as the aviation system the number of potential sources for surprise is unlimited. In other words, we will have to accept that pilots will be startled and surprised. Training should therefore be focused on providing pilots with tools to be able to adapt to and manage any startling/surprising situation.

That does not mean that no attention should be paid to prevention. Well-developed basic pilot skills and knowledge such as manual flying, CRM, and system knowledge can reduce the number of surprises or assist in recovery. Specific high risk surprise such as aircraft upsets can be targeted by specific training intervention. Training focussing on developing TEM skills and what-if planning can be useful. Another training solution targeting both prevention and recovery is the inclusion of more diverse simulation exercises compared to current practices. As these are addressed in many other aspects of training that is already in place, finding a recovery technique became the goal.

5.2.2 Desirable and Undesirable Behaviour

One of the starting points for the training design is the identification of desirable and undesirable behaviour. Desirable behaviour after a startle and/or surprise would be to stay calm, to remain communicative and to be able to think rationally. Undesirable behaviour would be an emotional (fear, panic), self-centred reaction and taking action without a clear, structured plan and TEM process. The unwanted and often present outcome of surprise is an unstructured decision making process, or no decision process at all.

Core ingredients for the training are derived from sport psychology and military mental training, leading to the 'startle effect management technique', which can be applied after a startle or surprise event. Three goals are leading in setting up the technique:

1. Controlling physiological and psychological reactions (emotions)
2. Being 'fail safe', i.e. not making things worse
3. Integration/connection with current practices

The fight-or-flight response needs to be controlled first. The emotions connected to the fight-or-flight response can be very intense because personal survival is involved. In the literature review, but also in the incident/accident review, we found that the startle effect implies that emotion is taking over at the expense of cognitive performance (sometimes called 'amygdala hijack').

The second goal is set to prevent 'wrong intuitive behaviour'. Surprise effects incite humans to take immediate action. The intuitive, non-analytical responses are often not the most effective in a flight deck environment and sometimes even aggravate the situation.

The third goal is to ensure the training has effect in operational flights. It is more likely that pilots will accept and apply a new technique when it is integrated or connected with current practices. Within KLM, the terms 'Unload, Roll, Power' (URP) are used to recover from any aircraft upset. Weighing the pros and cons it was decided using these terms in the event of a 'mental upset' would be better than the introduction of a new one. For other airlines, not familiar to URP tailoring would be required, using an existing acronym or introducing a new one. If a new acronym is required we suggest ROC: Relax, Observe, Confirm. These three words clearly describe the actions associated with each step.

5.3 The “Mental Upset” concept

5.3.1 Unload

To reach the first goal, an ‘active way of relaxing’, controlling emotions by doing something, is assumed to be more productive than telling people to do nothing, or at least nothing concrete (i.e. “take your time”). Expecting people to “do nothing” is counter-intuitive in a surprising -situation (particularly if a “fight-flight” response would be expected) because of the high physiological activity. And it might be counter-productive in a freeze situation where cognitive activity is already minimal. Different elements were combined in the first step, the “unload”-step of the URP technique:

1. taking physical distance
2. deep breathing
3. muscle relaxation
4. checking of colleague

First, fixation of attention needs to be avoided. Instead of focusing on the surprising event, focus should be on managing the effects of the event. Physical distance, or at least a focus on creating physical distance, can create mental distance. Specifically, pilots are instructed to push their backs into the back of the seat, preventing complete focus on one cue.

Second, deep breathing is used in other domains to manage emotions and to counteract physiological fight-or-flight reactions by focussing on breathing technique. Two of the more prominent physiological reactions are increased breathing rate and heart rate. Pilots are instructed to breathe in via the nose, using the diaphragm (deep breathing), and breathe out via the mouth. In this way breathing out can be controlled to last longer than breathing in, thereby slowing down the pace and creating physical relaxation, facilitating mental composure.

Third, another automatic physiological reaction is increased muscle tension. This physical tension does not assist pilots in achieving high cognitive performance and might even hinder fine motor skills. Muscle relaxation is achieved by squeezing the upper legs and then letting go of the tension in arms and shoulders. Conscious relaxing of the muscles is necessary after the automatic muscle tensioning response. This is a very short application of the widely used ‘Jacobson Progressive Relaxation Therapy’. Attention to physical relaxation is important because a calm, rational mind does not exist in a stressed, highly activated body.

All three first steps in the URP technique (taking physical distance, deep breathing and muscle relaxation are well known arousal control strategies. Amongst others these three are also mentioned in CAP 737 (The Civil Aviation Authority UK flight-crew Human Factors Handbook) as stress management and coping strategies. These three were selected for the application in surprise events on a flight deck for practical reasons, mainly the time constraints.

Finally, because of personal differences and due to noticing or not noticing of a surprising/startling cue, it is possible that pilots will not experience the same physiological and/or psychological effects. To minimise the time a crew requires aligning cognitive performance and starting working 'from the same page', a colleague-check is required. Preferably, this is done by using an open question and the other person's name to stop or prevent cognitive tunnelling (due to the 'cocktail party effect', a phenomenon that occurs when one may immediately detect words of importance, e.g. own name, originating from unattended stimuli). Asking is not just a procedural action, paying attention to the answer is equally important. Nonverbal signals may indicate the other pilot is not calm at all. In this case action can be taken to calm down the other pilot. In the extreme case where there is no response at all, it can be very effective to touch or lightly push the shoulder.

5.3.2 Roll

In the next URP step, "roll", the purpose is 'to start up' the cognitive process. Instead of trying to fully understand the situation at once, with the hazard of 'jumping to conclusions', observations are to be gathered and verbalised by each pilot. This can be structured by asking these questions:

- 'What do I see?',
- 'What do I hear?',
- 'What do I feel?', and
- 'What do I smell?'

This process uses all the applicable senses and can provide some solid ground in previously ambiguous conditions which can therefore have a calming effect as well. It can also provide a view of the big picture. An active search for other information is helpful to grasp the situation (e.g. N1 on engine #1 is zero, but N1 on engine #2 is still 98%). This 'taking action' provides humans with a sense of control. But focussing on the action 'observing and verbalising' prevents hasty decision making or performing incorrect and/or irreversible actions.

After the first pilot has done this, his colleague should correct and/or add his own observations. Only after establishing this level I SA, perception (Endsley, 1995) crews should try to give meaning to the situation, drawing conclusions (level II SA). A positive confirmation of the situation, if possible, should be called out to ensure a shared mental model. On this basis, an assessment of threats can begin to take place, together with an assessment of the available time and the ways to control the situation.

5.3.3 Power

In the "power"-step of the URP technique, a projection of the situation into the future is encouraged (level III SA) to foresee what mitigating measures can be taken to avoid eroding safety margins. This also applies the other way around: looking back at the decision making process by means of critical thinking to correct possible early-process-errors. This can be done by asking three critical questions:

- Do we miss information? (How sure are we about the information used?)
- Are there information conflicts/inconsistencies? (Among different sources)
- Are our assumptions correct?

This step was added to integrate the previous steps into the decision making processes. In recent years a special focus in KLM training was on Threat and Error Management (TEM) and one important aspect of this is to identify threats and foresee at what time these threats will influence flight operations. In anticipation of this, mitigating measures can be conceived. The other way to strengthen decision making is Critical Thinking (CT) which is common practice in military tactical decision making, but now is finding its way into aviation decision making. Critical thinking is disciplined process of actively and skillfully conceptualising, applying, analysing, synthesising, or evaluating information to be able to make clear and reasoned judgments. It has been described in many books and articles (Glaser, 1972) (Facione & Facione, 2007) as a metacognitive tool, with benefits such as avoiding plan continuation errors and jumping to conclusions.

5.3.4 Unload, Roll, Power as prevention

The main goal of the URP steps is to manage the fight-or-flight responses and assist in structured decision making after a startle or surprise. However, training the technique might have a secondary effect, creating a sense of increased self-efficacy in dealing with unexpected situations. This increased self-efficacy might reduce the effects of startles and surprises and might prevent fear-potentiated responses.

5.4 Training Design

The above mentioned reasons, theories, and decisions were used to create a three hour training program. A program of less than three hours is difficult to achieve when both classroom and simulator training are used. A program that takes more than 4 hours involves costs which will probably discourage the majority of airlines from implementing it. The next two paragraphs describe the broad set-up of the two training blocks, classroom and simulator, approximately 1.5 hour each.

5.4.1 Classroom

The content and goals of the main parts of the classroom training is described below.

- Experience: After an introduction, trainers, trainees and content, trainees are asked to perform a reaction time-test. However, during the test a surprise is presented and this has an effect on performance. What is experienced and why it affects performance is discussed.
- Theory: A short theoretical background on Startle & Surprise (S&S) is provided: definition of the two terms and psycho-physiological effects in combination with the fight-or-flight response. Non-aviation related video examples are shown and a link back to aviation is made via accidents/incidents. Then, the new URP-technique (mental upset) is introduced via the existing URP-technique (aircraft upset), explaining all the steps. Sport and military examples of similar mental training are used to minimise potential resistance.

- Personal experiences: Time is spent on discussing own experiences which have been written up in a homework assignment. This is important because S&S responses are different between individuals and this links the training to actual personal experiences, giving the training personal relevance.
- Practicing URP: The message is to always use URP if the need is there except in two situations in which immediate action is required: personal safety and flight path management. All letters of the acronym are practiced with a focus on the four steps of the Unload step. The usage of the simulator as a training environment for the startle effect management technique and not for correctly 'solving' scenarios is introduced here. Finally, a visualisation exercise is used to increase the transfer of training from classroom to simulator. Also, this part of the training is used to manage expectations for the simulator training: "you are not going to experience extreme S&S effects in the simulator and might not feel the urgency to apply the URP technique. However, try to imagine experiencing the simulator scenario in real life and what that might do to you, thinking about the personal responses you just described". This will also assist pilots in recognising S&S effects during operational flights.
- Training evaluation process: The usage of a self-scoring form is introduced and explained at the end of the classroom session. It asks pilots to rate the ease at which certain techniques can be applied and if they are perceived as successful in achieving the goal. Pilots are invited to use these forms to recognise which parts of the technique might require additional training and/or coaching. By doing this, trainees are also encouraged to take ownership of their own learning process and progress.

5.4.2 Simulator

The main goal of the simulator exercise is to practice the mental-upset URP technique in a flight deck environment, with focus on the Unload and Roll steps, as these are considered the most important and most novel. The Power step is part of the classroom session to ensure integration of the Unload and Roll steps in the decision making process, but in the simulator exercise, the focus is on startle effect management (U and R steps). With this in mind we designed a number of short scenarios, enabling ample opportunity for U and R practice. Some opportunity is provided to start the P, decision making, step. However, not the complete decision making process is played out due to time constraints and the fact that the focus of the training is on the novel, U and R, steps. The simulator is explicitly used as a training environment for the startle effect management technique and not for correctly 'solving' scenarios.

The second goal is to show that multiple, very diverse factors can cause S&S and that the startle effect management technique may be useful in all these situations. To design scenarios KLM's behavioural system SHAPE is used. 'S' stands for interaction with oneself, 'H' for human interaction, 'A' for interaction with the aircraft, 'P' for interaction with procedures, 'E' for interaction with the environment (based on the SHELL-model (Hawkins, 1987)).

Surprise category:	Description:	Examples:
S (Self)	Surprise concerning the pilots' individual perception or behaviour	- spatial disorientation/ visual illusions - surprise at own actions

H (Human Interaction)	Surprise concerning the interaction with other people in the aviation system	communication/ actions of: - other pilot - ATC - cabin crew - passengers (e.g. hijack, illness)
A (Aircraft)	Surprise concerning the aircraft's sudden behaviour	- automation surprise - aircraft upset - sudden controllability issues
P (Procedures)	Surprise concerning the use of procedures, resulting in an ambiguous situation	- multiple failures - conflict between procedures - unwanted outcome of procedures
E (Environment)	Surprise concerning elements in the environment outside of the aircraft, posing a direct threat to safety	- GPWS warnings/ terrain - change of runway conditions/ wind/ visibility - airport work in progress - birds - TCAS warnings/ other aircraft

During the simulator training we offer surprises from the H, A, P and E category, as a surprise from the S-category is very difficult to impose. The events in the simulator are focused on introducing different sorts of surprise and one startle (lightning strike), to demonstrate startle and surprise can be caused by many different types of events:

- Baseline measurement crew 1: automatic approach with an electronic failure causing the basic screens (Primary Flight Display & Navigation Display) of the pilot flying to blank.
- Baseline measurement crew 2 (training event for crew 1): an explosive decompression together with an engine severe damage.
- Training event 'human surprise': a non-precision approach during which the pilot monitoring would act confused and suggests to descent at the wrong waypoint (too early).
- Training event 'aircraft upset': sudden aircraft upset with 30 degrees nose up and 60 degrees of bank.
- Training event 'bird strike': take-off followed by one stalling engine and one engine with high vibration.
- Final measurement 'lightning strike': automatic approach with a lightning strike causing the mode control panel to freeze".

The details of the simulator scenarios are described in Appendix A. All these scenarios are about practicing the startle effect management technique in such a way that the chances of the technique being used during flight ops are maximised. In order to achieve this, the instructor maintains an active coaching role, where focus is on the use and the effects of the startle effect management technique. For example, instructors regularly ask trainees if they want to repeat (part of) the exercise. Deliberately, feedback will not be given on technicalities of the scenario, as this might distract from feedback on the technique. Another way to gain insight in their own learning progress is to provide trainees with self-scoring forms, rating their own application of the technique.

5.5 Training Evaluation Set Up

This chapter describes the way the measurements were set up to be able to evaluate training effectiveness.

5.5.1 Overview of simulator training and testing procedure

The table below lists the general set up of the simulator training elements and the specific assessment elements.

	Elements:
General	
	Short haul crew in B737 simulator, Long haul crew in B747 simulator Each 12 crews (10 + 2 margin) with range in experience
	Two crews in one simulator session of 3:30 hrs.
	Beginning of simulator session with baseline assessment of both crews, then training of URP[3]-strategy, ending with a test scenario to evaluate effectiveness of training
Baseline assessment	
	Baseline assessment consisting of a challenging surprise scenario
	Use of heart rate and eye tracking devices as quantitative measures (as a minimum, additional measures to be determined as required)
	Use of SHAPE 2.0-behavioural markers as qualitative measures
Training	*(Performance Based: How well can the trainee use what he knows?)*
	Training of crew in the separate elements of the URP-strategy, to be mastered sequentially (first U, then U + R, then U + R + P)
	Together with instructor assessment: subjective assessment of crewmembers to train some more or to move on to the next element
	Possibility of biofeedback (heart rate) to crews using low-tech devices (e.g. fit bit watch)
	Possibility of video feedback to crews
	Short surprise scenarios from different categories (SHAPE-categories)
	A few aircraft upset scenarios are introduced as well, to integrate the new strategy with existing strategy
Testing	
	Test consisting of a challenging surprise scenario to evaluate effectiveness of training
	Use of heart rate, eye tracking, and SHAPE 2.0, to be compared with the baseline assessment
	Use of subjective assessment of self-efficacy regarding surprise, but also of reactions to training

[3] Unload – Roll – Power: strategy for handling startle & surprise

Elements:

5.5.2 Baseline assessment

Crew classification	Part 1	Part 2	Part 3
CREW 1	Baseline assessment	Classroom training	Simulator training
CREW 2	Classroom training	Simulator training	

The table above shows the experimental design. Due to simulator availability, two crews train in one 3.5 hour simulator slot. One half of the crews (CREW 2) start the training in the classroom and afterwards move to the simulator. During the CREW 2's classroom training, CREW 1's start in the simulator for a baseline assessment before receiving classroom training. They are presented a surprise scenario to evaluate the crew's baseline performance using quantitative physiological data (heart rate and eye tracking), and qualitative behavioural data (SHAPE 2.0 behavioural marker system). This data will be compared with data from the test scenario at the end of the simulator session to evaluate the effectiveness of the total training (classroom + simulator). CREW 1's, who have already had the classroom training, are provided with the same surprise scenario at the start of their simulator training to evaluate the effectiveness of the classroom training.

With regard to SHAPE 2.0-components (see the SHAPE-table on page 14), the following will be observed to gather qualitative data:
(Note reference numbers refer to the SHAPE elements – e.g. S2 is "Self" #2, H3 is "Human Interaction" #3.)

SHAPE 2.0 aspect:	Component:	Reason:
UNLOAD		
S2 Self-Control	- 'Maintains self-control in rapidly changing and/or time-critical circumstances, regarding communication, decisions and/or actions'	Observe 'hyperactivity' Observe 'jumping to conclusions'
S5 Distraction Management	- 'Deals with interruptions, distractions and/or variations and effectively gets back to the task at hand'	Observe 'freeze'/ 'cognitive paralysis'
H3 Task Oriented Leadership	- 'Takes appropriate action upon signs of incapacitation of others'	Observe 'unload other pilot'
ROLL		
E1 Information Collection	- 'Searches actively and systematically for relevant information' - 'Uses available resources'	Observe 'calling out obervations' Observe 'tunneling'
E2 Threat Analysis & Identification	- 'Identifies threats and assesses accurately the state of the aircraft and its systems' - 'Identifies threats and assesses accurately environmental factors that may affect the operation'	Observe 'accurate interpretation of situation'
POWER		

SHAPE 2.0 aspect:	Component:	Reason:
E3 Planning & Anticipation	- 'Discusses expectations that can influence the operation' - 'Mentions appropriate mitigating measures'	Observe 'mental simulation' Observe 'mitigating measures'
E1 Information Collection	- 'Raises the question of correctness, relevance, value and completeness of information' - 'Seeks clarification when unsure about any aspect of the available information' - 'Compares new with existing information'	Observe 'critical thinking'

5.5.3 Simulator training

Simulator training consists of the three elements of the URP strategy: first the 'unload'-element will be trained by providing scenarios, but freezing the simulator after unloading by the crew (as illustrated/trained in classroom training). The same setup will be followed in training the 'Roll' (where 'Unload' must precede) and the 'Power' element (where 'Unload' and 'Roll' must precede).

Feedback will be given to the crew by the instructor, with the use of biofeedback (e.g. heart rate) and video feedback when considered helpful. Crews are asked whether this separate URP-element is mastered or if they feel need to train extra scenarios. For this purpose multiple short scenarios are prepared.

Care will be taken to provide for different sorts of surprise, using the interactions described in SHAPE: 'S' stands for interaction with oneself, 'H' for human interaction, 'A' for interaction with the aircraft, 'P' for interaction with procedures, 'E' for interaction with the environment. One or two aircraft upset scenarios are also incorporated to integrate the new strategy with already learned strategies.

Other examples of scenarios: sudden airport closure (E), map shift (A), conflicting procedures with unreliable airspeed and gear stuck in down position (P), other pilot calling twice 'you have to descend' when it is actually too early to start the approach (H), stuck thrust lever (A), sudden engine relight after a flameout (A), ATC is not cooperating in an emergency condition as there are other aircraft also in emergency conditions (H), a standby bus failure resulting in multiple black screens but procedure is not helpful in recovering (P), TOGA switch failure in a go around (A). A surprise from the 'S' category cannot be imposed, but can be expected when confronted with startle (lightning strike, vibration, noise, flash).

5.5.4 Simulator testing

At the end of the simulator session a challenging surprise scenario will be provided to evaluate effectiveness of training (total training or only classroom training). The scenario will be comparable to the initial scenario used in the baseline assessment. The same data as in the baseline assessment will be gathered, but also subjective assessment from the crew is asked to measure self-efficacy and reactions to training.

6 Results of the evaluation

The goal of the project was to develop and evaluate a training session that could be used to mitigate the effects of startle and surprise on flight crew. The aim of the training session is to demonstrate technique(s) that can be used to manage startle and surprise in the operational environment. The project should produce a practical tool or technique that could be trained relatively quickly, has a scientific foundation and can be applied to any surprise situation to mitigate and/or recover from the effects of surprise reactions: calm down, communicate and apply an appropriate decision making process.

Report Number	Title	Date
NLR-CR-2016-619	Literature review on startle, surprise and acute stress	March 2017
NLR-CR-2016-620	Review of startle and surprise role in incidents and accidents	April 2017
NLR-CR-2016-621	Review of existing training methods	April 2017
NLR-CR-2016-622	Definition of startle and surprise effect management training	April 2017
NLR-CR-2017-102	Cost benefit analysis of startle and surprise training	June 2017

This report describes the results from the simulator experiments that were carried out at the KLM training centre using a B737-800 simulator and a B747-400 simulator to evaluate the training technique that was developed by the project research team. The experiments mainly took place in May and June of 2016 with voluntary crews from KLM.

The aim of this report is to provide an overview of the experiment execution and experimental results. It will include instructor observations, feedback from the volunteers, follow-up questionnaire data, and data gathered during the experiment (eye-tracking, heart rate). The focus of this analysis lies on instructor observations and on self-reported changes in self-efficacy (in this case the pilot’s belief in handling Startle and Surprise), as physiological measurements of Startle and Surprise are well described in other experiments (e.g. Kochan, Breiter & Jentsch, 2005; Bürki-Cohen, 2010; and as described in Rivera, Talone, Boesser, Jentsch, & Yeh, 2014) .

6.1 Participants

A total of 44 volunteers took part in the training and testing, all active pilots within KLM. Three variables between the participants were taken into consideration: long-haul and short-haul; line pilots and instructors; and flight time experience. These were used to categorise the pilot participants.

To see if there was any difference between long haul and short haul pilots we looked at two aircraft types: 24 B747-400 pilots and 20 B737-NextGen pilots. Another possible point of interest was the difference between instructors and line-pilots: half of the volunteers was instructor or had experience in giving training, and the other half did not. To determine whether flight time experience had any noticeable effect a group of 20 pilots with more than 10000 hours was compared with a group of 24 pilots with less than 10000 hours. The similar group sizes for the instructor-line pilot group and the more than and less than 10000 hours groups is a coincidence. Predictably, the number of instructors in the more than 10000 hours group is higher than that of the line pilots. Instructor roles within airlines are usually performed by more experienced pilots. This can be seen in Table 1 where the average hours number for instructors is considerably higher.

Table 5 gives an overview of the descriptive statistics for the participants who responded to the evaluation questionnaire (n=42). Note that since these statistics are based on the responses for the evaluation questionnaire, where the descriptive data was collected, it does not include all of the participants (n=44). Two short-haul pilots did not complete the evaluation questionnaire, so their data is not included in this overview.

Table 5 Participant descriptive statistics (n=42)

	Number	**Experience**	**Instruction**	**Average Hours**
Short-haul pilots	18	10	8	8978
Long-haul pilots	24	9	12	8013
Experienced pilots	19	-	15	13116
Less experienced pilots	22	-	5	4552
Instructors	20	15	-	12010
Line-pilots	22	4	-	5168

6.2 General Setup of the Experiment

The experiment was set up to examine the effect of the training including a classroom session of 1.5 hours, as well as a simulator training and coaching session, or the effect of each session individually – just the classroom, or just the simulator session. As such the experiment was divided into blocks, and two crews were scheduled in for each simulator slot.

Table 6 Experiment Blocks

Experiment blocks	**Uninformed Group (n = 20)**	**Informed Group (n=24)**
Prior classroom training + questionnaire		x
Pre-test	x	
Classroom training intervention	x	x
Mid-test	x	x
Simulator training intervention	x	x
Post-test	x	x

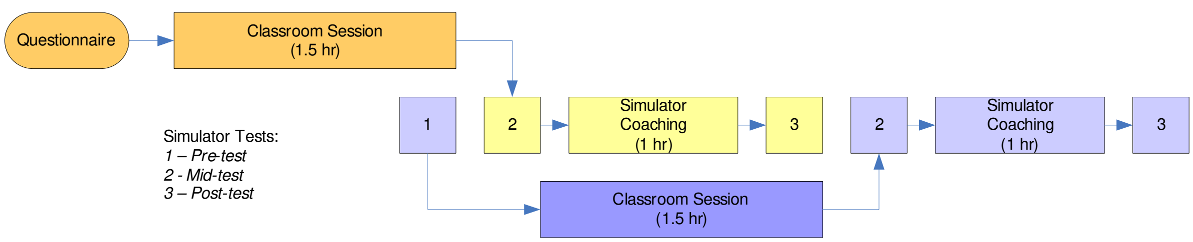


Figure 2 Experiment Training Session Timeline

In addition to the variables taken into consideration in describing the participants, the participants were split into two experimental groups.

- One group – the "Uninformed Group" - was given no information on the experiment prior to arriving for the experiment slot at the training centre and conducting a pre-test in the simulator. (Two 'Uninformed Group' training sessions were cancelled; hence the different number of participants.)
- The second group – the "Informed Group" –was sent an electronic questionnaire and classroom training material about the subject as preparation for the experiment prior to arrival. About 2 weeks before the experiment the participants in this group received an electronic questionnaire. Questions were asked about their own experiences with startle and/or surprise. This was done to open the minds for the subject and to have some material for the trainer to use during the classroom training

Both crews received the same classroom training intervention, mid-test simulator measurement, simulator training intervention and the same post-test simulator measurement. This was done to provide for two crews in one, 3.5 hour, simulator slot.

6.2.1.1 Training

The full description of the training as it was developed and delivered in the evaluation experiment is described in the separate report: NLR-CR-2016-622; Definition of startle and surprise effect management training. The training report documents the “URP technique” – Unload, Roll, Power – that was developed to manage the startle effect.

