2018

H-Workload 2018: 2nd International Symposium on Human Mental Workload: Models and Applications

Luca Longo
*Technological University Dublin*, luca.longo@dit.ie

Chiara Leva
*Technological University Dublin*, chiara.leva@dit.ie

Follow this and additional works at: https://arrow.dit.ie/hwork18

Part of the Computer Sciences Commons

Recommended Citation


This Conference Paper is brought to you for free and open access by the H-Workload:Models and Applications at ARROW@TU Dublin. It has been accepted for inclusion in H-Workload 2018: Models and Applications (Works in Progress) by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie, brian.widdis@dit.ie.
H-WORKLOAD 2018

the Netherlands Aerospace Centre,
Amsterdam, The Netherlands.

This booklet contains works in progress presented at the Second International Symposium on Human Mental Workload: models and applications
Preface

This book contains the work-in-progress research contributions presented at the Second International Symposium on Human Mental Workload, models and applications. It presents recent developments in the context of theoretical models of mental workload and practical applications. Additionally, it aims to stimulate and encourage discussion on mental workload, its measures, dimensions, models, applications and consequences. It is a topic that demands a multidisciplinary approach, spanning across Human Factors, Computer Science, Psychology, Neuroscience, Statistics and Cognitive Sciences.

From the content of this volume, it is clear that mental workload is a very complex construct, thought to be multidimensional and multifaceted, with no clear and accepted definition. This is confirmed by the different modelling approaches that consider different factors and employ distinct strategies for their aggregation. Despite this uncertainty in modelling mental workload, it is clear that this construct is key for predicting human performance. The nature and efforts required by the modern workplace is shifting towards more complex cognitive demands and predicting the influence of new interactive technologies on human work is a critical capability for the practitioners of the future. The capacity to assess human mental workload is a key element in designing and implementing processes capable of monitoring interactions between automated systems and the humans destined to use them. Similarly, mental workload assessment is key for designing instructions and learning tools aligned to the limitations of the human mind. Unfortunately, mental workload measurement is not trivial. Some of the articles presented during the symposium applied psychological subjective self-reporting measures, others made use of physiological or primary task measures and some a combination of these. We believe the adoption of a multidimensional approach is fundamental to further understand the complex construct of mental workload and to fully grasp its nature. A number of research articles presented at the symposium have started focusing on the development of novel models of mental workload employing data-driven techniques, borrowed from Machine Learning as sub-field of Artificial Intelligence, whose explanatory capacity over the topic is still to be explored. This last area is a field where traditional Human Factors approaches and novel data-driven modelling approaches can cross-paths and maybe trace a new fundamental research direction that should be further explored and promoted.

Luca Longo
M. Chiara Leva
Acknowledgment

We wish to thank all the people who helped in the organising committee for the 2nd International Symposium on Mental Workload, models and applications (H-WORKLOAD 2018). In particular the local chairs Rolf Zon, Wendy Duivestein, Tanja Bos and many more of the members of the scientific committee. We want to thank also the main sponsors of the event, the Netherlands Aerospace Centre and the Irish Ergonomics Society, without which neither the conference nor the book would have been realised. A special thank goes to the Dublin Institute of Technology as well as all the reviewers of the program committee who provided constructive feedback. A special thank goes to the researcher and practitioners who submitted their work and committed to attend the event and turn it into an opportunity to meet and share our experiences on this fascinating topic.