The classroom training consisted of the following elements:

- a reaction test, in which a small distraction illustrates the cognitive effect of surprise
- discussing of own experiences with startle and/or surprise, to align new information with the already known
- providing a brief theoretical background on Startle and Surprise, using PowerPoint
- creating urgency by showing a few accidents (Air France 447, Asiana 214) and serious incidents (Qantas 32, Hudson landing)
- training of the ‘startle effect management technique’ (“unload, roll, power”, URP technique)
- visualisation exercise of the technique to promote transfer between classroom and simulator

The events in the simulator were focused on introducing different sorts of surprise, yet without becoming unrealistic. For example:

- training event ‘human surprise’: a non-precision approach where the pilot monitoring would act confused and wanting to descent at the wrong waypoint
- training event ‘aircraft upset’: sudden aircraft upset with 30 degrees nose up and 60 degrees of bank
- training event ‘bird strike’: take off followed by one stalling engine and one engine with high vibration

6.2.1.2 Measurements

The following measurement tests were taken:

- pre-test measurement for the Uninformed Group: automatic approach with an electronic failure causing the basic screens (Primary Flight Display & Navigation Display) of the pilot flying to blank.
- mid-test measurement for both groups: an explosive decompression combined with an engine severe damage.
- end measurement ‘lightning strike’: automatic approach with a lightning strike causing the mode control panel to freeze.

An overview of the different scenarios that were used in the simulator session for measurements and coaching on the startle management technique is included in 0.

The effects of the URP technique were assessed by the instructors using the SHAPE behavioural markers (see Figure 1). The SHAPE model provides for a wide spectrum of technical and non-technical behaviours. The focus of the training and the assessment was on the initial handling of the scenario (the “unload”-step and the “roll”-step), not on problem solving or further decision making (the “power”-step). In addition to the instructor observations, video and simulator data were used to record the mid-test and post-test measurements.

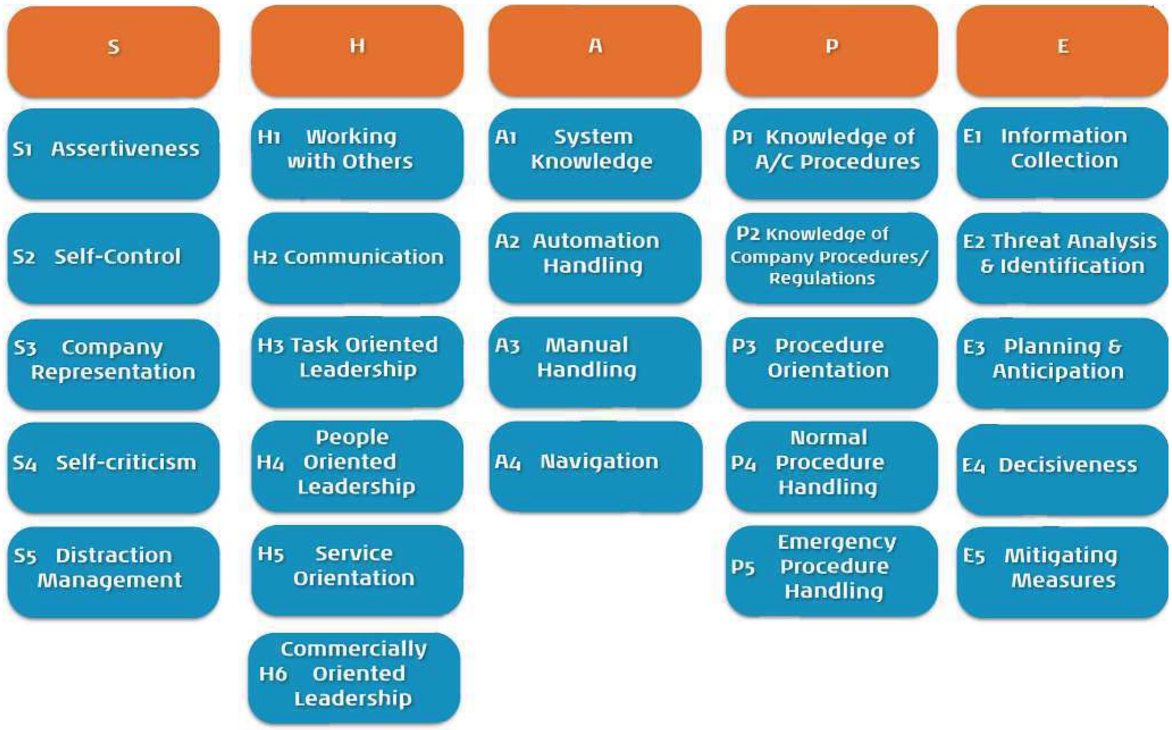


Figure 3 KLM SHAPE behavioural marker scheme

Eye-tracking and heart rate data was recorded for some participants to establish an indication of the usefulness of this data in the experiment assessment. The participants were asked to self-rate their performance at two points in the experiment. After the classroom training, the startle effect management technique was tested in the simulator – the mid-test scenario - using an “explosive decompression” scenario. A self-score form was provided where the pilots could rate their own progress on the separate steps of the technique. A second self-score form was provided after the last training event, the bird strike. In this way pilots were made aware of their own progress in a structured way.

Instructor observations are structured by four SHAPE indicators, to be able to compare performance with later scenarios:

Table 7 Instructor Simulator Observations

Indicator:	**Description:**
Self-Control	The pilot remains calm and does not do any sudden intuitive actions
Distraction Management	The pilot remains in an active mode and does not ‘freeze’
Checking of colleague	The pilot checks his colleague on signs of incapacitation
Information Collection	The pilot sticks to the facts by calling out his/her observations and does not interpret these facts right away by stating his opinion of the nature of the situation

We expect that pilots will not be fully ‘startled’ or ‘surprised’, as pilots expect things to happen in simulators. Therefore, ‘self-control’ would probably not reveal significant results. We also expect that pilots will remain active and not ‘freeze’ during the scenarios (indicator ‘distraction management’), as the scenarios were not extremely unusual and pilots are trained for similar scenarios. With the ‘checking of colleague’-indicator we expect to see progress, as this is not normal routine for pilots within KLM. In the “unload”-step the checking of colleague is explicitly mentioned and this was trained, both in the classroom and in the simulator. With the fourth indicator, ‘information collection’, we have no specific expectations, as calling out observations before interpreting (which Mica Endsley calls ‘Level I Situation Awareness’), has been addressed in previous training. However, it does not have a prominent place in instructor feedback. Communication about what pilots see, hear, feel and smell probably enhances shared situational awareness and increases chances of a correct situation assessment. This is the important foundation of further decision making, gathering information. Focussing attention on information gathering also decreases chances of jumping-to-conclusions and/or taking action without a proper assessment.

In addition to instructor observations, video observations will be applied to better observe the “unload”-step:

Table 8 Instructor Video Observations

Indicator:	**Description:**
Physical distance (+muscle tension release)	Moving backward with head and shoulders
Deep breathing technique	Deep inhalation via nose and out via mouth

With these indicators, together with the ‘checking of colleague’-indicator, we can observe whether the “unload”-step is performed as trained.

The complete training has been evaluated using a questionnaire, with questions about the training elements, about the startle effect management technique, and self-efficacy. Two months after the training, a follow-up questionnaire was distributed gathering information about startle/surprise awareness, about actual startle/surprise events, and about usage of the trained technique.

6.3 Descriptive Statistical Analysis

Paired samples t-tests were carried out to compare the participant's performance in the six critical instructor observations (Self-control, Distraction management, Information collection, Physical distance, Deep breathing and checking of colleague) between the pre-test, mid-test and post-test on the simulator.

The results indicate that there was a significant difference in information collection between the pre-test and post-test performance (M=0.42, SD=0.69); $t(18) = -2.650$, $p = .016$.

Table 9 Paired Samples Test Results - Information Collection

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pre - Post	-.421	.692	.159	-.755	-.087	-2.650	18	.016

This indicates that the complete training session led to a significant improvement in the information collection of the participants. It is interesting to note that there was also a significant improvement for information collection between the pre- and mid-test, and between the mid-and post-test.

Table 10 Descriptive Statistics - Information Handling

Descriptive Statistics – Information handling

	N	Minimum	Maximum	Mean	Std. Deviation
Pre	19	0	1	.42	.507
Mid	44	0	1	.66	.479
Post	44	0	1	.89	.321
Valid N (listwise)	19				

The differences for the other instructor observation parameters were not found to be significant, although there were trends observed that are further discussed combined with the qualitative analysis of the instructor evaluators below.

With respect to the information handling, a relationship was found between the information handling, and physical distance competencies observed during the simulator experiments , using a chi-square test – X^2 (1, N=33) = 6.42, p=0.01. This indicates that the physical distance that was trained during the session had the desired positive effect on the crew's information handling.

To further investigate the effect of the different elements of the training session, and the instructor observed competencies on each other, an analysis of the correlation between the different competencies was carried out. This analysis highlighted a number of significant correlations.

A chi-square test was performed and a relationship was found between the performance in the Deep Breathing and Physical Distance competencies X^2 (1, N = 34) = 12.1, p < .001. Similarly a relationship was found between the Checking of Colleague and Deep Breathing competencies X^2 (1, N=34) = 16.07, p< .001.

The trends observed between the different factors are reported below, using the descriptive statistics of the instructor observations of the competencies for each of the elements of the training – classroom, simulator and complete training session.

6.4 Baseline Performance

In the pre-test scenario (electrical bus failure on approach) for the uninformed group we observed 20 pilots and have 19 valid observations:

Table 11 Baseline (pre-test) Performance

Indicator	Total pre-test score	
Self-control	17	(89%)
Distraction management	17	(89%)
Information collection	8	(42%)
Physical distance	0	(0%)
Deep breathing	0	(0%)
Checking of colleague	4	(21%)

The table above indicates that:

- 17 pilots remained calm and did not perform any sudden actions (‘self-control’)
- 17 pilots remained in an active mode and did not freeze (‘distraction management’)
- 8 pilots called out observations instead of interpreting the situation at once (‘information collection’)
- 4 pilots checked their colleague on signs of incapacitation (‘checking of colleague’)
- None of the pilots used the ‘deep breathing’ technique, nor did they take ‘physical distance’

As expected, most pilots remained calm. The two pilots who did not remain calm immediately initiated a go-around the moment the bus failure showed itself. The same pilots did not engage in any activity focused on information gathering, nor did they communicate much. When asked, these pilots confirmed they were really surprised. Also, as expected, we did not see any of the pilots use the deep breathing technique or take physical distance, as this is not normal routine. However, we did see 4 pilots check each other, which is also not normal routine but apparently is already a personal strategy for some pilots. Regarding the ‘information collection’-indicator, we can see that 40% of the pilots do use a structured information collection process by first calling out their observations.

6.4.1 Pilot variable results

Taking into account the participant variables for these 19 pilots gives the following result (physical distance and deep breathing is not taking into account due to 0% scores):

Table 12 Baseline Performance: Short-haul vs Long-haul pilots

Indicator	short haul (n=8)	long haul (n=11)	Difference
self-control	6 (75%)	11 (100%)	25%
distraction management	6 (75%)	11 (100%)	25%
checking of colleague	2 (25%)	2 (18%)	7%
information collection	4 (50%)	4 (36%)	14%

Table 13 Baseline Performance: Experienced vs Inexperienced pilots

Indicator	experienced pilots (n=7)	inexperienced pilots (n=12)	Difference
self-control	7 (100%)	10 (83%)	17%
distraction management	7 (100%)	10 (83%)	17%
checking of colleague	3 (43%)	1 (8%)	35%
information collection	3 (43%)	5 (42%)	1%

Table 14 Baseline Performance: Instructors vs Line-pilots

Indicator	instructors (n=8)	line-pilots (n=11)	Difference
self-control	8 (100%)	9 (82%)	18%
distraction management	8 (100%)	9 (82%)	18%
checking of colleague	2 (25%)	2 (18%)	7%
information collection	4 (50%)	4 (36%)	14%

Regarding the indicators ‘self-control’ and ‘distraction management’ we see a high pre-test score of 89% for the total observed group. The group score of the long haul pilots is higher on each indicator (+25%) compared to the short haul pilots group score, experienced pilots group scores slightly higher (+17%) than the inexperienced pilots, and the instructors score slightly higher (+18%) than the line-pilots.

Regarding the indicator ‘checking of colleague’, the pre-test score for the total group is low (21%), probably because this is not incorporated in the standard operating procedures. The experienced pilot group scores considerably higher (+35%) than the inexperienced pilots.

Concerning the indicator ‘information collection’ we see that 42% of all pilots do call out their observations before analysing and that the other 58% does not. Within this group of 19 pilots the short haul pilot group scores slightly higher (+14%) than the long haul pilots, and the instructor group scores slightly higher (+14%) than the line-pilots.

6.5 Effects of classroom training

6.5.1 Instructor Observations

To measure the effects of classroom training the pre-test observations of the Uninformed Group are compared with the observations using the SHAPE behavioural markers from the first scenario after classroom training: the mid-test-scenario (explosive decompression combined with engine seizure, see appendix A for scenario overview). On the same indicators progress was achieved as indicated in the table and figure below (note that data for some crews was not observed as a result of failed video observation).

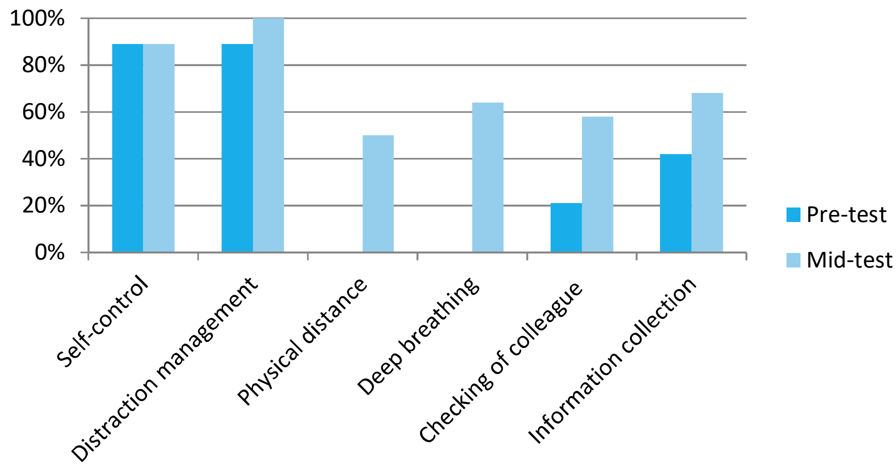


Figure 4 Pre-test vs Mid-test Instructor observations

Again, most pilots remained calm. The two pilots who did not remain calm initiated an emergency descent before properly assessing the situation (they rushed their communication and actions). At the same time, two pilots that remained calm during the pre-test were not calm in the mid-test. These results are indicated by (-2 +2) in the table and lead to a total group progress of 0. In the mid-test, all 19 pilots remained in an active mode; they actively searched for indications to clarify their situation. This implies that on 'distraction management total progress of +2 (-0 +2) was achieved. For information collection +5 (-2+7) was observed: two out of eight previously calling out pilot did not do so in the mid-test, but of the eleven pilots previously not calling out, now seven did do so. As the pre-test score was already high (42%), progress on this indicator is relatively low (+26%), but nevertheless a recognizable contribution of classroom training. Half of the pilots now take physical distance, leading to a +8 (-0+8), a notable +50% effect. Eleven of the seventeen pilots previously not practicing deep breathing did so in the mid-test. Progress on 'Deep breathing' therefore adds up to +11 (-0+11), a +64% classroom effect. As hoped for, more pilots (+38%; +7 (-3 +10)) checked their colleagues. We can conclude that classroom training had a positive effect.

Table 15 Classroom progress: Pre-test vs Mid-test

Indicator:	**Pre-test**	**Mid-test**	**Progress**	
Self-control (n=19)	17 (89%)	17 (89%)	0 (-2+2)	+ 0%
Distraction management (n=19)	17 (89%)	19 (100%)	+2 (-0+2)	+ 11%
Physical distance (n=16)	0 (0%)	8 (50%)	+8 (-0+8)	+ 50%
Deep breathing (n=17)	0 (0%)	11 (64%)	+11 (-0+11)	+ 64%
Checking of colleague (n=19)	4 (21%)	11 (58%)	+7 (-3+10)	+ 37%
Information collection (n=19)	8 (42%)	13 (68%)	+5 (-2+7)	+ 26%

Concerning the indicators 'self-control' and 'distraction management' we do not see much progress due to already high pre-test scores. The only notable effect is that short haul pilots make progress on distraction management (long haul pilots already score 100% on the pre-test).

On the indicator ‘physical distance’ we see a lot of progress, as expected, but especially with the group of long-haul pilots and with the instructors group. Interestingly, the group of short-haul pilots doesn’t show much progress on this indicator.

Considerable progress on the indicator ‘deep breathing’ is achieved, as was expected, but especially with inexperienced pilots.

Concerning the indicator ‘checking of colleague’ we see a lot of progress, especially with the group of inexperienced pilots (+50%) compared to the experienced pilots (+14%). The inexperienced group had lower pre-test scores, but was able to catch up after classroom instruction. The other groups do not show that much difference.

On the indicator ‘information collection’ considerable progress is achieved, especially with the group of short haul pilots (+50%) compared to the long haul pilots (+11%). In this case, both groups did not show very high pre-test scores (50% resp. 34%), so progress cannot just be attributed to this. (See Appendix E for a description of the analysis between pilot categories).

6.6 Effects of Simulator training

6.6.1 Instructor Observations

To measure progress after classroom training we compared the mid-test observations with the observations done in the last scenario of the simulator training: the post-test-scenario (lightning strike on approach). On the same indicators we observed the following with 41 pilots:

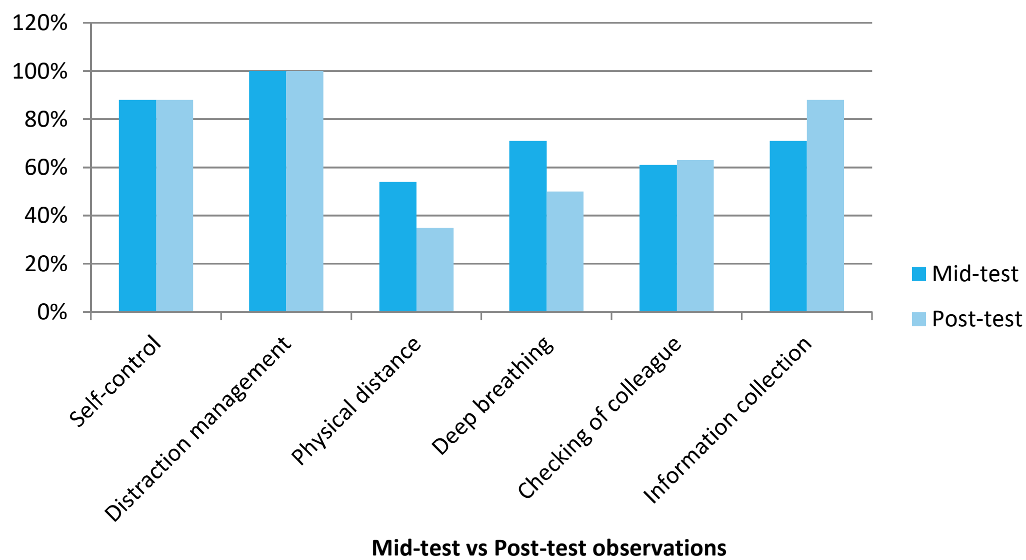


Figure 5 Simulator progress - instructor observations

What we see is stabile performance on self-control and distraction management, a drop in performance on physical distance and deep breathing, but an increase in performance on information collection. As this last indicator is most important for the decision making process we looked at possible correlations with the “unload”-step (physical distance, deep breathing, checking of colleague). It turns out that in the mid-test these correlations cannot be found, but in the post-test we do find a correlation between physical distance and information collection. For this we used Fisher’s exact test with result p=0.15. Furthermore, we find a correlation between physical distance and deep breathing (p<0.001, Fisher’s exact test). Apparently these steps combine naturally for the pilots.

Table 16 Simulator progress: Mid-test vs Post-test

Indicator	**Mid-test**	**Post-test**	**Progress**	
Self-control	36 (88%)	36 (88%)	0 (-5+5)	+ 0 %
Distraction management	41 (100%)	41 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=37)	22 (54%)	13 (35%)	- 5 (-11+6)	- 19 %
Deep breathing (n=36)	24 (71%)	18 (50%)	- 6 (-11+5)	- 21 %
Checking of colleague	25 (61%)	26 (63%)	+ 1 (-12+13)	+ 1 %
Information collection	*29 (71%)*	*36 (88%)*	*+7 (-4+11)*	*+17 %*

If we show these results with each indicator we can see what progress was made.

6.6.1.1 Self-control:

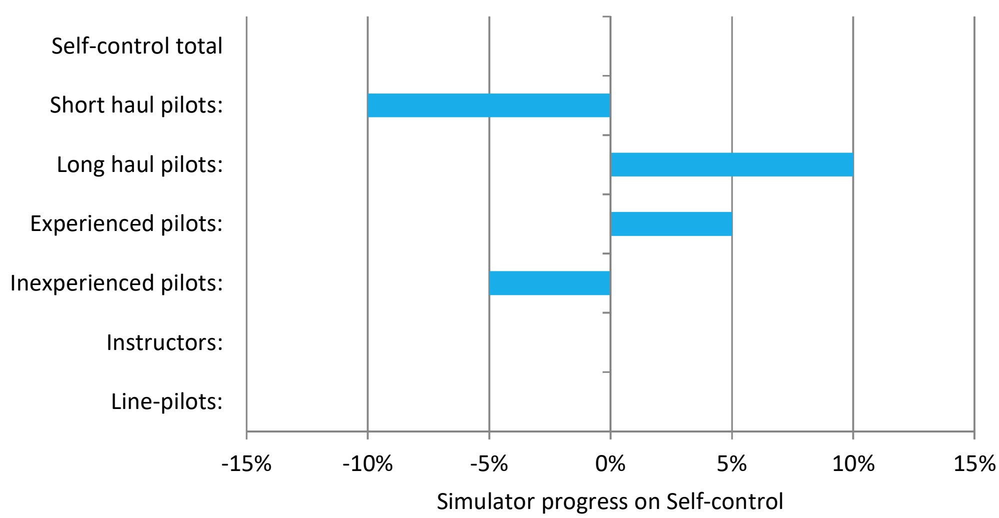


Figure 6 Simulator Progress: Self-control

This indicator showed no progress after the simulator session across the total participant group. However, some differences in progress can be observed within the different categories of pilots.

Table 17 Simulator Progress: Self-control observation

	Mid-test	Post-test	Progress	
Total :	**36 (88%)**	**36 (88%)**	**0 (-5+5)**	**+ 0 %**
Short haul pilots	17 (85%)	15 (75%)	-2 (-5+3)	- 10 %
Long haul pilots	19 (90%)	21 (100%)	+2 (-0+2)	+ 10 %

Experienced pilots	16 (84%)	17 (89%)	+1 (-2+3)	+ 5 %
Inexperienced pilots	20 (91%)	19 (86%)	-1 (-3+2)	- 5 %
Instructors:	19 (95%)	19 (95%)	0 (-1+1)	+ 0 %
Line-pilots:	17 (81%)	17 (81%)	0 (-4+4)	+ 0 %

6.6.1.2 Distraction management:

This indicator also showed no progress and no spread. 100% was scored in all observations.

To summarise, concerning the indicators ‘self-control’ and ‘distraction management’ we do not see any progress in the simulator training if we consider the total of 41 pilots. However, we do see some decrease in the indicator ‘self-control’ with short haul pilots (-10%) and inexperienced pilots (-5%). We see some increase in this indicator with long haul pilots (+10%) and experienced pilots (+5%). In the groups we do not see any difference.

6.6.1.3 Physical distance:

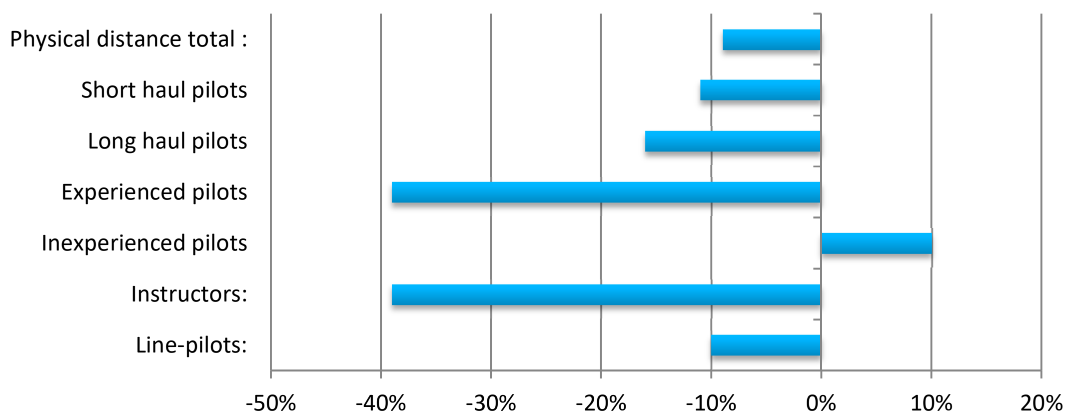


Figure 7 Simulator Progress: Physical distance

Apart from the inexperienced pilot group, we see a decrease in performance on this indicator. This indicates that training effect for this skill is fragile.

Table 18 Simulator Progress: Physical distance

	Mid-test	Post-test	Progress	
Total :	**22 (54%)**	**13 (35%)**	**- 5 (-11+6)**	**- 9 %**
Short haul pilots	6 (33%)	4 (22%)	-2 (-6+4)	- 11%
Long haul pilots	12 (63%)	9 (47%)	-3 (-6+3)	- 16%
Experienced pilots	11 (61%)	4 (22%)	- 7 (-9+2)	- 39 %
Inexperienced pilots	7 (37%)	9 (47%)	+2 (-3+5)	+ 10%
Instructors:	11 (61%)	4 (22%)	-7 (-9+2)	- 39%
Line-pilots:	7 (37%)	9 (47%)	+2 (-3+5)	- 10%

6.6.1.4 Deep breathing:

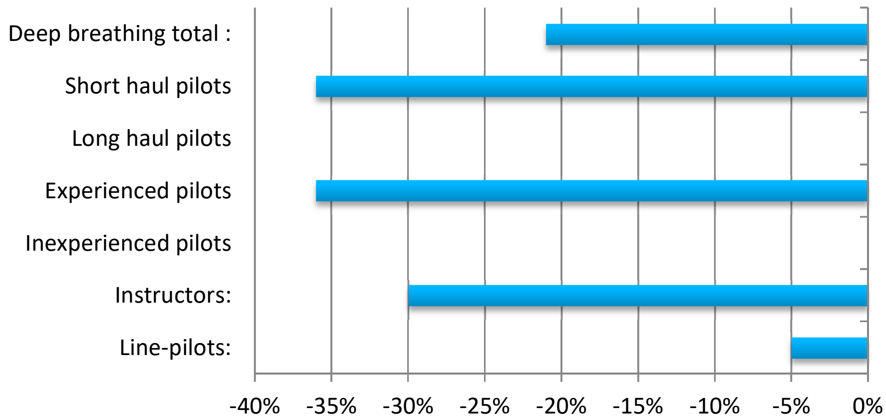


Figure 8 Simulator Progress: Deep breathing

Apart from the inexperienced pilot group, we see a decrease in performance on this indicator. This indicates that training effect for this skill is fragile.

Table 19 Simulator Progress: Deep breathing

	Mid-test	Post-test	Progress	
Total :	**24 (71%)**	**18 (50%)**	**- 6 (-11+5)**	**- 21 %**
Short haul pilots	11 (65%)	5 (29%)	-6 (-8+2)	- 36 %
Long haul pilots	13 (68%)	13 (68%)	0 (-3+3)	+ 0 %
Experienced pilots	12 (71%)	6 (35%)	- 6 (-8+2)	- 36 %
Inexperienced pilots	7 (37%)	7 (37%)	0 (-3+3)	+ 0 %
Instructors:	12 (71%)	7 (41%)	-5 (-7+2)	- 30 %
Line-pilots:	12 (63%)	11 (58%)	-1 (-4+3)	- 5 %

6.6.1.5 Checking of colleague:

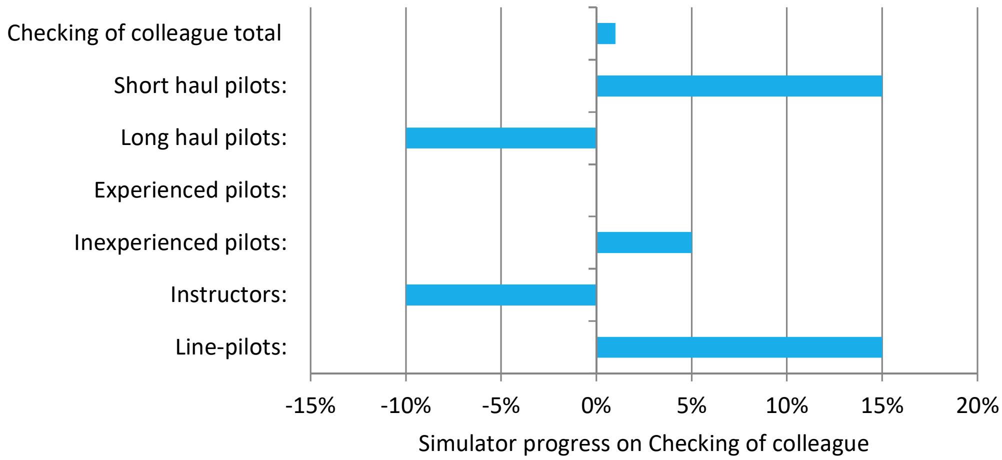


Figure 9 Simulator Progress: Checking of colleague

Concerning the indicator ‘checking of colleague’ we do not see any progress in the simulator training if we consider the total of 41 pilots (+1%), but we do see a lot of spread (pilots who performed well, did not do so later on (-12), and vice versa (+13)). Furthermore, we see an increase with short haul pilots and line-pilots (both +15%). We see a decrease with long haul pilots and instructors (both -10%). In the experienced pilots group and in the inexperienced pilots group there was not much progress, only a lot of spread. The large spread in total (and in every group) suggests that the training effect on this indicator is not very stable.

Table 20 Simulator Progress: Checking of colleague

	Mid-test	Post-test	Progress	
Total :	**25 (61%)**	**26 (63%)**	**+1 (-12+13)**	**+ 1 %**
Short haul pilots	11 (55%)	14 (70%)	+3 (-3+6)	+ 15 %
Long haul pilots	14 (67%)	12 (57%)	-2 (-9+7)	- 10 %
Experienced pilots	12 (63%)	12 (63%)	0 (-6+6)	+ 0 %
Inexperienced pilots	13 (59%)	14 (64%)	+1 (-6+7)	+ 5 %
Instructors:	14 (70%)	12 (60%)	-2 (-8+6)	- 10 %
Line-pilots:	11 (52%)	14 (67%)	+3 (-4+7)	+ 15%

6.6.1.6 *Information collection:*

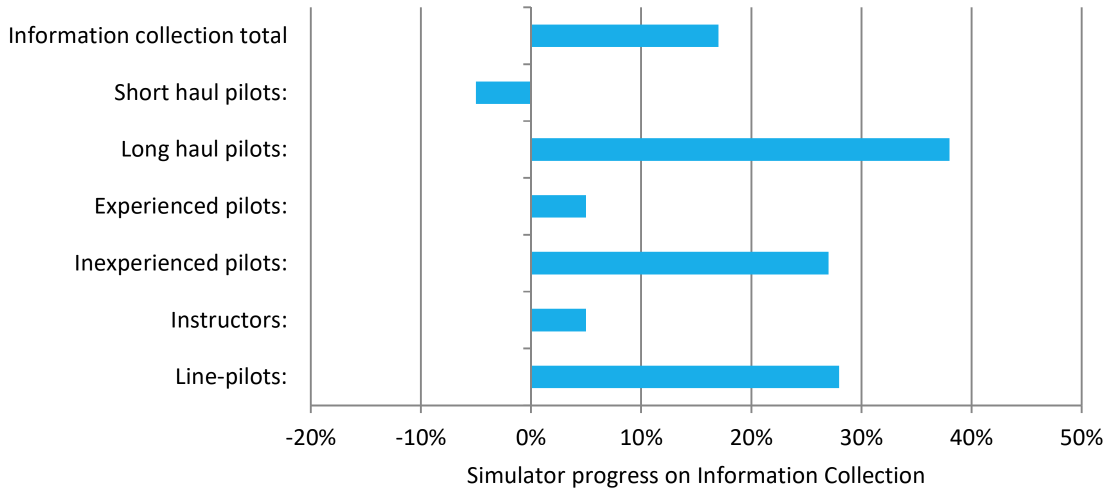


Figure 10 Simulator Progress: Information collection

Concerning the indicator 'information collection' we see a lot of progress in the simulator training if we consider the total of 41 pilots (+17%), with not much spread. However, the short haul pilots group showed some decay (-5%), but this should be seen in the light of already high scores in the mid-test. They end up with about the same end result as the long haul pilots, who did show remarkable progress (+38%). It must be said that these pilots came from low scores in the mid-test (52%). With the experienced pilots and the instructors we can see a small progress (+5%), but they already scored rather high in the mid-test (79% and 75%). With the inexperienced pilots and the line-pilots we see also quite some progress (+27% and +28%), but they also came from low scores in the mid-test.

Table 21 Simulator Progress: Information collection

	Mid-test	Post-test	Progress	
Total :	**29 (71%)**	**36 (88%)**	**+7 (-4+11)**	**+17 %**
Short haul pilots	18 (90%)	17 (85%)	-1 (-3+2)	- 5 %
Long haul pilots	11 (52%)	19 (90%)	+8 (-1+9)	+ 38 %
Experienced pilots	15 (79%)	16 (84%)	+1 (-2+3)	+ 5 %
Inexperienced pilots	14 (64%)	20 (91%)	+6 (-2+8)	+ 27 %
Instructors:	15 (75%)	16 (80%)	+1 (-3+4)	+ 5 %
Line-pilots:	14 (67%)	20 (95%)	+6 (-1+7)	+ 28 %

6.7 Effects of the full training package

6.7.1 Instructor Observations

To be able to measure total progress across the training session – including both the classroom and simulator training - a comparison was made between pre-test observations and post-test observations. For this we return to the pilots in the uninformed group. On the same indicators we observed the following for a total of 19 pilots:

- Two out of the 17 previously calm remaining pilots were not calm in the post-test scenario.
- All of the 19 pilots remained in an active mode, so two more than in the pre-test scenario.
- Three out of 4 previously checking colleagues did not do so in the post-test scenario, but of the 15 previously not checking; now 12 pilots did check each other, bringing the total score to 68%.
- Two out of 8 previously calling out observations did not do so in the post-test scenario, but of the 11 previously not calling out, now 10 pilots did do so, bringing the total score to 84%.

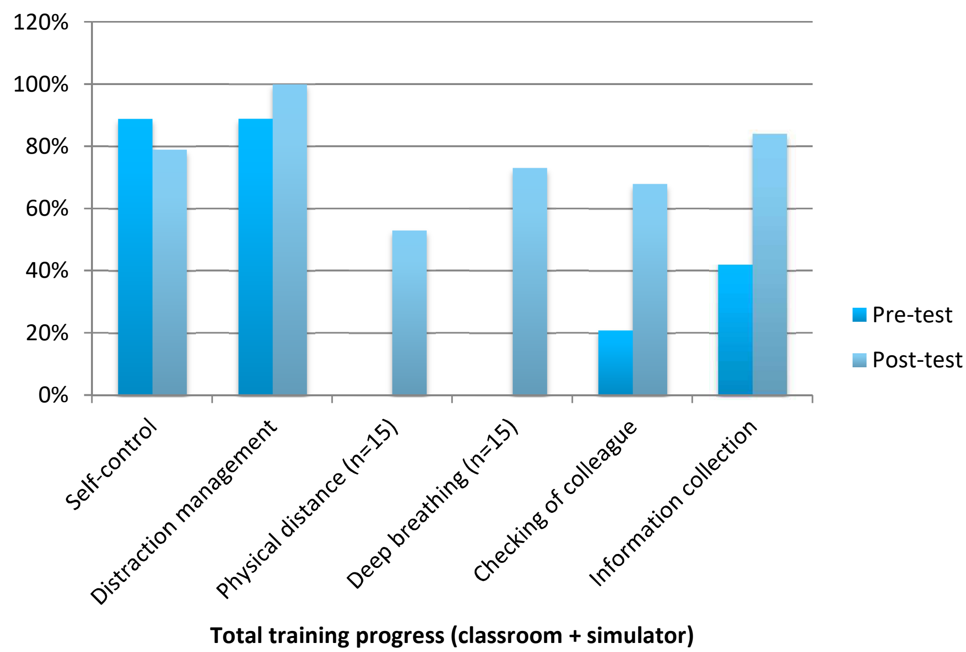


Figure 11 Total Training Observations

Table 22 Assessment of full training package

Indicator:	Pre-test	Post-test	Progress	
Self-control	17 (89%)	15 (79%)	-2 (-2+0)	- 10%
Distraction management	17 (89%)	19 (100%)	+2 (-0+2)	+ 11%
Physical distance (n=15)	0 (0%)	8 (53%)	+8 (-0+8)	+ 53%
Deep breathing (n=15)	0 (0%)	11 (73%)	+11 (-0+11)	+ 73%
Checking of colleague	4 (21%)	13 (68%)	+9 (-3+12)	+ 47%
Information collection	8 (42%)	16 (84%)	+8 (-2+10)	+ 42%

When using paired t-tests we found that the pilots scored significantly better on 'information collection' after receiving the training ($t(18) = -2.650$, $p = .016$).

6.7.2 Participant self-ratings

During the training session on the flight simulator, the crew were asked to rate their own performance in carrying out the startle effect management steps: Unload, Roll, Power. The self-rating was designed to give an indication of the crews own perception of how they applied the URP technique, and its potential usefulness. The participants (n=36) were asked to rate their performance after the first training scenario in the simulator training session (the mid-test scenario) and after the last scenario in the session (the post-test scenario).

The crew were asked to answer the following five questions:

Q1 It took little effort to pick the moment to apply the UNLOAD

Q2 It took little effort to execute the UNLOAD steps

Q3 By executing the UNLOAD steps I felt less haste/time pressure

Q4 It took little effort to perform the ROLL (observing without interpretation)

Q5 Executing the U and R positively influenced the POWER process (decision making)

Answers could be provided using a scale from 1 (Totally disagree) to 6 (Totally agree):

1	Totally disagree
2	Disagree
3	Slightly disagree
4	Slightly agree
5	Agree
6	Totally agree

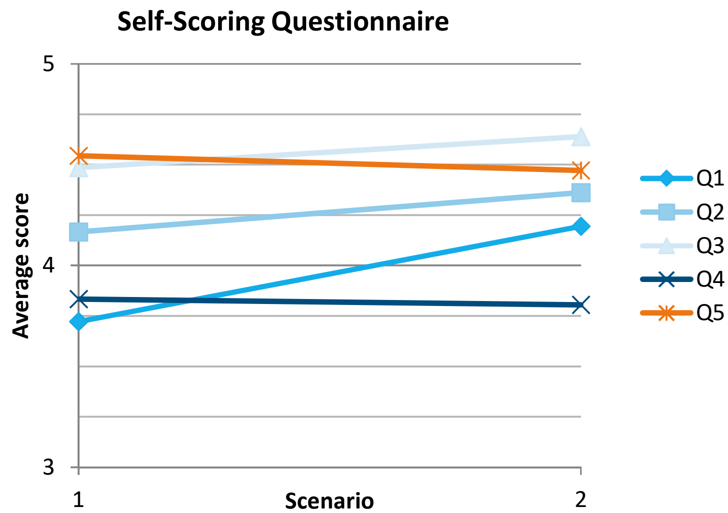


Figure 12 Participant self-scoring results

The results of the crews indicate that overall they perceived an improvement in their ability to apply the URP technique after the simulator training session. This is indicated by the improving trend for the scores in Q1, Q2 and Q3 that are related to the URP technique itself, particularly with reference to the initial management of the Startle effect.

For Questions 4 and 5, which are related to the later stages of the process, there is effectively no difference between the first and last scenario that they were asked to rate. Question 5 was focussed at the decision making process. However, due to time constraints, most scenarios were stopped before the full decision making process was completed. This might have affected the scores for Q5.

Questions 3 and 5 of the self-evaluation related to the outcome of performing the ‘URP’ process. The results indicate that the participants felt that the UNLOAD step had a positive effect on their management of Startle and Surprise. However, it is interesting to note that the results for Q5 are marginally lower after the second scenario compared to the first scenario directly after the classroom training, which may be due to the effect of carrying out the step several times on the simulator. The overall perceived effect of performing the U and R steps was still positive.

6.7.2.1 Analysis of Short-haul and Long-haul pilots

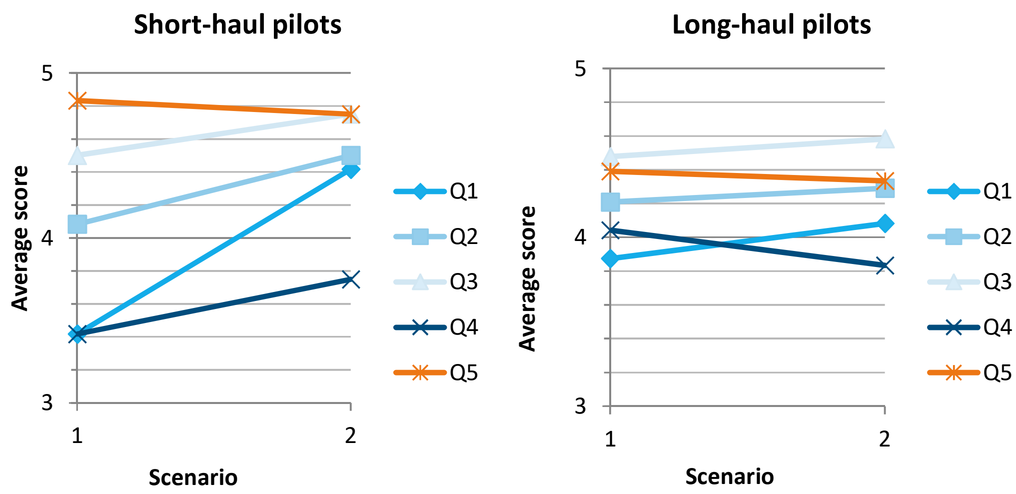


Figure 13 Participant self-rating results (Short-haul vs Long-haul)

It is interesting to examine the difference between short and long haul pilots. There was a more marked difference in perceived performance for the short-haul crews.

6.8 Training evaluation

The pilots were asked to complete a questionnaire to evaluate the training immediately after the training session (see 0). On a scale of 1 to 10 the pilots rated the training with an average of 8.3 (SD 1.0), which could be interpreted as pilots appreciating the training. It was interesting to see that experienced pilots (≥ 10000 flying hours) provided higher ratings than less experienced pilots: 8.7 (SD 0.9) compared to 8.0 (SD 1.0). Other discriminations between short- and long-haul pilots, or instruction experience, did not show this kind of difference.

When asked to rate the classroom session and the simulator session separately there was hardly any difference between the ratings: 8.0 (SD 1.0) for classroom, compared to 8.3 (SD 1.0) for simulator. Also, the connection between classroom and simulator was graded positive on a scale of 1-4, namely 3.7 (SD 0.6). From these ratings and small standard deviations the conclusion can be drawn that the training was received positively by all different subgroups within the complete group of participants.

Pilots were asked to score their perceived level of preparedness on a scale of 1 to 4, where ‘1’ means they don’t feel better prepared, and ‘4’ means they feel very well prepared for dealing with startle & surprise situations. Of the 44 pilots one (2%) was not feeling better prepared, five were feeling ‘a little bit better prepared’ (11%), thirty-one pilots felt ‘well prepared’ (71%) and seven felt ‘very well prepared’ (16%). From this it can be concluded that the training increased self-efficacy concerning dealing with startle & surprise by the majority of pilots.

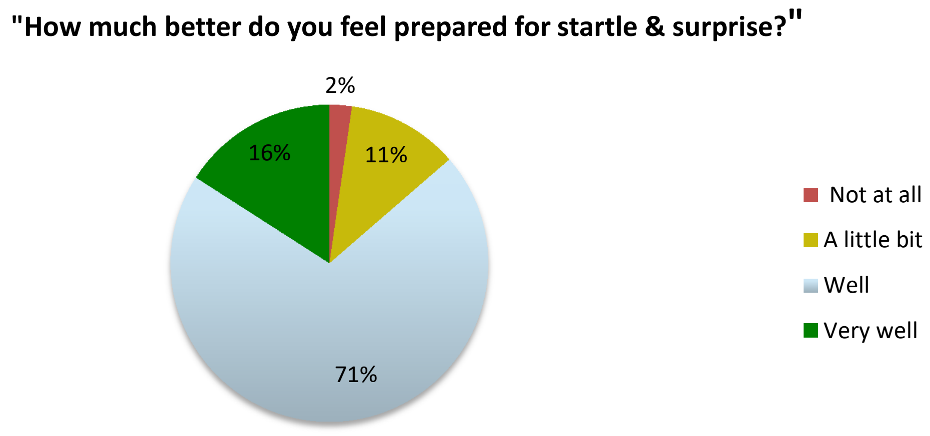


Figure 14 Participant self-efficacy assessment

Training was appreciated by pilots, self-efficacy was increased, but the question remains if they intend to use the trained startle management technique in the operation. Pilots were asked to score this intention on a scale of 1 to 4, where '1' means 'no' and '4' means a 'yes'.

Of the 44 pilots one (2%) did not think he would use the proposed technique (the same one that did not feel better prepared), four answered 'maybe' (9%), eleven thought they would 'probably' use the technique (25%), and 28 answered 'yes' (64%). This means that the majority of pilots found the technique valuable and practical enough to incorporate in their operational routine. Little variation was found between pilot groups.

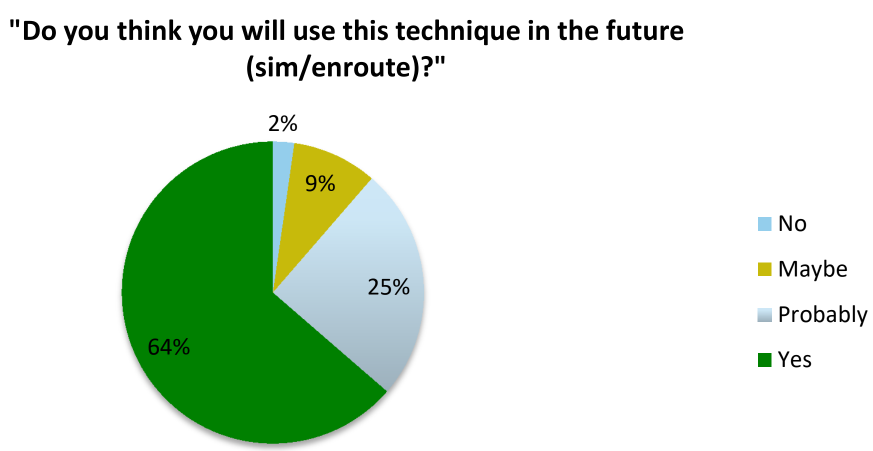


Figure 15 Participant proposed use of technique

Furthermore questions were asked about the different elements of the training:

Table 23 Participant assessment of training elements

Question:	**Average of all pilots:**
Do you think the reaction test was useful?	3.1 (SD 0.9)

(1 no… 4 yes)	
Do you think the PowerPoint presentation was useful? (1 no… 4 yes)	3.4 (SD 0.7)
Do you think the URP training was useful? (1 no… 4 yes)	3.6 (SD 0.5)
Do you think the visualization of the URP technique was useful? (1 no… 4 yes)	3.4 (SD 0.8)
Do you think the chosen approach in the simulator was effective (a few short events/scenarios with a focus on coaching and feedback)? (1 no… 4 yes)	3.9 (SD 0.3)

Approximately two months after the experiment follow-up questionnaires were sent to all pilots involved (see Appendix G), of which 18 responded. They were all sharing their training experience with colleagues. Effects of startle and surprise were discussed, as well as the trained URP-technique. Compared to the situation before the experiment, 13 out of 18 respondents reported an increased awareness of startle & surprise effects in the operation.

It was interesting to see that 50% of the respondents experienced a startle or surprise event in the two months after the experiment. Unfortunately, only one of them used the URP technique to control emotions (the others could not give a reason why they did not use the technique). However, the technique was used in other circumstances by seven respondents which indeed had the calming effect.

6.9 Discussion of the evaluation experiment

The aim of this experiment was to validate the training approach that was developed in the project by putting line crews through a realistic training session and evaluating their performance in startle and surprise management techniques. The training session was put together so that it would take up half of a typical recurrent training slot, using a combination of classroom and simulator instruction and including an element of coaching on the simulator. In this the training introduced a number of new aspects for the participants besides the S&S "URP" technique itself. The training session included a mental visualisation exercise, direct coaching in the technique from the instructor, and an emphasis on the process and technique (rather than completion of the event scenario). This combination of different elements may have made the training session a novel experience for some participants.

The intended effect of the technique that was developed in this project was to assist flight crew in managing the effects of a startle or surprise moment so that they can manage the event more effectively. Key elements of that are better information collection, crew teamwork, effective decision making and improved self-control. While the simulator and test scenarios in the experiment were not designed or intended to produce a major startle/surprise effect, they were intended to produce sufficient cues from the pilots such that the URP technique would assist the crews in managing the events.

The overall feedback from the crews that participated in the training sessions was positive at the end of the session. In some cases there had been scepticism about the usefulness of the training, and the techniques that were being presented; this was particularly expressed during the classroom training session. The pilots were enthusiastic about the technique that was introduced in the training session, describing it as a structured approach to older wisdoms such as "sitting on your hands" and "take a moment" that are heard in training.

The first training session taught us a very valuable lesson, namely to manage expectations. When hearing ‘startle’ and ‘simulator’ in one sentence most pilots will expect very serious scenarios they have never seen before. They expect to be startled. When this did not happen, one of the pilots in the first session was unpleasantly surprised; ‘where were those really heavy scenarios that provide the emotions that make me feel the need to use the URP technique?’ This was the only pilot who did not feel better prepared for startle and surprise situations and would not use the URP technique in the operation. Because pilots expect unexpected events to happen in the simulator, it is almost impossible to produce real life startle and surprise effects. Also, in the simulator potential life threatening situations will never produce the same emotions as in real life. That is the point of simulator training; a safe training environment. No matter how good the training content is, if expectations are not met, this will seriously affect acceptance and buy-in. After this first session we explained clearly what was and what was not going to happen during the simulator training and the reasons for doing so.

On a scale of 1 to 10 the pilots rated the training with an average of 8.3. This can be considered quite high for an experimental training, especially one focussing at non-technical skills. These skills are accepted in the flying community but in discussions it seems technical skills are still seen as the most important. Obviously, technical skills are the basis of a good pilot skill set. But is seems that more experienced and especially instructor pilots rate the importance of non-technical skills higher. This could be the reason for the slightly higher rating (8.9) from instructors compared to line pilots (8.0). Having more instructors in the sample (50%) than in the general airline population (around 10%) might have produced a slight positive bias in the overall rating. On the other hand, instructors might be expected to be critical about new training initiatives due to their experience both on the receiving and the providing end. The overall high rating from both instructors, and from the participants as a whole, bodes well for the introduction of a training session like this. It indicates that despite the focus on “soft skills”, and the slightly novel training methods applied (from a pilot’s perspective), the training has been accepted by the participants and one can expect it to be expected in the operation. The acceptance by the participants of a training program is an important contributing factor to the success and effectiveness of training.

It is widely accepted that physical training needs time and involves many repetitions to yield benefits and increase performance. However, when it comes to cognitive training, it seems this is not as widely accepted. Changing (cognitive) behaviour takes time just as building muscle strength and stamina. Trying to scientifically prove performance increase in an already well trained operational flow after a three hour training session is possibly a mission impossible. This is what a lot of the data shows us; the concept is accepted, after some practicing it is integrated in the way of working but it does not yet lead to significant performance improvements (except on ‘information collection’). On the other hand, it might lead to significant performance improvements outside the simulator, where it really counts. Many aspects of this experiment are promising: the evidence it works in other domains, the high level of acceptance by the pilot community and the proof that it is possible for pilots to integrate the trained techniques into their operational routine. We observed many pilots struggling with the timing of the UNLOAD steps. This was not just about learning something new but about unlearning; to not immediately assess the situation and/or execute a procedure. We see clear training effects from the self-ratings (2.4.2) where especially the short haul pilots show significant increases in perceived performance when it comes to the timing but also the execution of the UNLOAD steps (Questions 1 and 2). Also encouraging is the feedback about the effect from executing these steps; less haste (Question 3).

The overall agreement during the debrief discussions was that training these techniques should be started immediately during ab initio training. This is the only way to guarantee it becomes a natural and automatic way of dealing with startling and surprising situations. Just as has been done with the safety culture that is now an integral part of aviation.

6.9.1 Instructor observations

When comparing the observations from the pre-test scenario (before any training) with the observations from the post-test scenario (at the end of the simulator session), we can gain insight in training effectiveness. The most noteworthy result is the progress in ‘information collection’ (+42%), which we consider of utmost importance to prevent crews ‘jumping to conclusions’. This progress was seen more with long haul pilots than with short haul pilots, more with inexperienced pilots than with experienced pilots, and more with line-pilots than with instructors.

The indicator ‘checking of colleague’, which is important when we consider that working together is especially desired after a surprising or startling event. If no attention is paid to this factor a ‘split cockpit situation’ (where pilots are less coordinated in their response) might occur. Furthermore ‘checking of the colleague’ is desired when we consider that time is spent not toward the stressor (the situation), but away from the stressor. In this time no hasty actions can be performed which can be detrimental to the flight situation. Again, the progress was seen more with long haul pilots than with short haul pilots, more with inexperienced pilots than with experienced pilots, and more with line-pilots than with instructors.

We should be careful generalising these results, as simulator progress shows only a 1% increase with a lot of spread (pilots who performed well, did not do so later on, and vice versa). This suggests that the training effect on this indicator is not very stable after one training session. However, the goal of training is not to improve performance in the simulator, but on the aircraft in the operation. Longer term operational data will have to prove the effectiveness. The intention of the training proposed is for the startle effect management technique to be applied to a variety of different situations, and that it can therefore be repeated in all recurrent training sessions. In this way it is repeated and emphasised by the instructors, and the application of the technique can be increased over time.

6.9.2 Eye-tracking observations

The examination of the eye-tracking data for a number of participants was a useful addition to the instructor SHAPE observations and the participants’ own ratings of their performance. The eye-tracking data can be used to give an objective view on the performance of the crew. The full overview of this examination is reported in Appendix H. Although the sample sizes are low, and therefore it is not possible to make statistical conclusions, some general trends can be observed. The crews all clearly observed the different roles of the PF and PM after the failure event, which influences the pilot’s scan pattern as expected. In contrast to the PM, the “Unload” step did not appear to have an effect on the scan pattern of the PF: the area of interest remains focused on the PFD. The results from a number of the PM pilots indicate that the unload step was useful in the training, and did have an effect on their scan pattern in different ways. This suggests that it is possible to train this, and that it does have an effect, even after a relatively short training session.

6.9.3 Differences between the pilot categories

The setup of the investigation into Startle Effect Management included considering whether there was a difference between short-haul and long-haul pilots, and to take the pilot experience into consideration. Additionally, as this was an evaluation of the training, the instruction experience was also examined in the results. The results of the analysis between the pilot categories are reported in Appendix E.

Of the 44 participants, there was a relatively high percentage of instructors, which is not necessarily representative of the broader fleet at the airline. The same can be said for the distribution of experienced pilots, vs less experienced pilots. While a participant group of over 40 pilots is a good size for an analysis of this type, there are inevitable possible biases introduced into the data through the different participants, and the nature of asking for voluntary participation.

In general the results indicate that there was more progress seen in the results of the long-haul pilots compared to the short-haul pilots. Similarly, the results indicate that the training session benefitted inexperienced pilots and line-pilots more, which are indications that are in line with general expectations. However, the spread between the different groups was not so large. Also worth considering is the role the aircraft type may have in the differences. The Boeing 747-400 that was used for long-haul pilots is a highly redundant aircraft, so flight crew would anecdotally explain how the effect of many failures may be limited compared to the Boeing 737-NG aircraft (which we used for the short-haul crews). Therefore, 747 pilots perceive less time pressure and are less inclined to take immediate action.

Experienced pilots also often indicated during the classroom training that they agreed with the principles behind the training and that they were in line with techniques that they already used in their operations. That said, it was also true that some of the more experienced pilots had difficulty with the “UNLOAD” technique; timing the execution of the technique and applying it in the way that it was being trained.

6.9.4 Self-evaluation of URP process

The participants’ results indicate that the pilots were able to integrate the different steps of the Startle and Surprise “URP” process into their usual way of working, and that this has a positive effect in their opinion. Additionally, there is a general trend for the scores to increase across the training, supporting the perception that a learning process is taking place during the simulator session. As can be seen from the results, there is a slight difference in the participants’ ratings between the short-haul and long-haul crews.

The long-haul pilots start the first scenario with relatively high scores, and these hardly increase after the training. The short-haul pilots however started the first scenario with comparatively lower scores, increasing more over the training session to reach the levels of the long-haul pilots.

Looking at the self-evaluation questions related to the URP process itself, it is possible to see how much effort was required to time and perform the U and R steps appropriately in the scenarios. The results indicate that the pilots are able to select an appropriate moment to execute the ‘UNLOAD’ steps without too much effort. Here again it is interesting to note the difference between short-haul and long-haul crews, that may also be related to the specific aircraft types used in this experiment – the B737-NG and B747-400 respectively. The increase in scores on Q1 (Picking the moment for UNLOAD) is considerably higher for the short-haul pilots compared to the long-haul pilots. Reasons for this may be:

- Due to the higher redundancy of the long-haul aircraft (systems, engines), the perceived time pressure is less compared to the short-haul aircraft.
- Although the events in the first scenario are the same for the two types, the timing of the events differs. The uncontained engine failure almost immediately causes an explosive decompression in the short-haul aircraft, while in the long-haul aircraft (due to the larger fuselage) the decompression warning appears more than a minute after the engine failure. Having two failures at almost the same time might make it more difficult to pick a moment for the UNLOAD.

The expectation was that the 'UNLOAD' step of the process would require more effort than the 'ROLL' step, since it was presumed that the elements involved in this step (breathing, taking physical distance, explicitly checking your colleague) are slightly more unusual than the elements of the 'ROLL' step. However, the results indicated that the reverse was true and that the 'ROLL' step was less applied by the participants. This may be related to the training, since there was more effort on the management of the direct Startle and Surprise effects, placing more emphasis on the 'UNLOAD' step. These first results may indicate that the 'ROLL' step should also be actively included in the training.

6.9.5 Training evaluation

Participants were asked to evaluate the training session and how it could be applied broader within the airline. The overall rating from the participants was very positive. There were a couple of negative reactions, and several constructive feedback comments. Frequently heard comments were: "when can we expect to see this in the recurrent training?" and "when does everyone get to do it!"

In terms of the actual preparation for a startle and surprise event after this training, the self-efficacy plays an important role in this. A more important question is whether pilots feel better prepared for startle & surprise after this training, a question concerning self-efficacy (see (Bandura, 1977)). Self-efficacy could mitigate the effects of fear-potentiated startle, which can have 'severe implications for problem solving, decision making and situational awareness' (Martin et al. 2015). Other research shows a positive relationship between self-efficacy and task performance, for instance in the field of sports (Hepler & Chase, 2008). The results indicated that their self-efficacy was increased, which is a positive result for the project. The crucial last step in the evaluation by the participants was the result that the majority of the participants indicated that they would use this technique in the future during operations.

However, this positive intention did not seem to match the feedback that we received from some participants a few months after completing the training. Even though startle and surprise events were encountered, only a small portion of the participants applied the trained techniques. This may be due to several factors, such as the relative short and one-time training intervention, a second crew-member without the same training experience, and possibly due to some of the startle/surprise effect itself. This relatively low operational application-rate serves to underline the suggestion that the training is most effective when it is done by all of the crews, and when it becomes part of the standard response to events. Therefore, it should be repeated regularly during other training sessions as well.

6.10 Conclusions from the evaluation

Pilots demonstrated progress in applying the startle management technique after both the classroom and simulator sessions. The pilots viewed the training session in startle management techniques as a positive training experience and intended to apply the techniques learned in their daily operation. The result was that pilots were better able to manage an unexpected situation.

The evaluation of the startle management technique, and the training session that was developed, demonstrates that this technique can be trained in a relatively short training session and that the pilots believe that it is a useful technique. The pilots rated the training highly (average rating 8.3 on a scale of 1-10), and demonstrated a positive self-efficacy in handling startling and/or surprising events during and after the training session.

The training session was designed to fit inside “half” of a typical airline recurrent training session, so that either two crews (four pilots) could be trained within one session – as was the case in the experimental session – or so that additional material could be covered within the recurrent training session. After the classroom session pilots already demonstrated progress in key behavioural indicators such as “information collection”, which was further refined by the simulator session. Pilots scored significantly better in “information collection” after the training.

With respect to the specific technique to manage the effect of the startle itself – the “unload” step – the experiment demonstrated that this technique had a positive effect on the pilot’s information collection. Pilots were rated higher after they had taken physical distance in the unload step, and were more likely to take the breathing moment having taken physical distance. Both of these elements of the unload step are critical in the management of the cognitive effects of the startle and/or surprise, so it is very positive that the training experiment demonstrated progress in this area.

The intention of the technique that was developed, and the training session, is that this technique can be applied to all possible failures and unexpected situations, and will therefore be able to be repeated in all recurrent training sessions. Repetition should help to embed the technique in the operation and secure the benefits. Furthermore, if this technique is introduced in ab-initio pilot training, and repeated throughout the pilot’s career, it will become an integral part of dealing with startle and surprise in the cockpit.

7 Cost Benefit Analysis of Startle & Surprise Training

7.1 Cost of S&S training program

The costs for Startle and Surprise (S&S) training can be divided into initial and recurrent costs, some of which are fixed, some of which are variable depending on the airline size. In this case it has been assumed that the airline will invest in an in-house training for startle and surprise. Later in this report the exploitation of startle and surprise will also investigate the alternative possibility of an operator purchasing a training solution offered by another training provider or airline, which of course presents a different cost structure. This section will illustrate what the cost estimates compose of. Several base assumptions which have been used are listed below based on a standard legacy European carrier of about 3000 pilots, others are specified in the respective calculations:

- S&S training consultants cost €200,- per hour
- TRI[4]/TRE[5]/Pilot office days cost €200,- per person per day (8 hours)[6]
- TRI/TRE/Pilot training days cost €1000,- per person per day (8 hours)
- Simulators cost €500,- per hour dry
- 1 pilot FTE costs €300,000
- 1 pilot FTE is 180 working days
- S&S Training (both initial and recurrent) is in addition to existing training
- Airline maintains production capacity at all times
- Pilot office/training days impact production FTE requirements
- Airline has 3000 pilots (including TRI's & TRE's); 600 TRI's (ex. TRE's); 100 TRE's
- Airline does not grow, pilot refresh rate of 3% (assumed 35 year aviation career)

In order to facilitate in-house startle and surprise (S&S) training, an airline must make two initial investments. One is the development of the training, the other is the initial training of the pilot population. Training development will take a four person team a month to complete. This involves defining the recovery method, define a training scheme, design scenarios, and ensure incorporation into the airline training syllabus and training the other TRI's. Costs included are built up as follows:

[4] TRI: Type Rating Instructor
[5] TRE: Type Rating Examiner
[6] Office Day payment

Item	Assumptions	Cost estimate
S&S program development (own office hours)	2 TRI/TRE's; 20 office days;	€8,000
S&S program development (production FTE)	40 days;	€67,000
S&S program development (training consultants)	2 persons; 20 office days;	€64,000
Instructor training	150 instructors; 4 hours classroom; 2 hours sim pp; 10p per classroom;	€249,000
Instructor training (simulator costs)	75 crews; 4 hours per crew;	€150,000
Instructor training (production FTE)	195 days;	€325,000
	TOTAL	**€863,000**

Table 24. Initial costs S&S training development

After the initial training development, the operator must train the entire pilot population before the benefits of S&S training can truly be realised. This initial training is more extensive than the recurrent training in the following years, as it must introduce and manifest new pilot behaviour. Recurrent training must only refresh a pilot's familiarity with the concept.

Item	Assumptions	Cost estimate
Pilot initial training (ex. S&S instructors)	2850 pilots; 2 hour classroom; 1 hour sim pp; 2p per classroom;	€1,567,500
Pilot initial training (simulator costs)	1425 crews; 2 hours per crew;	€1,425,000
Pilot initial training (production FTE)	2137.5 days;	€3,562,500
	TOTAL	**€6,555,000**

Table 25. Initial training costs S&S training

After this initial training, pilots only receive a light recurrent program every year after their initial training. The recurrent training consists only of a classroom training session, without any explicit simulator training (assuming the practical S&S training is then integrated in the standard simulator training). However, as the airline has a refresh rate of 3%, only 97% of pilots receive a refresher training, and 3% receive an initial training. Also, the amount of S&S instructors refreshes with the same rate. All these items are included in the table below and sum to a steady state annual cost for the airline after the initial training in year 1.

Item	Assumptions	Cost estimate
Pilot recurrent training (ex. S&S instructors)	2764.5 pilots; 30 minute classroom; 10p per classroom;	€176,237
Pilot recurrent training (production FTE)	190 days;	€316,667
3% Initial training	90 new initial trainings; €1050 per initial training (including sim);	€94,500
3% Initial training (production FTE)	67.5 days;	€112,500
3% S&S instructor training	4.5 new S&S instructors; €2,660 per instructor; includes sim costs;	€11,970
3% S&S instructor training (production FTE)	5.85 days;	€9,750
	TOTAL	**€721,714**

Table 26. Recurrent training costs S&S training, including 3% pilot refresh costs

These costs will be used in the cost benefit analysis in section 4 of this report, and will be displayed across the span of several years to indicate how a return on investment develops.

7.2 Benefits of a S&S training program

Startle and surprise training can benefit the operator in three distinct ways. The largest benefit is the reduction in the risk of a major (fatal) accident. This is relevant as many startle and surprise cases have proven themselves capable of such disasters (Accident Analysis Report NLR-CR-2016-620). The second benefit is the reduction of unnecessary operational disruptions (e.g. unnecessary diversions) due to greater crew ability not overreact due to startle or surprise, and judge situations more objectively. The third benefit is the exploitation of S&S training by training pilots of other operators at marked up prices.

7.2.1 Reduction of fatal accident risk

S&S training may affect the total fatal accident risk that an airline is subject to at any given time. Although rare, the costs associated with such a major accident are tremendous, and as such a reduction in the risk may represent a significant reduction in the financial risk associated to it. As the calculation of such future effects is an exercise of approximation, both cost and risk calculations are subject to different calculation methods which combine to form a range in which the associated benefit may lie.

7.2.1.1 Current fatal accident risk of S&S

In order to determine the reduction of accident risk, first the current risk of an S&S related fatal accident must be determined. This has been performed in two ways. The first method makes use of database searches related to S&S events, and combines this with EASA accident rate statistics to form a heuristic estimate of the existing S&S fatal accident risk based on statistics. The second method performs a more contextual analysis of several accident modes and makes approximations how S&S training affects these modes and the resulting accident probabilities. By considering these two complementary methods (heuristic vs contextual), the estimate of the associated benefit becomes much more robust.

7.2.1.2 Database heuristic method

As introduced, the first method makes use of database statistics. Kochan et al. (2004)[7] analysed four databases for S&S events, resulting in the following numbers:

[7]

Database	Total no. entries	No. S&S related	Fraction
NTSB	10,597	131	0.0124
NASA ASRS	42,344	424	0.01
FAA AIDS	10,771	30	0.002
ASAP	Unknown	53	Unknown
TOTAL	63,712	583	0.0092

Table 27. Proportions of S&S accidents in various databases (Kochan et al., 2004)

Excluding the ASAP database[8], a total of 585 S&S occurrences in relation to 63,712 total occurrences indicates that around 1% of occurrences is a startle and surprise event. Hence also 1% of fatal accidents will be a S&S event.

The second part of this analysis method requires an estimate for the rate of fatal accidents. According to EASA's Annual Safety Review 2016, since 2005 the accident rate (fatal and non-fata) involving Commercial Air Transport aircraft involving an EASA Member State operator has been lower than 5 accidents per million departures ($5x10^{-6}$). The rate for *fatal* accidents has been below 0.4 per million departures ($4x10^{-7}$) since 2006. For this study, the analysis will limit itself to the occurrence of fatal accidents. Combining this with the above S&S occurrence statistics, it seems that 1% of the $4x10^{-7}$ accidents is related to S&S, hence the S&S fatal accident rate is $4x10^{-9}$.

It should be noted that the database analysis by Kochan et al. (2004) selected the S&S occurrences where S&S were explicitly reported by the crew, thereby not considering occurrences in which S&S may have played a role, but was not reported as such. Therefore, it is suspect that the actual S&S rate is higher than 1%.

7.2.1.3 Contextual analysis method

The second method to estimate the current S&S fatal accident risk involves an analysis of accident events which may be caused or exacerbated by S&S. This analysis makes use of the outcome of an NLR investigation[9] into quantifying commercial air transport risk for the Netherlands Ministry of Transport. This report contains several Event Sequence Diagrams (ESD's) which calculate the probability of benign, serious or fatal outcomes depending on both initiation and mitigation steps. An example of an ESD for spatial disorientation is presented in Figure 1.

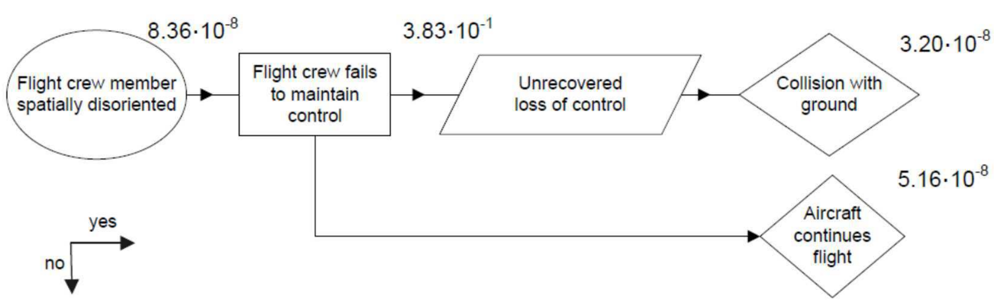


Figure 16. Quantification of an ESD for spatial disorientation (NLR, 2006)

[8] The ASAP database was excluded since a percentage cannot be calculated as it is unknown how many entries there are in the database.
[9] NLR rapport NLR-CR-2006-520

The ESD's which have been selected to be pertinent to S&S are presented in Appendix A:

- ESD11: Fire on board aircraft
- ESD12: Flight crew spatial disorientation
- ESD13: Flight control system failure
- ESD14: Flight crew incapacitation
- ESD15: Anti-ice
- ESD16: Flight instrument failure
- ESD17: Adverse weather
- ESD18: Single engine failure
- ESD19: Unstable approach
- ESD23: Wind shear
- ESD30: Aircraft encounters unexpected wind
- ESD35: Flight crew decision error/operation of equipment

Other ESD's have been omitted because they are either so acute (takeoff roll/landing roll) that S&S does not play a significant role in the outcome of the event, or that they pertain only very little to flight crew.

The outcome "collision with ground" has been summed for all the selected ESD's, as this outcome is representative of a fatal accident, totalling $7.23x10^{-7}$. However, the $7.23x10^{-7}$ rate is not solely due to S&S. Subsequently, all the selected ESD's are analysed to determine how S&S may affect them. This analysis is performed not by directly reasoning the percentage caused by S&S, but by determining how much reduction a hypothetically 100% effective S&S training would realise. S&S training could affect the ESD by mitigating the top-event (initial event), but it could also affect the probability of flight crew recovery from the top-event. The effect of the S&S training was estimated for every ESD and node independently and concluded that 100% effective training would realise a reduction of $4.76x10^{-7}$ of fatal accidents. This in turn concludes that 66% of all fatal accidents are related to S&S. This is based on worldwide accident rates, and thus this 66% must still be imposed on the European fatal accident rate of $4x10^{-7}$, resulting in an S&S accident rate of $2.64x10^{-7}$.

Comparing both methods of approximation, there is a large difference between $4x10^{-9}$ (likely to be an underestimation) and $2.64x10^{-7}$ (based on crude estimates, possibly inflated). The difference in approximations will be taken into accounted in a matrix in the last subsection.

7.2.2 Risk reduction due to S&S training

The risk presented by S&S must be reduced by the S&S training introduced by an airline. The question is how much the risk can be reduced by such training. There are several factors which can increase or decrease training effectiveness:

Increase effectiveness	Decrease effectiveness
Thorough training	Meagre training
Pilot acceptance	Low acceptance
Broad applicability	Scoped to limited situations
Proactive	Reactive

Table 28. Factors which affect training effectiveness

The accident and incident investigation basis of this program, as well as the resulting S&S training designed, both echo that S&S are events with a broad scope of causes and triggers, and hence the training is more generic than task or situation specific. Furthermore, the method is aimed at preventing S&S snowballing at an early stage, and therefor provides a near-proactive method to resolve situations. Feedback from the training sessions are overwhelmingly positive, with most test subjects remaining enthusiastic about S&S training, and feel that it may indeed help them in their operations. The training effectiveness may be affected by more or less investment in the training itself. If implemented as calculated in the previous costs chapter, it should results in superior transfer of training.

In the same article by Kochan et al. (2004), it is mentioned that in half of all S&S cases, training would have helped mitigate the consequences. This reflects similar training-desires mentioned in cases discussed in the S&S accident report of this project.

Reflecting on these considerations, a decision has been made to set the maximum achievable effectiveness of S&S training at 50% (it will never prevent more than 50% of cases). A further conservative factor of transfer or training, estimated at 70% based on Table 5, results in an estimation that the training will prevent or resolve 35% of S&S events. Integrating this factor into the S&S accident rates calculations from the previous section will show that the reduction in the total fatal accident risk lies between $-1.26\text{x}10^{-7}$ (with the contextual analysis) and $-1.4\text{x}10^{-9}$. This represents a reduction in the total fatal accident risk of 31% and 0.35% respectively.

7.2.3 Cost of a fatal accident

The last element required to calculate the safety benefit of S&S training is to estimate the costs for a fatal accident. Here also there are two difference analyses which may be compared in the matrix in the next subsection. The first method relies on an accident cost model developed by NLR in a project which analysed the costs associated with aviation safety[10]. This report indicates that the cost for a fatal accident can be approximated to €223 million.

The cost of an accident is built up from several heads of costs, including aircraft physical damage, loss of resale value, loss of use of the aircraft, costs related to passenger and crew fatalities and injuries and costs related to search and rescue services and accident investigation. Based on the different End States of the ESD, the costs have been determined based on the percentage of aircraft damage and the percentage of occupant fatalities. More details on the cost estimates of the End States of the ESDs can be found in NLR-CR-2008-307.

A second method involves a more airline-centric analysis of the costs of an accident. The notable difference between these methods is that this method assumes the airline is accident-insured for the most part, but sustains economic damage (brand tarnishing, loss of sales, etc.). The estimates are broad, but serve to compare with the previous estimate. Several assumptions[11]:

- 10% loss of net income for 2 years (brand image)
- €400 million income per year
- Total accident cost €200 million
- Insurance deductibles 10%
- Insurance premium 30 million
- Insurance premium +20% after accident, remains for 15 years

The table below indicates how several cost factors calculate and total for the accident costs for the airlines:

Item	Assumptions	Cost estimate

[10] NLR rapport NLR-CR-2008-307
[11] These estimates assume the same airline size as in the cost estimates

Loss of net income	-10%; 2 years; €400,000,000 income	€80,000,000
Insurance deductible	10%; €200,000,000 accident costs	€20,000,000
Insurance premium rise	+20%; 15 years; €30,000,000 premium	€90,000,000
	TOTAL	**€190,000,000**

Table 29. Estimate for fatal accident costs for an airline

The two estimates differ by about €30 million (about 15%), which is not a poor estimate. However, these two estimates will be treated as a high and low estimate in the matrix below.

7.2.4 Financial benefit range

The financial benefits of a risk reduction depend of the reduction of risk, the cost associated to the risk, and the exposure that a particular operator has to a risk. In this case, it is assumed that an airline with 3000 pilots and a net income of €400 million operates about 150,000 flights annually.
Combining the different reductions of risk, different cost estimates, and number of flights, the matrix below indicates a range of benefits, varying from about €40,000 to €4 million.

	Insured situation estimate (low impact) (€190,000,000)	NLR estimate (high impact) (€223,000,000)
Database estimate (low impact) (Δ -1.4x10-9)	**€39,900**	**€46,830**
NLR estimate (high impact) (Δ -1.26x10-7)	**€3,597,999**	**€4,222,915**

Table 30. Estimated financial benefits of a reduction in safety risk

It is highly likely that the database estimate is overly optimistic about the low pervasiveness of startle and surprise events. It is significantly more likely that the ESD analysis provides a better estimate about how many fatal accidents are actually caused or exacerbated by S&S (66%), in particular when combined with the EASA fatal accident rate. Added with a level of conservatism, it is not unrealistic to expect that the annual risk benefits are in the order of €1 million. Which is still only 25% of the high impact estimates. In the cost benefit analysis chapter, several risk benefit situations will be presented in relation to the corresponding return on investment (ROI) estimations. This provides the opportunity for a trade-off between ROI policy and realistic expectations of the reduction of risk.

7.2.5 Operational Cost

Next to the benefits of a reduction in accident risk, S&S training can also prevent crews from overreacting to a situation, and make unnecessary, and costly, diversions. Of course many diversions are fully warranted, and this calculation does not serve to promote not diverting. However, the effectiveness of S&S training may also resound in operational resilience. This analysis will apply the S&S risk exposure and S&S training effectiveness numbers to the context of unscheduled landings (diversions).

In EUROCONTROL's Standard Input (Ref 1), a table is provided with the cost for a diverted aircraft.

Type of flight	Cost of flight diverted (€)
Regional flight	810 - 5,780
Continental flights	1,160 - 8,680
Intercontinental flights	5,780 - 63,600

Table 31. Flight diversion costs (Ref. 1)

Additionally, it is stated in this document that "typical values (based on expert judgement from the SESAR Definition Phase) would be €5,780 for a 120 seat narrow body and €19,670 for a 400 seat wide body". For this analysis an average cost of €10,000 per diversion is assumed based on the above figures. Furthermore, in 2012 0.22% (19,177) of all flights in the EUROCONTROL Network Manager Area made unscheduled landings. In 2014 0.22% (19,390) of flights landed at other airports than initially planned. Using the estimation of 150,000 flights for our model airline, this comes to 330 diversion annually.

From the previous risk analysis, between 1% and 66% of these events could be S&S related, and between 0.35% and 31% of events would be preventable with S&S training. This amounts to approximately between 1 and 102 diversions being preventable by S&S training. Similar to the reflections on the risk analysis, the estimation of 100 mitigated diversions is more likely than 0. A conservative estimate of 25 mitigated diversions results in a savings of €250,000 annually.

7.2.6 Exploitation/outsourcing benefits

The last benefit that an airline may have from S&S training is selling the in-house developed program to external pilots from other airlines. This is an interesting aspect because it provides a direct income which may be used to pay back the initial and recurrent investments in training, and makes an airline less reliant on risk and safety estimations for its investment case. In addition from a purchasing airlines perspective, it may be an interesting solution for airlines which are too small to have a reasonable ROI for the up-front investments of an in-house program. This section will cover both aspects of this exploitation concept.

7.2.6.1 Hosting airline perspective

From a hosting airlines perspective there are again costs associated with providing the training to pilots. Both the initial and recurrent training are provided. For the hosting airline the training costs are mainly instructor, simulator and production FTE costs (for the instructor). The following assumptions are made:

- Training programs are equivalent to in-house program
- 500 initial pilots every year
- 2000 recurrent pilots every year

- Pilots are training in pairs minimum
- Instructors earn 400,- per day for instruction (double fees)
- Simulator costs 500,- per hour (dry)
- Instructor training days impact production FTE requirements
- 1 pilot FTE costs €300,000
- 1 pilot FTE is 180 working days

Item	Assumptions	Cost estimate
Initial training (instructor fees)	500 pilots (250 crews); 2 hour classroom; 2p per classroom; 1 hour sim pp;	€50,000
Initial training (simulator costs)	250 crews; 2 hours per crew;	€250,000
Initial training (production FTE)	125 days;	€208,333
Recurrent training (instructor fees)	2000 pilots; 30 minute classroom; 10p per classroom;	€5,000
Recurrent training (production FTE)	12.5 days;	€20,833
	TOTAL	**€534,166**

Table 32. Total costs for outsourcing training (500 initial, 2000 recurrent)

The respective costs are €1017,- per pilot for initial training, and €13,- for the recurrent session. With a sales markup raising the prices to €1400,- per pilot for an initial and €20,- for a recurrent session, the benefits for the airline are €740,000, resulting in a net profit of €205,834. This may seem like a large markup but the prices are not very unrealistic as seen from a purchasing airline's perspective. Note that this exploitation may likely only be possible after 1 or 2 years after the own pilot body has been initially trained.

7.2.6.2 Purchasing airline perspective

From the perspective of a purchasing airline, the cost basis is simply the training fee's paid and the production FTE's to be offset. In order to illustrate the case for a smaller airline, the following assumptions are made:

- The airline has 500 pilots requiring initial and recurrent training
- The airline does not train internal instructors
- Training days impact production FTE requirements
- 1 pilot FTE costs €300,000
- 1 pilot FTE is 180 working days
- Pilot training days cost €1000,- per person per day (8 hours)
- Travel costs not included

The table below illustrates the cost base for the airline in case they decide to outsource their S&S training, both initial and recurrent:

Item	Assumptions	Cost estimate
Initial training	500 pilots; €1400,- per pilot training; 4 hour sessions;	€700,000
Initial training (production FTE)	500 pilots; 4 hour sessions;	€416,666

TOTAL	€1,116,667

Table 33. Total cost for outsourcing initial S&S training for 500 pilots

Item	Assumptions	Cost estimate
Recurrent training	500 pilots; 20,- per pilot; 30 minute session;	€10,000
Recurrent training (production FTE)	500 pilots; Assume 1 hour;	€52,083
	TOTAL	**€114,166**

Table 34. Total cost for outsourcing recurrent S&S training for 500 pilots

Comparing this to a mock calculation of this airline developing their own training, they will incur the following costs, based on quick ratio's from the calculations in chapter 1:

Item	Assumptions	Cost estimate
S&S program development	Own pilots; Consultants; FTE;	€139,000
Instructor training	20 instructors; FTE;	€96,533
Initial training	480 pilots; sim; FTE;	€1,104,000
Recurrent training	500 pilots; FTE;	€89,149

Table 35. Quick calculation of in-house costs for 500 pilots

The one time net savings from the initial training is €12,667 and the net savings annually for recurrent training is €25,017 (this even assumes that pilots are spending more time on this than an in-house solution). However, the additional costs of training development are €235,533 so it takes nine years for such an in-house solution to pay off. Assuming that after four years the recurrent sessions will be hosted in-house, it is more effective for this airline to outsource. Note that many of these decisions also depend on the FTE demand that training incurs. If pilots work more than 180 days or are paid less, then this may in turn warrant an in-house solution.

7.3 Cost benefit analysis

Combining the costs from chapter 2 and the benefits from chapter 3, this chapter will attempt to illustrate how airlines are able to make a decision about whether investing in S&S outweighs the costs involved. There are several core variables which are to be interpreted in this analysis:

- What is a realistic estimation of S&S exposure? 2.64×10^{-7} or 4×10^{-9}?
- Related, how many fatal accidents will be mitigated by S&S training? 31% or 0.35%?
- How does S&S training exploitation affect a return on investment?

As illustrated in chapter 3, there are two estimates of the effectiveness of startle and surprise training, varying by two orders or magnitude (31% and 0.35%). Most likely the actual value will lie somewhere in between these two values. From a financial perspective, for a given return on investment, the risk exposure must be much higher to warrant a 0.35% effective training, than a 31% effective training. In other terms, the period between fatal S&S accidents must be very small for a 0.35% effective training to have enough financial impact to warrant the investment, while a 35% effective training will be financially balanced at a much larger period between fatal S&S accidents. For example, for a return on investment of 5 years, at 31% the airline must have an S&S fatal accident every 40 years, but at 0.35% they would require one every 139 *days* (0.38 years)! Most likely the value lies somewhere in between. This will be illustrated in the table below.

The table below also considers the effect of S&S exploitation, and considers three scenarios:

- low outsourcing (100 initials; 500 recurrent; €41.8K annual profit)
- medium outsourcing (250 initials; 1000 recurrent; €102.75K annual profit)
- high outsourcing (500 initials; 2000 recurrent; €205.5K annual profit)

In addition there is a scenario involving only In-House Training (IHT), and a scenario considering the purchase of outsourced training at the priced indicated in chapter 3. All scenarios consider an airline of 3000 pilots with an initial training in year 1 and recurrent training annually thereafter, and 150,000 movements per year.
The table is an aid in determining whether an investment should be made or not. The airline in question determines the scenario most appropriate, as well as a predetermined ROI (based on policies) and can look up what the maximum allowable period is between fatal S&S accidents (expressed in years). For example (as indicated in green): an airline requiring an ROI within 5 years and intends to heavily exploit S&S training, finds a period of 40 years. If an airline expects to have such an accident in less than 40 years, then they must invest. If they expect to have less S&S fatal accidents, they should not invest. The floor threshold to invest (based on 0.35% effectiveness) varies between 31 (ROI 1 year) and 271 days (ROI 20 years) and as such is regarded to be zero. Therefor only the ceiling limit (determined by 31% effectiveness) is used to base the investment decision on.

Note that the different exploitation scenarios do not differ in large amounts, and therefor may increase the maximum tolerable S&S period by a few years, but will not make tremendous changes in the ROI period. This is due to the fact that the contribution of such exploitations are only a fraction of the mitigated accident costs.

ROI Year	Only IHT	IHT + low OT	IHT + med OT	IHT + high OT	Purchase Training
1	9	9	9	9	10
2	16	17	17	17	18
3	23	24	24	25	26
4	30	30	31	33	33
5	36	36	38	40	39
6	41	42	44	47	45
7	46	47	50	54	51
8	50	52	55	60	55
9	54	57	60	66	60
10	58	61	65	72	64
11	62	65	69	78	68
12	65	68	73	83	71
13	68	72	77	89	75
14	71	75	81	94	78
15	74	78	85	98	81
16	77	81	88	103	83
17	79	84	91	107	86
18	82	86	94	112	88
19	84	89	97	116	91
20	86	91	100	120	93

Table 36. Minimum S&S fatal accident period, given a specific S&S training regime and ROI year

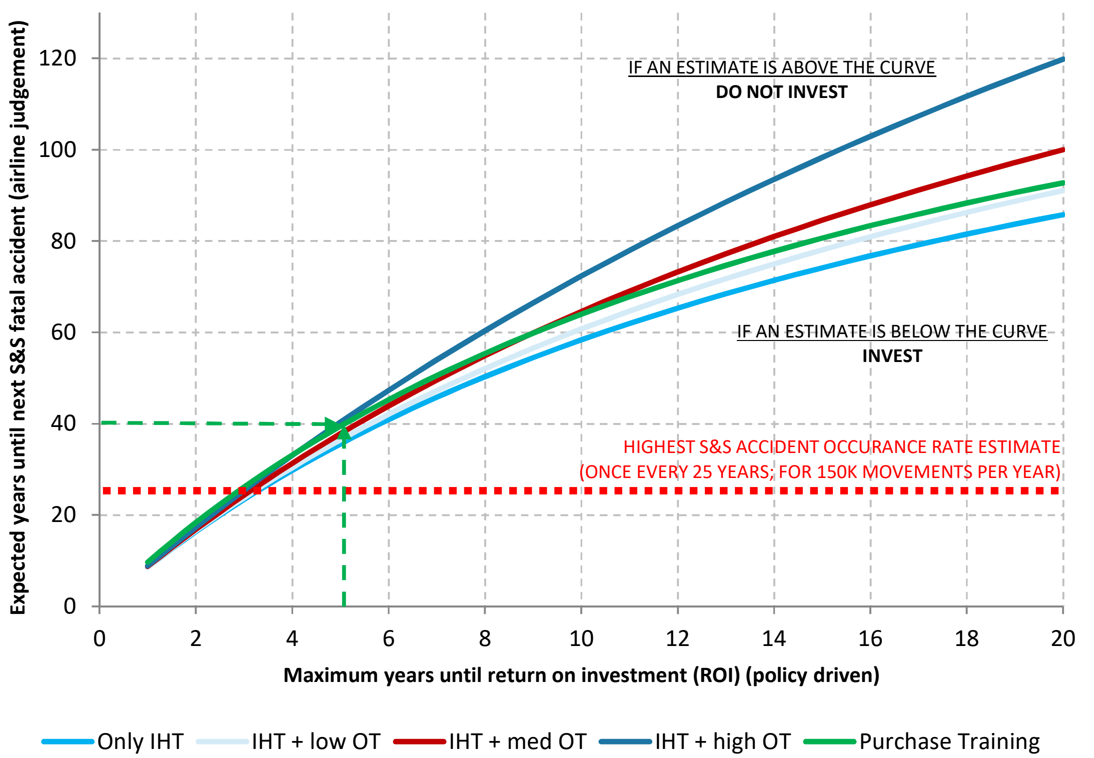


Figure 17. S&S training ROI vs expected S&S fatal accident period

A further illustration of this analysis can be found in figure 2. In this figure the same five scenarios are plotted against ROI (x-axis) and the maximum period between S&S fatal accidents. As indicated, if an airline is determined to lie below the curve, then investment is recommended. If the airline lies above the curve, then it is unlikely that a return on investment will happen *given the associated ROI expectations and S&S period expectations*.
The green value indicated in Table 13 is also depicted with two green dotted arrows. Note that the same logic applies: if the airline reasons that they might have a fatal S&S accident before 40 years, they then lie under the curve and should invest in S&S training.

Another important consideration is that, given the highest estimate of the S&S accident rate, the average period between fatal S&S accidents for an average European airlines is 25 years ($2.64x10^{-7}$; 150,000 movements per year), and is indicated by the dotted red line. The lowest estimate of the S&S accident rate is 1667 years ($4x10^{-9}$; 150,000 movements per year). This latter limit is not indicated as it is by many accounts not a good estimate. In light of these estimates, it may be concluded that an average European carrier will feature a return on investment no earlier than about 3 years. This is a lower limit, the airline is free to judge whether they believe a higher period is suitable, and determine if they accepted the associated ROI or not.

7.4 Cost-Benefit Analysis Conclusions

In light of the above chapters, this report concludes that S&S training can indeed have a return on investment. However, there is a distinct tradeoff between the airline's financial policies and endurance, and the perception of the S&S risk. This report has attempted to build an approximation model to permit well founded discussions about the implications of S&S and mitigation efforts. As such, the assumptions in this report and the tables and graphs in chapter 4 are intended to provide insight into decision making about training, rather than mandate implementation; this is not possible as both risks and finance of airlines differ so greatly it cannot be generalised.

The development of future aircraft, Communication Navigation and Surveillance - CNS, procedures and training, all impact the presence and risk of startle and surprise. Evidence from accidents in the past 10-15 years do indicate that the presence and risk of S&S are more likely to be on the rise than on the fall, and as such this chapter will conclude with the advice that airlines do not observe the above cost benefit analysis as a set of static estimations, but must seriously consider that the proportion of fatal accidents due to startle and surprise will much more likely approach the high limit of 31%, instead of the low limit of 0.35%.

In the end, the tangible investments in training today may primarily reduce operational risk, and therefore only be accounted by the fact that some negative event does not occur in the future. This is difficult to fully appreciate, and an even more difficult case to found financially. However, one certainty may be stated: as the operational risk of S&S increases and the high end estimations in this chapter become less unrealistic, it is sure that today's training will mitigate less and less of the risks. Therefore, the period until the next S&S fatal accident decreases is like a candle burning from two ends: the natural progression of time on one side, and the increase of S&S risk on the other. Will an airline wait until this candle is burned and only at that terrible instant realise the large and growing ROI which was present all this time, or will the airline perceive change as a constant, and invest in S&S training before the ROI has solid, yet terrible, proof of existence?

8 Bibliography

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bürki-Cohen, J. (2010). Technical challenges of upset recovery training: Simulating the element of surprise. *AIAA Guidance, Navigation, and Control Conference.*

Casner, S. M., Geven, R. W., & Williams, K. T. (2013). The Effectiveness of Airline Pilot Training for Abnormal Events. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*(3), 477-485.

EASA. (2016). *Annual Safety Review.* Cologne.

Ekman, P., Friesen, W., & Simons, R. (1985). Is the startle reaction an emotion. *Journal of Personality and Social Psychology, 49*(5), 1416-1426.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

EUROCONTROL. (2013). *Standard inputs for EUROCONTROL Cost Benefit Analyses, version 6.0.* EUROCONTROL.

Facione, N., & Facione, P. (2007). *Thinking and Reasoning in Human Decision Making: The Method of Argument and Heuristic Analysis, .* Milbrae, CA: California Academic Press.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human perception and performance, 1*(3), 288.

Glaser, E. M. (1972). *An experiment in the development of critical thinking.* Columbia University, Teachers College. New York: AMS Press.

Hawkins, F. (1987). *Human Factors in Flight* (2nd ed.). Aldershot, UK: Ashgate.

Hepler, T. J., & Chase, M. A. (2008). Relationship between decision-making, self-efficacy and the performance of a sport skill. *Journal of Sports Sciences, 26*(6), 603-610.

Horstmann, G. (2006). Latency and duration of the action interruption in surprise. *Cognition and emotion, 20*(2), 242-273.

Klein, G. (1998). *Sources of power: how people make decisions.* Cambridge, MA: MIT Press.

Koch, M. (1999). The neurobiology of startle. *Progress in Neurobiology, 59*, 107-128.

Kochan, J. (2005). *The role of domain expertise and judgment in dealing with unexpected events, doctoral dissertation.* USA, Florida, Orlando: University of Central Florida.

Kochan, J., Breiter, E., & Jentsch, F. (2005). Surprise and unexpectedness in flying: Factors and features. *13th International Symposium on Aviation Psychology.*

Lang, P., Bradley, M., & Cuthbert, B. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*(3), 377-395.

LeDoux, J. (1997). Emotion, memory and the brain. *Scientific American, 7*(1), 68-75.

Martin, W., & Murray, P. (2013). Training Interventions for Managing Startle During Unexpected Critical Events. *66th International Air Safety Summit.* Flight Safety Foundation.

Martin, W., Murray, P., & Bates, P. (2012). *The effects of startle on pilots during critical events: A case study analysis.* D. H. Hogreffe. (Ed.).

Martin, W., Murray, P., Bates, P., & Lee, P. (2015). Fear-Potentiated Startle: A Review from an Aviation Perspective. *The International Journal of Aviation Psychology*(25(2)), 97-107.

Meyer, W., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion, 21*(3), 251-274.

Rivera, J., Talone, A., Boesser, C., Jentsch, F., & Yeh, M. (2014). Startle and surprise on the flight deck: Similarities, differences, and prevalence. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting*, (pp. 1047-1051).

Roelen, A. L., van Doorn, B. A., Smeltink, J. W., Verbeek, M. J., & Wever, R. (2006). *Quantification of Event Sequence Diagrams for a causal risk model of commercial air transport, NLR-CR-2006-520.* Amsterdam: NLR.

Roelen, A., & J.W.Smeltink. (2008). *Accident Costs for a Causal Model of Air Transport: NLR-CR-2008-307.* Amsterdam: NLR.

Sapolsky, R. (1994). *Why zebras don't get ulcers.* USA, New York: WH Freeman.

Simons, R. (1997). *FACS in the Study of the Latah Syndrome.* Paul Ekman Erika L. Rosenberg, 38.

Wickens, C. D. (2001). Keynote address: Attention to safety and the psychology of surprise. *11th International Symposium on Aviation Psychology* (pp. 1-11). USA, Ohio, Columbus: The Ohio state university.

Project reports

Boland, E.J. & Field, J.N., (March 2017), Literature review on startle, surprise and acute stress, NLR-CR-2016-619

Mohrmann, J.F.W. & Field, J.N., (April 2017), Review of startle and surprise role in incidents and accidents, NLR-CR-2016-620

Boland, E.J. & van Rooij, J.M., (April 2017), Review of existing training methods, NLR-CR-2016-621

Boland, E.J., van Rooij, J.M. & Field, J.N., (April 2017), Definition of startle and surprise effect management training, NLR-CR-2016-622

Mohrmann, J.F.W. & Smeltink, J.W., (June 2017), Cost benefit analysis of startle and surprise training, NLR-CR-2017-102

Appendix A ASRS Incident Synopses Startle Effect

The following 15 selected ASRS incidents are those illustrating the startle effect, as described in Table 3.

ACN: 1291766

Synopsis:
“A320 series flight crew reports a Runway Overrun Prevention System (ROPS) warning upon landing in BOS on Runway 27. The call out is "runway too short, brake max braking", which the First Officer applies along with maximum reverse. The aircraft is stopped well before the end of the runway and the crew does not understand why the alert was activated.”

Excerpts:
“…Having never received this warning and not anticipating it, the command was startling but followed dutifully. At no point in time did it feel like the runway was insufficient for a safe landing…. The aircraft is stopped well before the end of the runway and the crew does not understand why the alert was activated…”

Analysis:
Onset: False Runway Overrun Prevention System warning (verbal call out “runway too short, brk mx”)
Effect: Immediately adjusted autobrakes (per call out)
Reflection: Still no understanding of the reason for the warning

---

ACN: 1268324

Synopsis:
“ERJ-175 First Officer reported encountering wake turbulence at touchdown in MIA that resulted in a left and right roll so a go-around was executed.”

Excerpts:
“…as I listened to the radar altimeter count down from 20 and 10 at a normal descent rate. That's when I began to experience irregularity with the aircraft handling. Suddenly, the left wing dropped aggressively causing the left main gear to slam to the runway….. At that point, I decided the aircraft was not in control to land and I initiated a go-around. Honestly, it happened so fast that in the heat of the moment I don't remember if I hit the TOGA button or not, but I must have missed the button because I don't recall the TOGA thrust mode becoming active. I did however firewall the thrust levers because of how fast the event was unfolding, and the need for thrust immediately…. We agreed we had probably encountered wake turbulence in the flare. A heavy 767 had landed just prior to our arrival….. The event itself and the go around were a lot more intense than I expected, and I think that adds to part of the confusion my mind faced during the event…. I also learned to always anticipate a go-around from a landing. Even down as far as 10' above the ground - in truly the worst place I would have to go-around, it could happen. Practicing a go around in this condition may be helpful as a sim scenario in the future….. Also, the heat of the moment in the go-around itself. I would have liked to be more "automatic" with the go around procedure. I need to review that mentally more often so when I have to do it again, I'll be completely ready to act than be startled.”

Analysis:
Onset: Wake turbulence during touchdown, gear contacted ground violently
Effect: Immediate Go-around action (“firewall throttles”), go-around was “intense”, no TOGA mode act
Reflection: Too unprepared for very late go-around, desire for automatic pilot GA response also in worst case scenario

---

ACN: 1268324

Synopsis:
“An EMB-175 flight crew reported overshooting their cleared altitude during a go-around that was executed because of an unstable approach flown by the new-hire First Officer.”

Excerpts:
“In the process of conducting New Hire First Officer (FO) Initial Operating Experience (IOE) we were cleared for the visual approach to Runway 18R in DFW…. I called "1,000 feet - UNSTABLE" but the student did not react. He was frozen-up and not reacting as expected. Couple of second later I repeated the "1,000 feet - UNSTABLE" call and as the student again did not respond I called "Go-Around, Flaps 2" and selected flaps 2. Then I called "Positive Rate - Gear Up!" and selected gear up. Student finally contributed to the Go-Around by pressing the Takeoff Go Around (TOGA) buttons. As he did not call for any lateral or vertical modes I announced and selected "HDG" and "FLCH"…… As I was working hard to complete PM and at least half of the PF duties at the same time I did not notice that student was not levelling off at 3,000 feet……. During simulator training students should be exposed more to unplanned and unscripted Go-Around events so they can overcome the startle factor easier and be prepared for a Go-Around on any approach.”

Analysis:
Onset: Unstable approach and “UNSTABLE” callout
Effect: Student frozen, inaction. Late TOGA action, no mode callouts
Reflection: Train more unplanned and unscripted go-arounds, prepare automatic response

---

ACN: 1230172

Synopsis:
“An aircrew mistakenly landed at Flaps 20 instead of Flaps 25 after a miscommunication between the two crewmembers. They received an aural warning "too low flaps", but elected to continue to an uneventful landing.”

Excerpts:
“At low altitude we got a "too low flaps warning" aural warning. This startled me, but I was tired and dialled in on this approach….. The FO then pointed out that the flaps were actually indicating 20. He asked if I wanted to drop them to 25. I felt that would be destabilizing to the landing and I was carrying ref 25 plus about 12 knots at the time so I said no. I added a little power for 3 or 4 extra knots and touched down softly and uneventfully on the 12,000-foot runway….In hindsight the proper thing to do would have been to go around and set up a new and proper approach. I always double check gear down on short final, and will now include a final flap position check. I will make a great effort in the future to go around when the situation calls for it.”

Analysis:
Onset: “too low flaps” alert
Effect: Quickly adjusted speed for less flaps
Reflection: Go-around would have been proper reaction, poor flaps awareness (to be improved)

ACN: 1268057

Synopsis:
“Medium transport flight crew, distracted by weather conditions and lightning on final approach, missed the setting of flaps 30 until reminded by a too low flaps warning. Crew selected flaps 30 and landed.”

Excerpts:
“This combined with storms scattered all along our route, plus fatigue from a very long day, definitely put me in the "Yellow", and I think my FO was there too…. At approximately 400 ft-300 ft AGL, we got the caution, "too low flaps," which startled us and I immediately looked at the flap indicator (25) then gear (down, three green), brakes (armed green light)…. I believe we were both fixated on just seeing the runway, and distracted by the ATC call, and neglected to run the before landing checklist.”

Analysis:
Onset: “too low flaps” warning
Effect: Immediately scanned flaps/gear/brakes
Reflection: Too fixated on other tasks, missed flaps awareness

---

ACN: 1227427

Synopsis:
“On approach to HOU at about 700 feet, flight crew was struck by a green laser light flash. They were able to continue the approach and landed.”

Excerpts:

“An industrial strength type BRIGHT PULSE green laser struck my cockpit…. The laser was a distraction and startled both Pilots. The light burst did interfere with my vision, and both Pilots immediately confirmed the experience to be that of a green laser.”

Analysis:
Onset: Green laser flash in the cockpit
Effect: Whole crew distracted, vision affected
Reflection: none

ACN: 1219870

Synopsis:
“A321 Captain reports experiencing a terrain warning during a night visual approach to Runway 25L at LAS. Evasive action is taken then a left 360 is requested to lose altitude. During this manoeuvre the flying Captain switches of the automation off, becomes task saturated, and passes control of the aircraft to the First Officer.”

Excerpts:
“Passing through 4200 ft we got the "caution terrain alert" we levelled off and confirmed visually that the high terrain was off to our 2 o'clock position and we still had the airport visually. Shortly after we got the "too low terrain, pull up alert". We still had the terrain and airport insight…. It took a minute or so for me to get back in the "green"… The event occurred mainly from fatigue, the "startle factor" of the terrain warning putting us in the "yellow", and accepting a visual approach at night with high terrain…. It was a short, high workload flight completely flown when our circadian rhythm was at its lowest point.”

Analysis:
Onset: Two terrain warnings despite visual terrain contact
Effect: Level off, took time to recover from startle
Reflection: Combined with fatigue the startle had a clear and prolonged disruption on task performance

---

ACN: 1168197

Synopsis:
“ERJ-170 Captain encountered "severe" wake turbulence just before touchdown at DFW in trail of an Airbus A320. A somewhat confused and non-SOP go-around was executed.”

Excerpts:
“The aircraft gained between 20 to 30 feet of altitude and experienced a desire to roll…. This entire moment took both of us by complete surprise. With the speed so low and the attitude of the plane upset I forced the throttles to max power without regard for the go-around button…. I would suggest that more unanticipated balked landings occur during training. The "startle factor" of a wake turbulence upset near the runway was intense. I believe this "startle factor" was what led me to abandon my duty to push the go-around button, thus upsetting the normal flow for a go-around.”

Analysis:
Onset: Wake turbulence during landing, intense movement
Effect: Immediate throttle up, no TOGA button (flow disruption)
Reflection: Better training in late go-arounds, preparedness

ACN: 1157235

Synopsis:
“When a B767-300 was struck by as many as three large birds shortly after take-off the right engine had to be shut down due to high EGT and vibration. As they returned for landing they learned the flap extend/retract mechanism had been damaged forcing a less than normal landing flap setting and high right side brake temperatures after roll out. CFR suggested they evacuate when the fuse plugs starting melting. The crew complied with their suggestion.”

Excerpts:
“we struck at least three large birds probably black vultures. One hit the fuselage near me, one went into the RH engine causing excessive EGT 1050 and high vibration, [and] I am not sure what happened to the last one. I was very startled and didn't immediately start doing the memory items. The Captain started to do them so I decided to transfer control of the airplane to him and assumed the role of pilot not flying…. We did the Evacuation Checklist and evacuated the airplane.”

Analysis:
Onset: Impact with large birds, engine vibrations
Effect: No immediate action (no memory items)
Reflection: None

ACN: 1142116

Synopsis:
“B757 First Officer reported he attempted a go-around because of an unstable approach due to either wake vortex or wind shear, and the go-around was sloppily executed.”

Excerpts:

“Around two miles prior to the FAF, the aircraft began handling as if it was encountering wake turbulence and the speed began to increase beyond 190 KTS…. Having been on the B-737 for six years prior, I instinctively reached ahead of the throttle quadrant for the TOGA buttons. The fact that they weren't there surprised me and I spent the next few seconds reaching around with my fingers to find them. After a few seconds, the Captain repeated "Go-around, let's go!"…. So, I manually overrode the throttles and proceeded to execute the go-around without flight director guidance….. he realized I still had not engaged the TOGA modes and reached over and hit the TOGA button for me…. I hand flew the new heading but felt like I was behind the airplane and spending too much focus on hand flying. I engaged the autopilot and then checked to see what speed we were at for the next flap retraction. I only focused on the airspeed indicator for a few seconds but when I came back to the rest of my instruments, I saw we were around 1,900 FT and the autopilot had stopped climbing and was beginning a descent. That startled me so much, I disconnected the autopilot and autothrottles, aggressively pulled back on the yoke and applied full power. We quickly recovered, cleaned up the airplane, levelled off, reengaged the autopilot and autothrottles, and realized the gear had never been retracted…. To say I was startled to see us descending would be an understatement and I know I overreacted with the controls. Though correct, it was not a smooth response. Add rumbling gear and multiple full-power applications and I'm sure it alarmed some passengers in the back…. Other regrets: I wish I had asked the Captain for help finding the TOGA switches. I also wish I had disconnected the autothrottles immediately to keep them from fighting me on the go-around.”

Analysis:
Onset: Wake turbulence, speed increase
Effect: TOGA buttons could not be found, manual override after verbal command. Rough steering, behind aircraft
Reflection: Early autothrottle disconnect (simplify), too much overreaction

---

ACN: 1045570

Synopsis:

“Air Carrier Captain reports descending early during arrival into SKBO in VMC due to misunderstanding what the First Officer read back to the Controller. An EGPWS warning is triggered. The First Officer was on his second trip after IOE and was having difficulty with foreign accents.”

Excerpts:

“We were over a valley but had a ridge line close off the left wing. Maintained course and terrain clearance visually and continued approach. The terrain warning did startle me and I disconnected the autopilot and auto throttles but did not pull up, as I realized I had safe visual terrain clearance along my flight path.”

Analysis:

Onset: Terrain warning despite visual contact with terrain

Effect: AP disconnect, but no manual action (realized margins were OK)

Reflection: None

ACN: 1017650

Synopsis:

“A B737 First Officer responded incorrectly to a TCAS RA after he increased the aircraft's rate of descent when the resolution was to decrease the descent rate.”

Excerpts:

Captain:

“Our traffic alert quickly changed to a TCAS Resolution Advisory (RA), telling us to "Monitor Vertical Speed." At that time, the Co-pilot turned off the autopilot and responded to the RA…. When I looked at the TCAS information on the Vertical Speed indicator, it seemed to me the Co-pilot was not reacting properly for the information displayed. The green arc was a band about 500 FPM… I told the Co-pilot I had the aircraft while I grabbed the control yoke and pulled up. This action stopped the aircraft's descent and started a shallow climb….. My decision was delayed because of my expectation that my experienced Co-pilot would respond correctly to the TCAS display; that made me hesitate and question my correct interpretation.”

First Officer:

“An RA was triggered approximately three seconds later, saying, "Monitor Vertical Speed."… I mistakenly read the TCAS and initially increased the descent…. He started to pull back on the yoke about the time I realized that I needed to shallow the descent…. I learned after the event that they had called after the RA had gone off. At that point, I was fully trying to comply with the RA and tuning the radio out so I didn't hear the traffic call…. Another threat was being startled. It is a fairly low work environment in the mid FL300s and an RA at that altitude gets your attention. I was startled and it took a few seconds to realize what was happening. I happened to misinterpret what I saw…. Another threat was the TCAS instrument itself. In the -300 it is a round dial. When the RA occurred, we were at -3,000 FPM, which is about the four to five o'clock position on the dial. It was all red until about the eight o'clock position. It was not immediately apparent to me that the TCAS wanted me to shallow the descent looking at the instrumentation. I think in the heat of the moment I looked down and swapped the dial from the left being zero to the right being zero and felt the urge to push through the red to the green on the left. I recommend that all TCAS instruments in the future be vertical so pilots, in the heat of the moment, would see clearer pictures of what to do. Another threat is the lack of communication of the threatening aircraft prior to the event. In most of my other RA events, I could see a situation developing. In this case, I was startled and made a wrong decision."

Analysis:
Onset: TCAS warning
Effect: Incorrectly responded (descend vs climb), blank for several seconds
Reflection: TCAS dial prone to misinterpretation, poor SA prior to TCAS

ACN: 1004144

Synopsis:
"An Air Carrier First Officer reported a near miss with what appeared to be a weather balloon at 11,000 FT between BRAND and KORRY Intersections on the LGA KORRY 3 Arrival."

Excerpts:
"The balloon appeared to be about 4 to 5 FT in diameter and equally as long, it passed about 50 FT off the nose…. I heard the Captain yell and saw him duck. It startled me quite a bit as did it him….. It was so startling that we did not get down to 10,000 FT by KORRY."

Analysis:
Onset: Sudden physical object in front of aircraft
Effect: Physical jerk reaction and scream, induced startle in other pilot, descent mismanaged
Reflection: None

ACN: 795767

Synopsis:
"MOMENTARY AFT LAV SMOKE AND MASTER WARNING LIGHT ON TKOF ROTATION RESULTS IN AN EMERGENCY DECLARATION AND OVERWEIGHT LNDG AT DEP ARPT. NO EVIDENCE OF FIRE IS FOUND."

Excerpts:
"RIGHT AT ROTATION ON TKOF, WE RECEIVED A MASTER WARNING LIGHT AND A SMOKE AFT LAVATORY MESSAGE. I WAS THE PF AND WAS JUST RAISING THE NOSE. I MOMENTARILY REDUCED PWR, STARTLED, BUT THEN PUSHED IT BACK IN AND CONTINUED THE TKOF…. THE INITIAL ERROR IN REDUCING THE PWR MOMENTARILY WAS JUST A SCREW-UP ON MY PART. I WAS REALLY SURPRISED THAT I DID THAT. AS FOR THE LATE DECISION TO RETURN TO FIELD, I TAKE RESPONSIBILITY FOR A BAD DECISION MAKING PROCESS. THAT SAID, WE DON'T HAVE MUCH GUIDANCE FOR MOMENTARY INDICATIONS AND MESSAGES."

Analysis:
Onset: Master caution (smoke) during rotation
Effect: Momentary pause of rotation, reduced power
Reflection: Surprised at own actions, currently poor guidance on momentary indications & messages

ACN: 704645

Synopsis:

“A B767-300 FLT CREW ABORTED THEIR TKOF AT EDDF BECAUSE OF EXCESSIVE VIBRATION. ACFT CHKED OUT NORMAL, CREW DETERMINED VIBRATION WAS DUE TO ROUGH RWY SURFACE.”

Excerpts:

“ON TKOF ROLL FROM RWY 7L AT EDDF AT ABOUT 70-75 KTS, FRANKFURT'S ROUGH RWY MADE ME WONDER IF SOMETHING WAS NOT RIGHT WITH LNDG GEAR OR TIRES. I BEGAN THE ABORT BELOW 80 KTS…. WE HAD DISCUSSED HOW BUMPY EDDF RWYS WERE THE DAY BEFORE AND PRIOR TO TKOF AS WELL. BUT THE ROUGH PAVEMENT STILL STARTLED ME ENOUGH TO STOP THE TKOF…. WE CONCLUDED THE SHAKE WAS DUE TO VERY ROUGH RWY.”

Analysis:

Onset: Loud runway friction noise (assumed gear/tire failure)

Effect: Aborted take-off

Reflection: Later discussion pointed to the rough runway surface conditions

Appendix B Appendix B: ASRS Incident Synopses Surprise Effect

The following 13 selected ASRS incidents are those illustrating the surprise effect, as described in Table 4.

ACN: 1223274

Synopsis:
“An A321 flight crew was surprised to receive an ECAM memo of "FOB LESS THAN 6600," much below that required, despite regular previous howgozits of fuel in excess of that planned. They soon determined that they had no access to the fuel in the ACT (Additional Centre Tank) which should have fed into the centre tank as it emptied. They diverted, whereat a fueler discovered the transfer mechanism was improperly configured, preventing fuel transfer from the ACT to a tank accessible by the engines.”

Excerpts:
“We noticed a green memo "FOB LESS THAN 6600 lbs"…. I immediately went to the fuel page and noticed that our wing tanks were down to approximately 5000 lbs. of usable fuel…. This action was taking place during the descent, while setting our altimeter, thru transition, while changing our destination, selecting a runway change as the winds were shifting, and my IPAD suddenly said I had less than 10% power remaining! I was then connecting my backup battery, and an altitude warning horn brought my attention to a discrepancy of approximately 1000 ft. I had a message to consider selecting air data on 3, which I did, but this didn't fix that problem so we compared to the STBY, and with air traffic control, thus confirming our proper altitude of 15,000. My altimeter was incorrect, thus ruled out for the remainder of the flight… The altimeter error was caused by pilot error, during a very high task loading event. I had set 28.96, instead of 29.86, thus causing a 1000 ft error, but we did eliminate the use of this altimeter, almost immediately.”

Analysis:
Onset: Unexpected low fuel state (caused by centre tank disabled via wing fuel interface)
Effect: Current tasks (altimeter setting) were obfuscated, task saturation with other menial issues
Reflection: None

ACN: 1174038

Synopsis:
“An A319 crew was surprised by the aircraft transitioning to Managed Speed and accelerating to overspeed the flaps while on approach at 3,000 FT in the HDG and SPD SEL modes, APP not activated, so the crew executed a go-around, but returned for a normal landing.”

Excerpts:
“We had already flown several flights together both on this sequence and on previous sequences in the past…. The aircraft was in selected Speed Mode and open descent thus allowing the crew to enter a visual pattern…. The aircraft was in selected speed mode, the flaps were at 1 and the First Officer was hand flying the aircraft using flight director guidance, manual stick and autothrust…. At 3,000 MSL and turning left to final approach both pilot immediately noted the aircraft engines spooling up and the aircraft beginning to accelerate toward 200 KIAS. I immediately called "watch your speed" and simultaneously looked at the speed select window. I noted the speed had reverted to "managed" mode so I immediately pulled the speed select knob and spun the speed back to below 150 KIAS. As I spun the speed select knob I noted the airspeed indicator was just at or slightly below the barber pole but the aircraft and the airspeed trend arrow bouncing erratically due to the wind gust of up to 30 knots…. Notes concerning this issue can be found in Operation Manual, though the Manual fails to explain why the approach mode will inadvertently activate. I have concluded that the A319 will automatically activate the approach mode when below 7,200 MSL (Ops Man Vol II), but if not established on a published segment of the approach, within a yet to be determined radial distance of the IAF. The issue is that since the aircraft is not on the approach segment but below 10,000 MSL, the aircraft attempts to accelerate to the default managed speed of 250 KIAS. This aspect must be explored in more depth by the appropriate Company department and subject matter experts to be verified.”

Analysis:
Onset: Inadvertent approach mode activation caused a sudden unexpected acceleration during descent
Effect: Immediate action to revert to sel. Mode continued descent
Reflection: Confusing mode switching, poorly documented, should be clarified

ACN: 1137763

Synopsis:

"The flight crew of a Boeing RNAV capable, GPS equipped, twin jet cleared for the RNAV (GPS) Runway 23 at FWA were alerted by an EGPWS warning and a low altitude alert from the Tower that they had descended excessively below the RNAV generated glide path. Simultaneous with the warning the crew was surprised to see that the autoflight system had somehow reverted from VNAV PATH to FLCH. An immediate EGPWS escape manoeuvre was executed by the pilot flying who applied insufficient pressure to the go-around button to activate go-around guidance and autothrottle response and a short duration stick shaker warning occurred, silenced quickly by manual thrust lever advancement. A subsequent approach under close observation repeated the anomaly with VNAV/FLCH."

Excerpts:

"A number of factors; weather, experience level, system malfunction, and less-than-ideal decisions lead to a situation where an EGPWS Caution was experienced as a go-around was initiated…. I am very familiar with this approach and had flown it as the pilot not flying the night before in better weather conditions in the same aircraft. The aircraft flew it flawlessly as it usually does…. For about one second my brain attempted to make sense of the outside picture, and then I called for a go-around. Almost simultaneously, we received an EGPWS caution "TOO LOW, TERRAIN" as the go-around was in progress. Later during the debrief, the First Officer said that once the go-around was commanded, he hit the G/A switch but does not think it engaged (due to lack of force from his finger). He then reacted to the GPWS by abruptly pulling back on the yoke to initiate the climb, but with the power back, we briefly entered "the foot" on the airspeed tape, and he recalls momentarily getting the stick shaker before manually adding full power….) I noticed the airspeed out of the corner of my eye… Approaching BABAC I noticed the MCP VNAV switch annunciator extinguish and the FLCH switch illuminate. What the heck?... Since we had degraded situational awareness, I was beginning the GPWS escape manoeuvre because I was not sure exactly how much vertical separation we had from the ground and I wanted to get the aircraft climbing as quickly as possible. However, I initially pitched the aircraft too quickly without having sufficient power on the aircraft and we intermittently received stick shaker warnings."

Analysis:

Onset: Suddenly clear of clouds, runway not as expected, followed by EGPWS warning
Effect: Abrupt PF action (pull back), without power = very low airspeed
Reflection: Too abrupt action in GPWS escape manoeuvre

ACN: 1104105

Synopsis:
“An A300-600 flight crew was caught by surprise when the autoflight system initiated an abrupt descent and accelerated after crossing FNCHR waypoint on the FNCHR RNAV STAR into MEM. The jet descended ~500 FT below the next minimum crossing restriction of 9,000 and accelerated to about 320 KTS before the autopilot was disconnected and a recovery completed.”

Excerpts:
“We were coming in well rested and having planned ahead. Everything seemed to be going fine…. As we crossed FNCHR the airplane suddenly dived and picked up speed to about 320 KTS. I was the pilot monitoring and called airspeed…. As we were accelerating the First Officer (pilot flying) reached down to [set] TACT [an FMS speed intervention mode] 290 KTS (we were around 310-320 KTS at this point). [At this time] I took the airplane, disconnecting the autopilot and pulling the nose up…. I have 15 years in the jet and have never seen it dive on its own like that. It can get confused if above profile but I thought we were fine. Many lessons learned…. The First Officer is very diligent, helpful, no reason to expect he would not have corrected ASAP. I was surprised when the solution he came up [with] for the diving and increasing speed was to go to the box…. I was [task] saturated and missed the altitude of 9,000. I should have taken it sooner….. by total surprise with the dive in the arrival environment, coming out of--what I thought was--nowhere….. I will be more aggressive and forward thinking in all phases in the future.”

Analysis:
Onset: Aircraft suddenly dived and accelerated
Effect: FO attempts automation intervention, captain reacts with manual control (pull up, disengage)
Reflection: Surprised by automation, puzzled by FO automation reaction

---

ACN: 1032254

Synopsis:
“Despite having programmed ATC's descent crossing restriction, the flight crew of a B737-400 failed to comply when the pilot flying disconnected the autoflight system and initiated a descent rate less than that required to comply.”

Excerpts:
“Just as the autopilot was about to begin the descent, ATC called to ask if we were going to make the crossing restriction. I replied with a yes but the First Officer was apparently startled and reacted by taking the aircraft controls and disengaging the autopilot which resulted in the aircraft not beginning its planned idle descent. I told the First Officer to use the speedbrake to get the aircraft down but he was not responding so I activated the speedbrakes. The First Officer was still not responding adequately enough. I told him to push the nose over and use increased speed to get it down but he still did not respond quick enough, seeming to do a gradual descent, resulting in us missing the crossing restriction by 1,500 FT.”

Analysis:
Onset: Untimely ATC call
Effect: PF disables automation and assumes manual control, PF frozen (unresponsive)
Reflection: None

ACN: 1030005

Synopsis:

"An MD-11 flight crew was surprised when the aircraft entered a climbing right turn and the autothrottles advanced commensurately when the approach/land mode autoflight mode was selected. They disconnected the autopilot and autothrottles but gained nearly 1,000 FT before regaining flight path control."

Excerpts:

"When the Captain selected approach/land mode the aircraft started an uncommanded high pitch/high power right turn. The Captain re-selected 3,000 FT, but we continued the high power/high pitch turn.
The Captain disconnected the autopilot and started a descent with the autothrottles still commanding high power. The First Officer selected 2,800 FT in control panel multiple times but the autothrottles did not respond. The Captain then disconnected the autothrottles."

Analysis:

Onset: Uncommanded high pitch high power turn
Effect: Reselected alt on MCP, no effect. AP disconnected, A/THR still high power, late A/THR disconnect
Reflection: None

ACN: 1028073

Synopsis:

"Although they believe they programmed the FMS properly to comply with ATC's clearance to cross HALIE at 6,000 MSL the flight crew was surprised to find they had never left 10,000 MSL when ATC queried them about making the restriction. An acknowledged contributing factor was the flight crew's failure to monitor their flight path, anticipating that the FMS would do what they thought they had programmed."

Excerpts:

"A few miles from HALIE, ATC asked us in a sarcastic tone of voice: "Do you remember that crossing restriction at HALIE!?" I looked at the FMA and we were in VNAV PATH. I looked at the Descent page on my FMC/CDU and saw HALIE at 6,000 FT. I looked at the MCP and saw 6,000. I looked at the Cruise page and saw 10,000 FT. I cannot explain why the jet did not initiate a descent earlier and on proper descent profile… I disengaged the autothrottles (I believe) and deployed the speedbrakes,… What bothers me so much is that we verified the programming, verbalized the callouts, and monitored until we were distracted with our own discussion of VNAV crossing restrictions. All the while this jet was not about to make the restriction even though, I believe, all programming and VNAV functions were done correctly.
I'll be the first to admit my mistakes, but I need to know what they were so I (and others) can learn from this! If the jet was not going to make HALIE at 6,000 why did it stay in VNAV PATH and not switch to VNAV SPD like we see a lot of times when we are not going to make the next restriction and further action is required? This is what puzzles me so much!... However, it has always bothered me that when the jet drops out of VNAV PATH and switches to VNAV SPD, there needs to be more than just an FMA change; a light, an FMC annunciation, or something. For an action as critical as not making an altitude restriction, we need a better warning system in place for the jet to tell us that it is not going to make an altitude restriction unless action is taken."

Analysis:

Onset: Jet did not descent as programmed
Effect: Disengaged A/THR and used SPDBRK
Reflection: Bothered about all precautions and dutiful flying, still gross error. Limited FMA change annunciation

---

ACN: 990058

Synopsis:
"Level, on the localizer and expecting a routine glideslope capture followed by an autopilot controlled ILS approach, the flight crew of a B737-300 was surprised and momentarily disoriented when the airspeed began to decay rapidly despite thrust increases as high as max continuous and indications the aircraft was descending on the glideslope."

Excerpts:
"Glideslope captured and speed started to decay. Pilot flying announced, "Correcting airspeed," and added power. Pilot not flying announced "airspeed" as airspeed decayed through Vref. Aircraft was still [indicating] on glideslope with airspeed decaying rapidly.... This time we briefed our plan of action to override the autopilot immediately if it attempted to climb.... The situation evolved rapidly from mind numbingly normal to something I have never seen before, appearing to be descending on glideslope at MCT yet losing airspeed [because we were actually climbing]. The threat of continued loss of airspeed forced me out of believing the instruments into saving the aircraft from a potential stall."

Analysis:
Onset: Speed decay during glideslope capture
Effect: Go around and second approach better prepared for behaviour
Reflection: Rapidly changed from "mind-numbing" standard to "something never seen before"

ACN: 964358

Synopsis:
"After planning a 10 degree flap take-off due to their aircraft's planned weight and a short runway, the B737-700 flight crew, out of habit, set five degrees instead. Shortly after the start of the take-off roll the First Officer realized the error and reset the flaps to 10 degrees and the take-off was continued."

Excerpts:
"This was the first Flaps 10 take-off that I have done in literally several years. It's always Flaps 5.... We were cleared for take-off and, just about the time I hit the TOGA button, my First Officer said, "S***, Flaps 10" and instantly slapped the lever to 10. That startled me and delayed what should have been an abort. After one or two seconds (while my brain was trying to process what just happened), the Takeoff Warning horn beeped about two or three times. I looked up saw the flaps at 10 and the speed going past 80 knots (spring loaded to only abort for fire or failure) and I elected to continue the take-off.... I do not feel the threat/problem to be complacency so much as habit. 99.9% of my take-offs have been Flaps 5 so I was totally spring loaded into that routine. Kudos to my First Officer for catching the mistake."

Analysis:
Onset: Realized incorrect flap settings during take-off run

Effect: Delayed reaction (several seconds), continued take-off
Reflection: Hard-coded habits lead to surprises in alternative configurations/situations

---

ACN: 897215

Synopsis:
“A B737-300 flight crew was surprised when the autoflight system initiated a hard right turn during an arrival sequencing.”

Excerpts:
“The aircraft started to turn sharply to the right nine miles before we reached the fix. The number two FMC blanked out just before this, but came back with no errors. We thought the plane was unsure of its position…. If we had glass in the plane, it would not have been a problem.”

Analysis:
Onset: Sudden uncommanded turn to the right during arrival
Effect: Recovered and landed
Reflection: Likely preventable with glass cockpit

ACN: 881125

Synopsis:
"A B737-700 flight crew on vectors to intercept and ILS was surprised when the autopilot commanded a right turn and pitched up while the autothrottles failed to advance to maintain programmed airspeed."

Excerpts:
"Captain then selected flaps 15 manoeuvring speed on the MCP. As he did so, the autopilot pitched the aircraft up, and started a climbing right turn. The airspeed bled off rapidly.
We had a Company Captain in the jumpseat. He was the first to call out airspeed…. Unfortunately, I am uncertain as to what the FMA annunciations were at that point; the MASI (Mach Airspeed Indicator) and ADI had my full attention at that time…. After much discussion with Captain and the Captain in the jumpseat, we were unable to ascertain exactly why the autopilot and autothrottles responded the way they did."

Analysis:
Onset: Uncommanded right turn and pitch up, losing airspeed
Effect: Jumpseat captain first to call airspeed (delay reaction)
Reflection: Later discussion explained behaviour

ACN: 845610

Synopsis:
"Given a surprise, close in visual following lengthy vectors for weather and traffic, the Flight Crew of a B737-300 completes an unstabilized approach to a safe landing."

Excerpts:
"I asked the First Officer to ask him our sequence. He replied we were number one and cleared the visual. To say that this was a surprise is an understatement. I should have denied the approach at this point. It was awkward because the turn was given and being accomplished with no reference as to whether we had the field visually. Then there was an exchange between the First Officer and the Controller to clarify his intention. The clearance was given in mixed IMC at a distance and altitude that made a visual very difficult. In short, we were out of position for a stabilized approach. Unfortunately, I accepted the approach and fell into 'how can I make this work' mode…. It was my poor judgment that allowed the situation to continue….. It's a whole lot easier to say "no" or "go-around" than it is to do a bunch of reports and try to explain your lapse in judgment."

Analysis:
Onset: Surprised by approach situation, compounded by poor visual contact
Effect: Fell into can-do mentality
Reflection: Would have been better to go-around and re-ascertain situation

---

ACN: 973820

Synopsis:
“A B737 Captain reported a near collision with a Beech 1900 at night at the DEN airport, citing difficulties in picking out the aircraft position lights from the construction lights.”

Excerpts:
“After landing at night on 17R, we exited M5. The taxi instructions were to proceed to gate… The Beech did not come into view until we were less than 25 FT behind it. I stopped 10 FT to 15 FT behind the Beech using light braking. The Beech was almost impossible to see from behind as it sat low to the ground…. Shaken up by the close call, we continued our taxi in, once the Beech started moving towards 17R. This is when I turned west on CS, instead of CN. As soon as I entered the ramp, we realized our mistake and I stopped the aircraft. . We did well by taxiing slowly and using the landing light. This alone saved us from a possible disaster. Where I went wrong was in not taking time to recover from the shock, and instead, continuing to taxi at a very complicated airport under night conditions. Would have been better to set the parking brake, and take a minute to see where we were and where we needed to go from there. When you are startled by something, you tend to lose the big picture and instead focus on what just happened. Taking a minute to settle your nerves and think can keep you out of further trouble. I will continue to taxi slowly at night, use all available lighting, and be mindful that small aircraft can be very difficult to see from behind at night.”

Analysis:
Onset: Small aircraft suddenly appeared during taxi at night
Effect: Distracted, boggled, missed (slow) taxiing at night, lost big picture
Reflection: Better to take a minute to recover instead of pushing on

Appendix C Airline Questionnaire

- Are Startle and Surprise specific subjects handled in current training (e.g. recurrent, type), or is it included as a part of training?
- If yes, in which training are they addressed (theoretical and/or practical training).
- If theoretical training addresses these issues, are the causes and consequences (e.g. physiological, behavioural, and cognitive) explained in relation to performance on the flight deck?
- Is practical training focused on dealing with specific Surprising events (such as upsets, stalls or specific failures) or with the training of skills which can be used in any Surprising event?
- How do you determine which events would be useful for practical training where Startle and Surprise are handled?
- Are Startle and Surprise related subjects addressed in current Instructor training?
- How is ‘What if’ planning and thinking encouraged (before, during or after flights)?
- Are stress management techniques (breathing, managing thoughts, biofeedback) provided and/or practiced in standard training or otherwise on an individual level?

Appendix D Simulator scenarios

Exercise	Conditions	Remarks
(00) Automatic ILS 27 approach with stby bus failure	• CAPT is PF • On intercept heading • Cloudbase at AMS at 300’ • Visibility 1200m • 45 minutes of fuel	Pre-Test Uninformed Group
(0 + 1) Explosive decompression with engine severe damage	• FO is PF • Flight at FL350 • Location in TMA of AMS • IMC • Random traffic TCAS traffic on display	Mid-Test Informed & Uninformed Group
(2) VOR approach 36C AMS with simulated surprise	• FO (pilot in RHS) is PF • Cloudbase at 700’ • Visibility 5000m • Approach programmed in FMS with the extra waypoint before point D	
(3) Sudden A/C upset of 30 degrees nose up and 60 degrees of bank.	• CAPT is PF • Reposition to FL100 • Clean aircraft and speed 250 kts. • IMC	
(4) Birdstrike after Take-off 24 followed by one stalling engine and one engine with high vibration (indication)	• CAPT is PF • Average aircraft weight: B737: ZFW 58t and fuel 8t B747: ZFW 235t and fuel 85t • Cloudbase at 700’ • Visibility 5000m • Wind 330/15 • Recoverable stall @100’	
(End) Automatic approach 36R with lightning strike and frozen MCP	• CAPT is PF • On intercept heading • CAT III weather • Showers in the vicinity • Wind 330/15 • -RA • Lightning strike @ 1900’	Post-test Informed & Uninformed Group

Appendix E Analysis between pilot categories

In addition to examining the effectiveness of the training as a complete training session, the analysis has included an assessment of the difference between the uninformed and informed groups. A further detailed analysis can be carried out on the results of the experiment to take into account the additional pilot variables identified in the experimental set-up:

- Short-haul vs Long-haul
- Experienced vs Less experienced
- Instructors vs Line-pilots

Appendix E.1 Classroom training results

Appendix E.1.1 *Short-haul vs Long-haul*

Splitting out these 19 pilots in short haul pilots and long haul pilots gives the following result concerning progress:

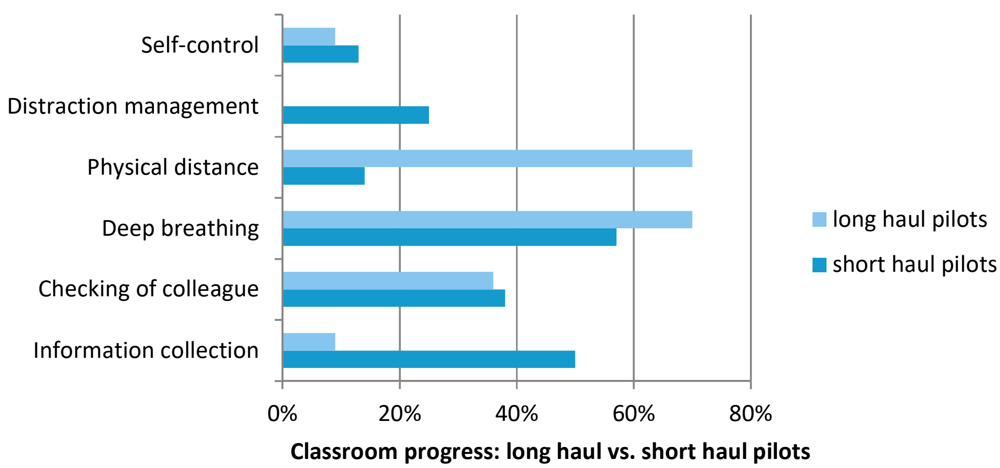


Figure 18 Classroom progress: Long-haul vs Short-haul

Table 37 Classroom progress: Short-haul pilots

short haul pilots (n = 8)	Pre-test	Mid-test	Progress	
Self-control	6 (75%)	7 (88%)	+1 (-1+2)	+ 13%
Distraction management	6 (75%)	8 (100%)	+2 (-0+2)	+ 25%
Physical distance (n =7)	0 (0%)	1 (14%)	+1 (-0+1)	+ 14%
Deep breathing (n = 7)	0 (0%)	4 (57%)	+4 (-0+4)	+ 57%
Checking of colleague	2 (25%)	5 (63%)	+3 (-2+5)	+ 38%
Information collection	4 (50%)	8 (100%)	+4 (-0+4)	+ 50%

Table 38 Classroom progress: Long-haul pilots

long haul pilots (n = 11)	**Pre-test**	**Mid-test**	**Progress**	
Self-control	11 (100%)	10 (91%)	-1 (-1+0)	- 9%
Distraction management	11 (100%)	11 (100%)	0 (-0+0)	+ 0%
Physical distance (n = 10)	0 (0%)	7 (70%)	+7 (-0+7)	+ 70%
Deep breathing (n = 10)	0 (0%)	7 (70%)	+7 (-0+7)	+ 70%
Checking of colleague	2 (18%)	6 (54%)	+4 (-1+5)	+ 36%
Information collection	4 (36%)	5 (45%)	+1 (-2+3)	+ 9%

Appendix E.1.2 *Experienced vs Inexperienced*

Splitting out the 19 pilots in experienced pilots and inexperienced pilots gives the following result

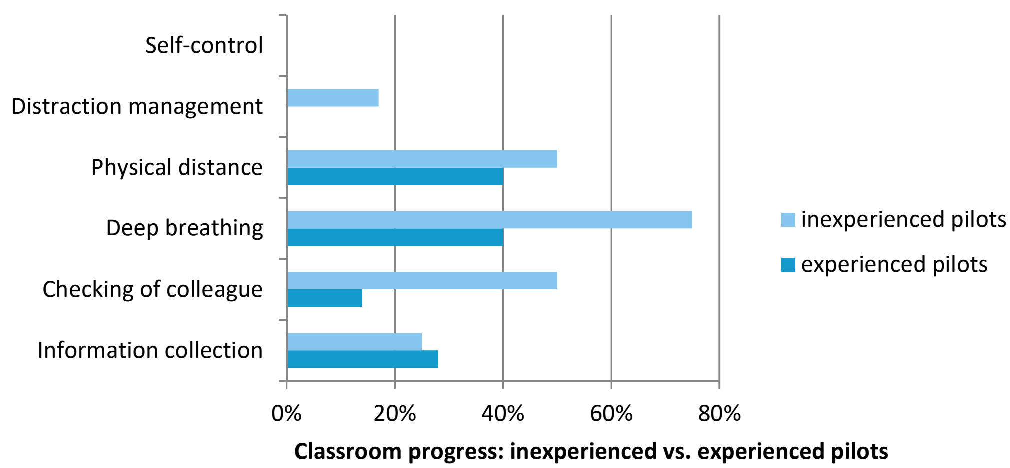


Figure 19 Classroom progress: Inexperienced vs Experienced pilots

Table 39 Classroom progress: Inexperienced pilots

inexperienced pilots (n = 12)	**Pre-test**	**Mid-test**	**Progress**	
Self-control	10 (83%)	10 (83%)	0 (-2+2)	+ 0%
Distraction management	10 (83%)	12 (100%)	+2 (-0+2)	+ 17%
Physical distance	0 (0%)	6 (50%)	+6 (-0+6)	+ 50%
Deep breathing	0 (0%)	9 (75%)	+0 (-0+9)	+ 75%
Checking of colleague	1 (8%)	7 (58%)	+6 (-1+7)	+ 50%
Information collection	5 (42%)	8 (67%)	+3 (-1+4)	+ 25%

Table 40 Classroom progress: Experienced pilots

experienced pilots (n = 7)	**Pre-test**	**Mid-test**	**Progress**	
Self-control	7 (100%)	7 (100%)	0 (-0+0)	+ 0%
Distraction management	7 (100%)	7 (100%)	0 (-0+0)	+ 0%
Physical distance (n = 5)	0 (0%)	2 (40%)	+2 (-0+2)	+ 40%
Deep breathing (n = 5)	0 (0%)	2 (40%)	+2 (-0+2)	+ 40%

Checking of colleague	3 (43%)	4 (57%)	+1 (-2+3)	+ 14%
Information collection	3 (43%)	5 (71%)	+2 (-1+3)	+ 28%

Appendix E.1.3 *Line-pilots vs Instructors*

Splitting out the 19 pilots in instructors and line-pilots gives the following result:

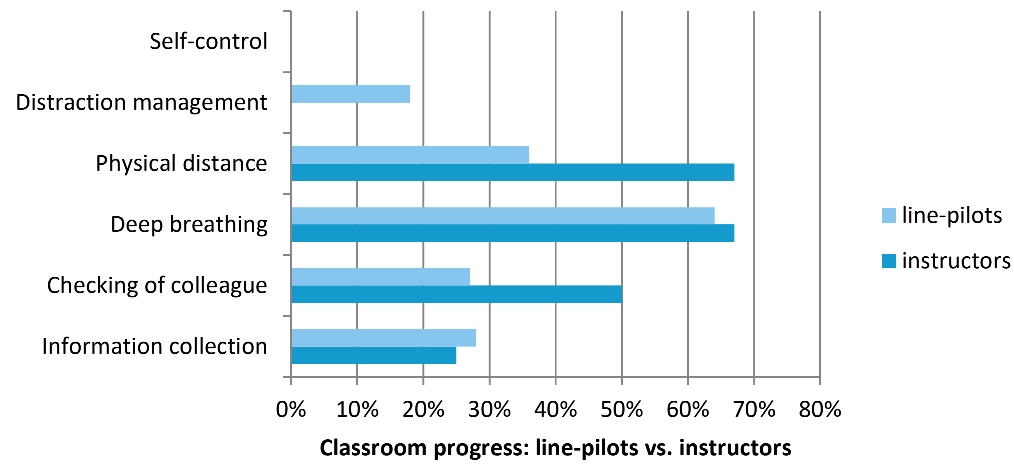


Figure 20 Classroom progress: line-pilots vs instructors

Table 41 Classroom progress: Line Pilots

line-pilots (n = 11)	**Pre-test**	**Mid-test**	**Progress**	
Self-control	9 (82%)	9 (82%)	0 (-2+2)	+ 0 %
Distraction management	9 (82%)	11 (100%)	+2 (-0+2)	+ 18 %
Physical distance	0 (0%)	4 (36%)	+4 (-0+4)	+ 36%
Deep breathing	0 (0%)	7 (64%)	+7 (-0+7)	+ 64%
Checking of colleague	2 (18%)	5 (45%)	+3 (-2+5)	+ 27 %
Information collection	4 (36%)	7 (64%)	+3 (-1+4)	+ 28 %

Table 42 Classroom progress: Instructors

instructors (n = 8)	**Pre-test**	**Mid-test**	**Progress**	
Self-control	8 (100%)	8 (100%)	0 (-0+0)	+ 0%
Distraction management	8 (100%)	8 (100%)	0 (-0+0)	+ 0%
Physical distance	0 (0%)	4 (67%)	+4 (-0+4)	+ 67%
Deep breathing (n=6)	0 (0%)	4 (67%)	+4 (-0+4)	+ 67%
Checking of colleague	2 (25%)	6 (75%)	+4 (-1+5)	+ 50%
Information collection	4 (50%)	6 (75%)	+2 (-1+3)	+ 25%

Appendix E.2 Simulator training results

Appendix E.2.1 *Short-haul vs Long-haul pilots*

Splitting out these 41 pilots in 20 short haul pilots and 21 long haul pilots gives the following result:

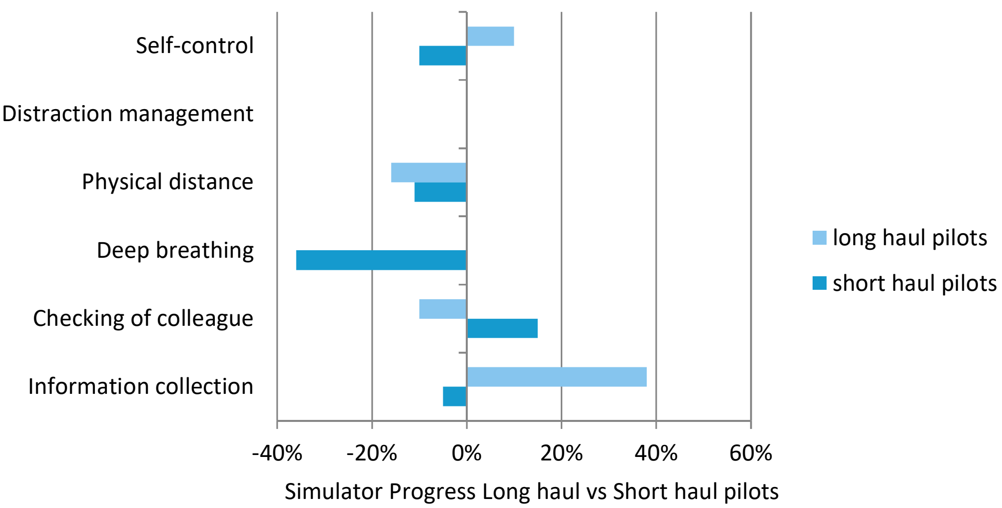


Figure 21 Simulator progress: Long-haul vs Short-haul

Apart from the indicator 'distraction management', it shows that short haul and long haul pilots show quite some difference in progress on the other indicators, especially with 'information collection'. Long haul pilots start with low scores on this indicator (52%), compared with short haul pilots (90%), but end up with about the same end result.

Another remarkable difference is on the indicator 'deep breathing' where short haul pilots score much less in the post-test than long haul pilots (29% compared to 68%). This difference is hard to explain, but could be attributed to differences in instructor guidance, to differences in aircraft (B737 gives a much louder wailer when the autopilot disconnects than the B747, which could give more distraction), or to the more dynamic operations of the B737.

With the indicator 'checking of colleague' we see a big spread with long haul pilots (-9+7), especially if we compare this with short haul pilots (-3+6). This indicates a fragile training effect on this indicator for long haul pilots (9 of them forgot to check their colleague compared to the mid-test).

Table 43 Simulator Progress: Short-haul pilots

short haul pilots (n = 20)	**Mid-test**	**Post-test**	**Progress**	
Self-control	17 (85%)	15 (75%)	-2 (-5+3)	-10 %
Distraction management	20 (100%)	20 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=18)	6 (33%)	4 (22%)	-2 (-6+4)	- 11%
Deep breathing (n=17)	11 (65%)	5 (29%)	-6 (-8+2)	- 36 %
Checking of colleague	11 (55%)	14 (70%)	+3 (-3+6)	+ 15 %
Information collection	18 (90%)	17 (85%)	-1 (-3+2)	- 5 %

Table 44 Simulator Progress: Long-haul pilots

long haul pilots (n = 21)	Mid-test	Post-test	Progress	
Self-control	19 (90%)	21 (100%)	+2 (-0+2)	+ 10 %
Distraction management	21 (100%)	21 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=19)	12 (63%)	9 (47%)	-3 (-6+3)	- 16%
Deep breathing (n=19)	13 (68%)	13 (68%)	0 (-3+3)	+ 0 %
Checking of colleague	14 (67%)	12 (57%)	-2 (-9+7)	- 10 %
Information collection	11 (52%)	19 (90%)	+8 (-1+9)	+ 38 %

Appendix E.2.2 *Experienced vs Inexperienced Pilots*

Splitting out the 41 pilots in 19 experienced pilots and 22 inexperienced pilots gives the following:

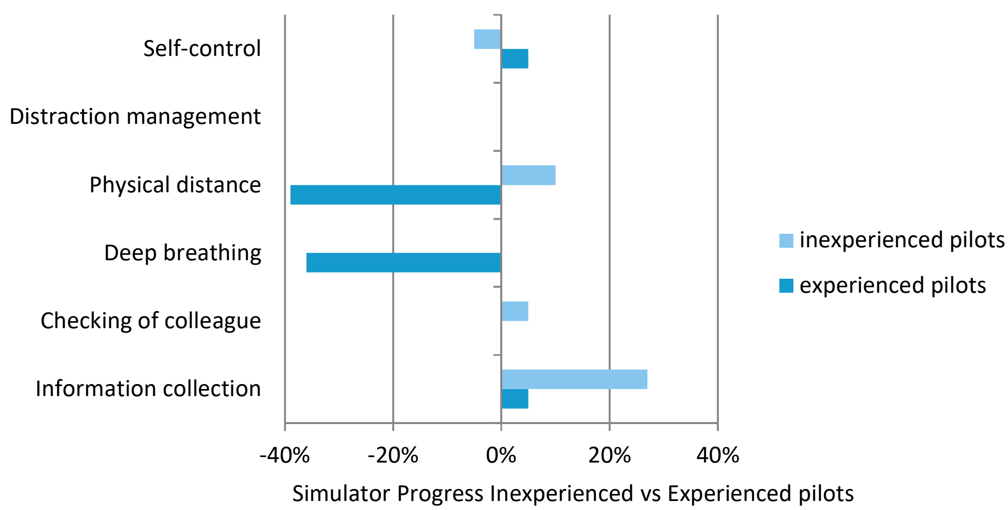


Figure 22 Simulator Progress: Inexperienced vs Experienced

Comparing these two groups shows about the same progress on ‘self-control’, ‘distraction management’ and ‘checking of colleague’, but again with ‘information collection’ we see quite some difference. Inexperienced pilots start with lower scores than the experienced pilots, but end up with higher scores in the end. This accounts for the progress of 27%. In comparison, the experienced pilots scored rather high in the mid-test, but gained only 5%.

A remarkable difference is found between the two groups on the indicators ‘physical distance’ and ‘deep breathing’, with inexperienced pilots making more progress, but also scoring higher in absolute sense. It seems that classroom training was not enough for the inexperienced pilots and that they needed simulator practice to train the skills. For experienced pilots, this turned the other way around in the sense that simulator practice was not beneficial for their performance.

Table 45 Simulator Progress: Experienced Pilots

experienced pilots (n = 19)	**Mid-test**	**Post-test**	**Progress**	
Self-control	16 (84%)	17 (89%)	+1 (-2+3)	+ 5 %
Distraction management	19 (100%)	19 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=18)	11 (61%)	4 (22%)	- 7 (-9+2)	- 39 %
Deep breathing (n=17)	12 (71%)	6 (35%)	- 6 (-8+2)	- 36 %
Checking of colleague	12 (63%)	12 (63%)	0 (-6+6)	+ 0 %
Information collection	15 (79%)	16 (84%)	+1 (-2+3)	+ 5 %

Table 46 Simulator Progress: Inexperienced Pilots

inexperienced pilots (n = 22)	**Mid-test**	**Post-test**	**Progress**	
Self-control	20 (91%)	19 (86%)	-1 (-3+2)	- 5%
Distraction management	22 (100%)	22 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=19)	7 (37%)	9 (47%)	+2 (-3+5)	+ 10%
Deep breathing (n=19)	7 (37%)	7 (37%)	0 (-3+3)	+ 0 %
Checking of colleague	13 (59%)	14 (64%)	+1 (-6+7)	+ 5 %
Information collection	14 (64%)	20 (91%)	+6 (-2+8)	+ 27 %

Appendix E.2.3 *Line-pilots vs Instructors*

Splitting out the 41 pilots in 20 instructors and 21 line-pilots gives the following:

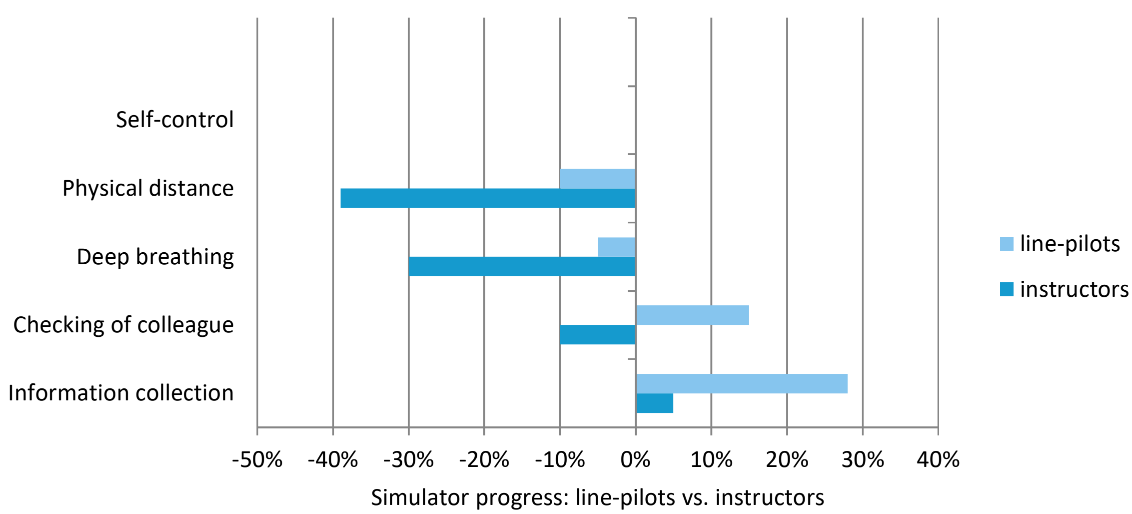


Figure 23 Simulator Progress: Instructors vs Line Pilots

With these two groups, the indicator 'checking of colleague' shows a decrease with instructors and an increase with line-pilots. Instructors scored higher in the mid-test than line-pilots (70% vs. 52%), but ended up lower in the post-test observations (60% vs. 67%). This accounts for the 25% difference in progress between groups. Again, a lot of spread is seen with both groups, with more instructors forgetting to check their colleagues in the post-test (-8) than the line-pilots do (-4).

Also on the indicators ‘physical distance’ and ‘deep breathing’ we find again big differences. We see the instructors performing much less in the post-test than in the mid-test.

The indicator ‘information collection’ shows an increase with both groups, but more with line-pilots (+28%) than with instructors (+5). The line-pilots start with lower scores in the mid-test (67% vs. 75%) but end up with much higher scores in the post-test observations (95% vs. 80%). This accounts for the 23% difference in progress between the two groups.

Table 47 Simulator Progress: Instructors

instructors (n = 20)	**Mid-test**	**Post-test**	**Progress**	
Self-control	19 (95%)	19 (95%)	0 (-1+1)	+ 0 %
Distraction management	20 (100%)	20 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=18)	11 (61%)	4 (22%)	-7 (-9+2)	- 39%
Deep breathing (n=17)	12 (71%)	7 (41%)	-5 (-7+2)	- 30 %
Checking of colleague	14 (70%)	12 (60%)	-2 (-8+6)	- 10 %
Information collection	15 (75%)	16 (80%)	+1 (-3+4)	+ 5 %

Table 48 Simulator Progress: Line-pilots

line-pilots (n = 21)	**Mid-test**	**Post-test**	**Progress**	
Self-control	17 (81%)	17 (81%)	0 (-4+4)	+ 0 %
Distraction management	21 (100%)	21 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=19)	7 (37%)	9 (47%)	+2 (-3+5)	- 10%
Deep breathing (n=19)	12 (63%)	11 (58%)	-1 (-4+3)	- 5 %
Checking of colleague	11 (52%)	14 (67%)	+3 (-4+7)	+ 15 %
Information collection	14 (67%)	20 (95%)	+6 (-1+7)	+ 28 %

Appendix E.3 Full training results

Appendix E.3.1 *Short-haul vs Long-haul Pilots*

Splitting out these 19 pilots in short haul pilots and long haul pilots gives the following result:

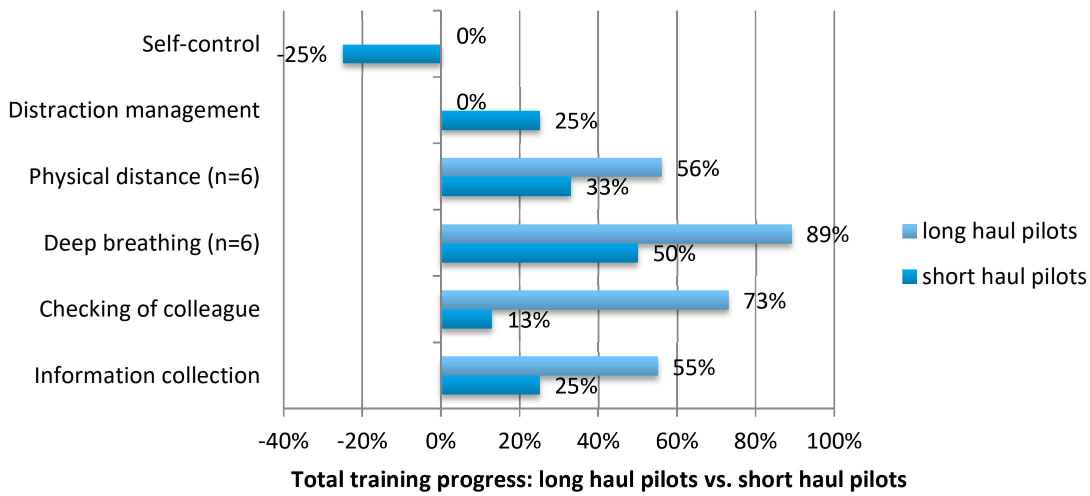


Figure 24 Training Progress: Long-haul vs Short-haul

In this group we see a big total progress on the indicator 'checking of colleague' with 11 long haul pilots (+73%). Compared to progress made in classroom training (+36%), we have to conclude that this progress could not be made without simulator training. Compared to the 21 long haul pilot group used to measure simulator progress (-10%), we should take care of generalizing this result.

Again, the remarkable aspect is the total progress made on the indicator 'information collection'. Both groups show big progress, especially the long haul pilots (+55%). Compared to progress in classroom training (+9%), we have to conclude that this progress could not be made without simulator training. This holds true when comparing with the larger group to measure simulator progress (+38%). With short haul pilots we see that the biggest progress is made after classroom training (+50%).

Table 49 Training progress: Short-haul pilots

short haul pilots (n=8)	**Pre-test**	**Post-test**	**Progress**	
Self-control	6 (75%)	4 (50%)	-2 (-2+0)	- 25%
Distraction management	6 (75%)	8 (100%)	+2 (-0+2)	+ 25%
Physical distance (n=6)	0 (0%)	2 (33%)	+2 (-0+2)	+ 33%
Deep breathing (n=6)	0 (0%)	3 (50%)	+3 (-0+3)	+ 50%
Checking of colleague	2 (25%)	3 (38%)	+1 (-2+3)	+ 13%
Information collection	4 (50%)	6 (75%)	+2 (-2+4)	+ 25%

Table 50 Training progress: Long-haul pilots

long haul pilots (n=11)	**Pre-test**	**Post-test**	**Progress**	
Self-control	11 (100%)	11 (100%)	0 (-0+0)	+ 0 %
Distraction management	11 (100%)	11 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=9)	0 (0%)	5 (56%)	+5 (-0+5)	+ 56%
Deep breathing (n=9)	0 (0%)	8 (89%)	+8 (-0+8)	+ 89%
Checking of colleague	2 (18%)	10 (91%)	+8 (-1+9)	+ 73 %
Information collection	4 (36%)	10 (91%)	+6 (-0+6)	+ 55 %

Appendix E.3.2 *Experienced vs Inexperienced Pilots*

Splitting out the 19 pilots in 7 experienced pilots and 12 inexperienced pilots gives the following result:

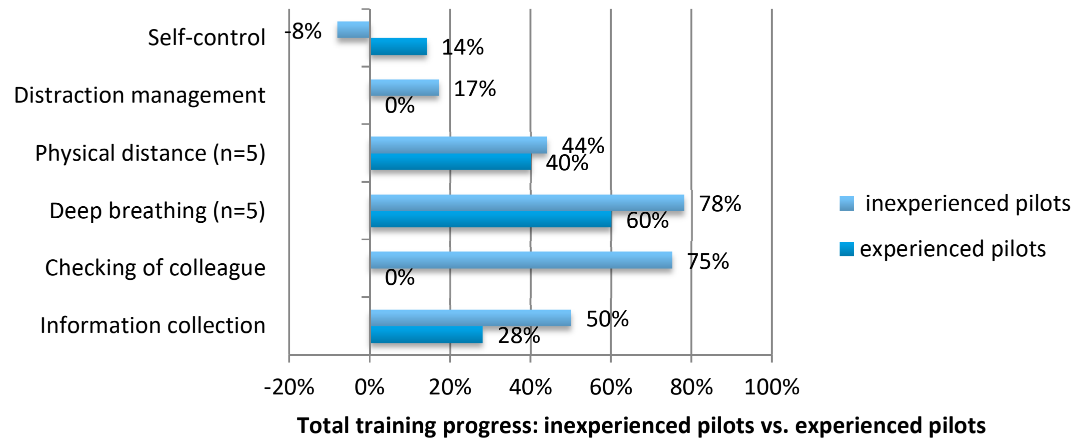


Figure 25 Training Progress: Inexperienced vs Experienced

With these groups we see much total progress with the inexperienced pilots on the indicator ‘checking with colleague’: from an 8% score in the pre-test to 83% in the post-test scenario. This is especially remarkable compared to the experienced pilots, who showed no progress at all (stayed at 43%). So even if the experienced group has higher base rates, they are not susceptible to training on this aspect (i.e. in the single training we used in the experiment). And the inexperienced group is extremely susceptible. Again, we should be careful to generalize these results, as the simulator progress scores show a progress of only 5% for 22 inexperienced pilots (and 0% for experienced pilots).

For the indicator ‘information collection’ we see both groups made progress, again more with the inexperienced group than with the experienced group. Starting with the same base rate (43%), the inexperienced group scored 92% (compared to 71% of the experienced pilots). This effect can be confirmed when comparing with the higher simulator progress scores for inexperienced pilots (+27%) in contrast with experienced pilots (+5%).

Table 51 Training progress: Experienced pilots

experienced (n=7)	**Pre-test**	**Post-test**	**Progress**	
Self-control	7 (100%)	6 (86%)	-1 (-1+0)	+ 14%
Distraction management	7 (100%)	7 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=5)	0 (0%)	2 (40%)	+2 (-0+2)	+ 40%
Deep breathing (n=5)	0 (0%)	3 (60%)	+3 (-0+3)	+ 60%
Checking of colleague	3 (43%)	3 (43%)	0 (-2+2)	+ 0 %
Information collection	3 (43%)	5 (71%)	+2 (-1+3)	+ 28%

Table 52 Training progress: Inexperienced pilots

inexperienced (n=12)	**Pre-test:**	**Post-test:**	**Progress:**	
Self-control	10 (83%)	9 (75%)	-1 (-1+0)	- 8 %
Distraction management	10 (83%)	12 (100%)	+2 (-0+2)	+ 17%
Physical distance (n=9)	0 (0%)	4 (44%)	+4 (-0+4)	+ 44%
Deep breathing (n=9)	0 (0%)	7 (78%)	+7 (-0+7)	+ 78%
Checking of colleague	1 (8%)	10 (83%)	*+9 (-1+10)*	+ 75 %
Information collection	5 (42%)	11 (92%)	*+6 (-1+7)*	+ 50 %

Appendix E.3.3 *Line-pilots vs Instructors*

Splitting out the 19 pilots in 8 instructors and 11 line-pilots gives the following result:

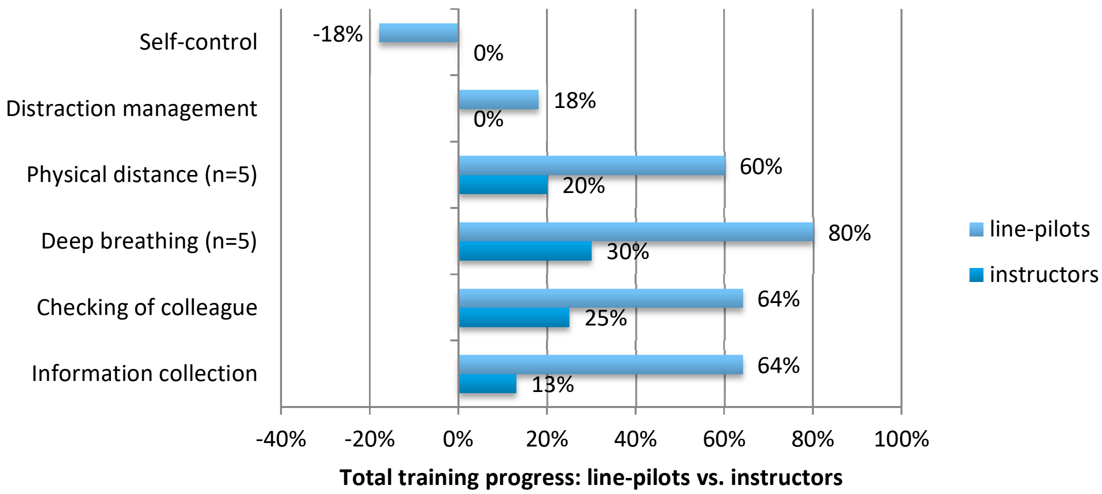


Figure 26 Training Progress: Line-pilots vs Instructors

Here we see about the same things as in the other groups: 'checking of colleague' and 'information collection' have more progress with line-pilots than with instructors, but coming from lower pre-test scores. Interesting to see though, is that all (!) line-pilots ended up using the 'information collection' technique in the post-test scenario, but that instructors scored only 63%. We can conclude that line-pilots are a lot more susceptible to this aspect of training than instructors. This can be confirmed by the simulator progress scores in the 41 pilot group (21 line-pilots +28% compared to 20 instructors +5%).

Table 53 Training progress: Instructors

instructors (n=8)	**Pre-test**	**Post-test**	**Progress**	
Self-control	8 (100%)	8 (100%)	0 (-0+0)	+ 0 %
Distraction management	8 (100%)	8 (100%)	0 (-0+0)	+ 0 %
Physical distance (n=5)	0 (0%)	1 (20%)	+1 (-0+1)	+ 20 %
Deep breathing (n=5)	0 (0%)	3 (60%)	+3 (-0+3)	+ 30 %
Checking of colleague	2 (25%)	4 (50%)	+2 (-1+3)	+ 25%
Information collection	4 (50%)	5 (63%)	+1 (-2+3)	+13 %

Table 54 Training progress: Line-pilots

line-pilots (n=11)	**Pre-test**	**Post-test**	**Progress**	
Self-control	9 (82%)	7 (64%)	-2 (-2+0)	- 18 %
Distraction management	9 (82%)	11 (100%)	+2 (-0+2)	+ 18 %
Physical distance (n=10)	0 (0%)	6 (60%)	+6 (-0+6)	+ 60%
Deep breathing (n=10)	0 (0%)	8 (80%)	+8 (-0+8)	+ 80%
Checking of colleague	2 (18%)	9 (82%)	*+7 (-2+9)*	+ 64 %
Information collection	4 (36%)	11 (100%)	*+7 (-0+7)*	+ 64 %

Appendix F Evaluation questionnaire

The questionnaire below was given to participants via an email link to a Google Form when the training session was completed. The participants were invited to complete the questionnaire immediately on their tablet/laptop and submit the responses to us, which is what most participants did. Some participants had to complete the questionnaire and submit the responses at a later moment.
The questions were all rating questions, with a Likert Scale, a space was included for individual comments.

1. <u>Which grade do you give for the whole training? (1 waste of time…. 10 great)</u>
2. <u>How much better do feel prepared for startle & surprise after this training? (1 not.. 4 a lot)</u>
3. Which grade do you give for the classroom training (1 bad… 10 great)
   a. Why?
4. Do you think the reaction test was useful? (1 no… 4 yes)
   a. Why?
5. Do you think the powerpoint presentation was useful? (1 no… 4 yes)
   a. Why?
6. Do you think the URP training was useful? (1 no…. 4 yes)
   a. Why?
7. Do you have a good idea of what the URP technique is about? (1 not at all… 4 very good)
8. <u>Do you think you will use this in the future (sim/enroute)? (1 no.. 4 yes)</u>
   a. Why?
9. Do you think the visualization of the URP technique was useful? (1 no… 4 yes)
   a. Why?
10. Do you think you would be fully trained without classroom training? (1 no… 4 yes)
    a. Why?
11. Which grade do you give for the simulator training? (1 bad… 10 great)
    a. Why?
12. How much did the simulator training connect with the classroom training? (1 not good… 4 very good)
    a. Why?
13. Do you think the chosen approach in the simulator was effective (a few short events with feedback)? (1 no… 4 yes)
    a. Why?
14. Which of the following events contributed to learning the URP technique?
    - Engine seizure plus explosive decompression
    - Confused colleague
    - Aircraft upset
    - Stalling engine plus vibration on the other engine
    - Lightning strike plus MCP failure
15. If you found one or more events not useful, could you elaborate why?
16. Do you think you would be fully trained without simulator training (so only classroom training)? (1 no… 4 yes)
    a. Why?
17. Do you have any further comments?

Appendix G Follow up questionnaire

The questionnaire below was emailed as a link to a Google form to the participants several months after completing the Startle & Surprise training exercise. The aim of the questionnaire was to gather information on their perception of the training and the application of the technique in the operation.

1. Did your impression of startle and surprise change after the training, if yes: how?
2. Did your opinion about the training change compared to the evaluation questionnaire?
3. Did you talk about the training and research with colleagues, if yes: which aspects did you talk about?
4. Did you experience anything after the training that you call startle or surprise, if yes: what?
5. Did you use the URP technique in flight operations (sim also possible), if yes: in what circumstances?
6. If you used the URP technique, was there a calming effect in the cockpit (individually or crew)? If yes: how did you notice?
7. If you didn’t use the technique while there was reason to, can you elaborate what you did do and if this was succesful?
8. Do you have any further comments about the URP technique?

Appendix H Eye-tracking Observations

For eye-tracking analyse observations, several steps were taken to identify the eye-gaze behaviour of one of the eye-tracker equipped crew members. This was done for crew members were a combination of both successful eye-tracking data- and front facing video recordings were made. Initially, analysis of the final lightning strike scenario video recordings were done to identify unload exercises as part of the URP strategy (Table 55). The video camera system positioned at the windshield facing the crew made it possible to isolate in which part of the recordings the crew was unloading. This was based on face expressions, body movements or vocal expressions. These identified timings were subsequently used to determine the behaviour in the eye-tracking data what the resulting eye behaviour was.

Table 55: Unload observation in the lightning strike scenario

Crew	Unload observations
313-314	No front facing video recording to identify unload signs
323-324	Captain (PF) announces unload 72 seconds after lightning strike, and gazes relaxed towards to PFD according the eye-tracking data. FO vocally announces unload (not eye-tracked) 15 seconds after lightning strike.
333-334	Crew did not show any signs of unloading
381-382	Crew did not show any signs of unloading (audio inaudible making it difficult to identify unload signs)
391-392	Captain PF (eye-tracked) tells FO he could unload 123 seconds after lightning strike. The FO (not eye-tracked) does some exercises with his shoulder while take a few deep breaths.
413-414	No front facing video recording to identify unload signs
471-472	Captain PF (Eye-tracked) did not show signs of unloading. First officer (Not eye-tracked) did unload 25 seconds after the lightning strike. He showed signs of slow breathings for seven seconds after the captain agreed to take over his tasks temporarily.
481-482	COCO (PF, not eye-tracked) and FO (eye-tracked) both directly start breathing deeply for about six seconds two seconds after the lightning strike. The COCO meanwhile continues with his PF task. The FO meanwhile scans the PFD, NAV, EICAS, MCP in a high rate for about 6 seconds were after he closes his eyes for a second.
491-492	No front facing video recording to identify unload signs

The eye-tracking behaviour of the crew was assessed using their instrument dwell time, dwell percentage over a predefined amount of time, and dwell frequency. The dwell time is defined as the time spend looking at a specific cockpit instrument. While an eye dwell towards an instrument do not inherently mean that information from this instrument is mentally processed, sudden changes in the scanning pattern suggest a change in cognitive processing. A sudden increase in amount of dwells per minute suggests a change in cognitive processing, e.g. to search for, or keep up with new information. On the other hand a sudden decrease in dwells could indicate a potential cognitive unload of the brain.

During the experiment most captains equipped with an eye-tracker were instructed to fly as PF. The eye-tracking data shows, as anticipated that the PF is mostly focused on managing the aircraft flight path using the PFD and NAV displays. In the 20 seconds leading to the lightning strike most PF's scanned their instruments at a rate between 15 and 20 dwells per minute, and their biggest area of interest area was the PFD with dwell percentages ranging between 50 and 84%. In the 20 seconds after the lightning strike, both the PF's that showed and that did not showed unloading signs showed similar eye behaviour. All PF's either maintained or lowered their scan rate ranging from 15 to 18 dwells per minute, while increasing their total dwell time on the PFD ranging from 90 to 95% of the time. In other words, focussing less on secondary instruments. This eye behaviour corresponds with the CRM role separation were the PF focuses primary on flight path management, whereas the PM has more resources to gather and process information on the sudden lightning strike consequences.

While most eye-tracking recordings were done on the PF, some recordings were made on the PM to further investigate unload eye behaviour within a different role. Again, using both the camera facing the crew and eye-tracking videos the behaviour could be observed in more detail. The eye-tracking data from first officer 482 who performed the experiment as PM was of extra interest. While most pilots, acting as PF, were mainly focussing on the PFD before and after the lightning strike event; the eye-tracking data from PM 482 clearly demonstrated (Figure 27) a sudden scanning pattern change after the lightning strike and additionally performed the trained unload technique. The PM's behaviour was observed for twenty seconds before the lightning strike until twenty seconds after the lightning strike. In the twenty seconds before the lightning strike the PM was busy with the checklist while scanning different parts of the cockpit at a normal scanning rate of 24 dwells per minute. Right after the lightning strikes flash and bang the PM demonstrated signs of startle (face and vocal expression) and started scanning the primary and secondary instruments, at a high scanning rate of 72 dwells per minute for five seconds. Within this same period the captain announced to start unloading and five seconds after the lightning strike both the crew members started unloading themselves using the breathing technique for about 5 seconds. Subsequently he gazed to the PFD and partially closed his eyes (1 dwell per minute). This eye-tracking data suggests that the PM did not just partially imitated the trained unload technique to show compliance, but also tried to cognitively unload himself by stopping the high frequent scanning pattern. After the unload session the scanning frequency ramps up immediately back to the high scanning rate of 72 dwells per minute for the next ten seconds.

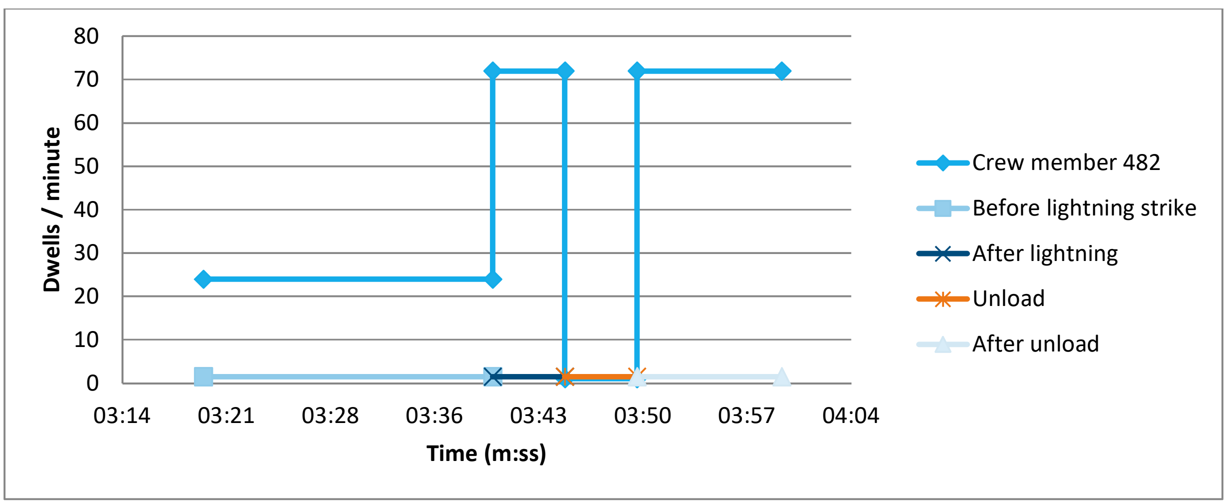


Figure 27: Lightning strike eye-tracking behaviour (instrument dwells/ minute) of crew member 482 acting as pilot monitoring

Crew 481-482, in contrast to the lightning strike scenario, did not demonstrate any signs of unloading in both his eye behaviour, face and body expressions during the bus failure baseline scenario. This suggest that the PM's unload in the lightning strike scenario was not part of his pre-experiment routine but successfully learned and implemented from his URP training.

In the baseline scenario the COCO (481) acted as pilot flying until the bus failure and consequently gave the controls to the FO (482). The eye-tracking data also suggests there was no sign of cognitive unloading. In the twenty seconds before the bus failure the FO scanned the instruments at a rate of 42 dwells per minute. His instrument scanning rate increased to 54 dwells per minute during the twenty seconds after the failure and during his assignment as PF. During this period no significant ‘breaks’ were observed in his scanning pattern which could indicate a possible unload.

Although sample sizes are too low to make any conclusions on the eye-tracking data several findings were found. First, the eye-tracking data shows that the after a failure the scan pattern between the PF and PM role is very distinct, but this was expected. In contrast to PM, it was for the PF more difficult to identify any significant unload related changes in their scan pattern as the PF’s area of interest is mostly on the PFD before and after any failure. Second, eye-tracking data from crew member 482 suggests that it was possible to both train and perform the unload technique within the scope of the experiment. Crew member 482 did not only show body language, like deep breaths, that suggests unloading, but his eye behaviour also suggested a corresponding cognitive unloading. Third, in the one of the unloading crews it was observed that the pilot closed his eyes temporarily. Although this might help to unload, in the potentially forthcoming unload training session there can be considered to inform the pilots to close their eyes or not when unloading. As alternative pilots could gaze towards the PFD with their peripheral vision. Whereas closing the eye complete could block the pilots from missing clues like potential upsets or other events as result from the initial failure, keeping the PFD in the peripheral vision could prevent this.

Intentionally left blank

EASA
European Aviation Safety Agency

European Aviation Safety Agency
Safety Intelligence & Performance Department

Postal address
Postfach 10 12 53
50452 Cologne
Germany

Visiting address
Konrad-Adenauer-Ufer 3
50668 Cologne
Germany

Tel. +49 221 89990 - 000
Fax +49 221 89990 - 999
Mail info@easa.europa.eu
Web www.easa.europa.eu