Organising Committee

General chairs, editors and programme committee chairs

– Dr. Luca Longo
– Dr. Chiara Leva

Local chairs

– Rolz Zon
– Tanja Bos
<table>
<thead>
<tr>
<th>Name</th>
<th>Surname</th>
<th>Affiliation</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aghajani</td>
<td>Haleh</td>
<td>Baylor Scott &amp; White Research Institute</td>
<td>United States</td>
</tr>
<tr>
<td>Balfe</td>
<td>Nora</td>
<td>Trinity College Dublin</td>
<td>Ireland</td>
</tr>
<tr>
<td>Barge</td>
<td>Roland</td>
<td>Rolls-Royce</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Bellini</td>
<td>Emanuele</td>
<td>LOGOS Ricerca &amp; Innovazione, Florence</td>
<td>Italy</td>
</tr>
<tr>
<td>Bos</td>
<td>Danja</td>
<td>Netherlands Aerospace Centre</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Brookhuis</td>
<td>Karel</td>
<td>Rijks Universiteit Groningen</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Byrne</td>
<td>Aidan</td>
<td>ABMU Health Board</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Cahill</td>
<td>Joan</td>
<td>Trinity College Dublin</td>
<td>Ireland</td>
</tr>
<tr>
<td>Cain</td>
<td>Brad</td>
<td>CS Human Factors</td>
<td>Canada</td>
</tr>
<tr>
<td>Castor</td>
<td>Martin</td>
<td>GEISTT</td>
<td>Sweden</td>
</tr>
<tr>
<td>Cerrato</td>
<td>Loredana</td>
<td>EITDigital</td>
<td>Sweden</td>
</tr>
<tr>
<td>Charles</td>
<td>Rebecca</td>
<td>Cranfield University</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Costa</td>
<td>Antonio</td>
<td>Aveiro University</td>
<td>Portugal</td>
</tr>
<tr>
<td>Edwards</td>
<td>Tamsyn</td>
<td>NASA Ames/ San Jose State University</td>
<td>United States</td>
</tr>
<tr>
<td>Genga</td>
<td>Glauco</td>
<td>Institute of Aerospace Medicine</td>
<td>Italy</td>
</tr>
<tr>
<td>Galicic</td>
<td>Matjaz</td>
<td>Arcadis</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Golightly</td>
<td>David</td>
<td>HFRG, University of Nottingham</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Greiner</td>
<td>Birgit</td>
<td>University College Cork</td>
<td>Ireland</td>
</tr>
<tr>
<td>Hines</td>
<td>Andrew</td>
<td>University College Dublin</td>
<td>Ireland</td>
</tr>
<tr>
<td>Kane</td>
<td>Bridget</td>
<td>Karlstad University</td>
<td>Sweden</td>
</tr>
<tr>
<td>Keunyn</td>
<td>Fiona</td>
<td>Metro Trains Melbourne</td>
<td>Australia</td>
</tr>
<tr>
<td>Konstandinidou</td>
<td>Myrto</td>
<td>NCSR Demokritos</td>
<td>Greece</td>
</tr>
<tr>
<td>Maij</td>
<td>Anneloes</td>
<td>Netherlands Aerospace Centre</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Massaiu</td>
<td>Salvatore</td>
<td>IFE</td>
<td>Norway</td>
</tr>
<tr>
<td>Mijovic</td>
<td>Pavle</td>
<td>mBrainTrain</td>
<td>Serbia</td>
</tr>
<tr>
<td>Mills</td>
<td>Ann</td>
<td>Railway Safety Standards Board</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>O'Sullivan</td>
<td>Leonard</td>
<td>University Limerick</td>
<td>Ireland</td>
</tr>
<tr>
<td>Orru</td>
<td>Giuliano</td>
<td>Dublin Institute of Technology</td>
<td>Ireland</td>
</tr>
<tr>
<td>Pediconi</td>
<td>Maria</td>
<td>University of Urbino</td>
<td>Italy</td>
</tr>
<tr>
<td>Pickup</td>
<td>Laura</td>
<td>University of Exeter</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Raduntz</td>
<td>Thea</td>
<td>Federal Institute for Occupational Safety and Health</td>
<td>Germany</td>
</tr>
<tr>
<td>Ratkai</td>
<td>Melinda</td>
<td>University of Malaga</td>
<td>Spain</td>
</tr>
<tr>
<td>Raufi</td>
<td>Bujar</td>
<td>South East European University</td>
<td>Macedonia</td>
</tr>
<tr>
<td>Rizzo</td>
<td>Lucas</td>
<td>Dublin Institute of Technology</td>
<td>Ireland</td>
</tr>
<tr>
<td>Rocha</td>
<td>Mariana</td>
<td>Dublin Institute of Technology</td>
<td>Ireland</td>
</tr>
<tr>
<td>Roelen</td>
<td>Alfred</td>
<td>Netherlands Aerospace Centre</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Smith</td>
<td>K Tara</td>
<td>Human Factors Engineering Solutions Ltd</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Teixeira</td>
<td>Carlos</td>
<td>Dublin City College</td>
<td>Ireland</td>
</tr>
<tr>
<td>Waldmann</td>
<td>Thomas</td>
<td>University of Limerick</td>
<td>Ireland</td>
</tr>
<tr>
<td>Wilkins</td>
<td>Maurice</td>
<td>Yokogawa</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Young</td>
<td>Mark</td>
<td>Loughborough, United Kingdom</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Zon</td>
<td>Rolf</td>
<td>Netherlands Aerospace Centre</td>
<td>Netherlands</td>
</tr>
<tr>
<td>De Waard</td>
<td>Dick</td>
<td>University of Groningen</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Moustafa</td>
<td>Karim</td>
<td>Dublin Institute of Technology</td>
<td>Ireland</td>
</tr>
<tr>
<td>Van Dijk</td>
<td>Henk</td>
<td>Netherlands Aerospace Centre</td>
<td>Netherlands</td>
</tr>
</tbody>
</table>
# Table of Contents

Joan Cahill, Paul Cullen and Keith Gaynor - Estimating the Impact of Work Related Stress on Pilot Wellbeing and Flight Safety 7

Marielle Plat Robain - Human Resource Recovery: Does Hypnosis Help Manage Fatigue When Workload is Increased? 33

Maria Gabriella Pediconi and Glauco Maria Genga - On the Earth as Well as in the Sky Facing the Risks Between Workload and Resilience: a Psychoanalytic Point of View 64

Avner Y. Bendheim - Automation Interaction from a Design-Thinking Perspective: The Case of Cockpit Automation in Commercial Aircrafts in Israel 93

Héla Mourali (Corresponding author) and Lassaad Lakhal - Mental workload measurement, The case of stock market traders 114

Rolf Zon - Eye tracking overview 139

Alison Kay, Katie O’Byrne, Siobhan Corrigan, Tiziana C. Callari, Dubhfeasa Slattery, David Smyth, Michael Turner, Deirdre Bennett, Jamie Logan and Sam Cromie - Development of an Observational Protocol for Reducing and Mitigating Workload and the Risk of Retained Foreign Objects 154

Tanja Bos, Antoine de Reus, Marieke Suijkerbuijk, Jaap Groeneweg, Wilfred Rouwhorst - Interactive Head Up Display in the cockpit to reduce Crew Workload 177

Henk van Dijk and Alfred Roelen - Pilot Fatigue Measurement 187

Adrian Kelly and Maria Chiara Leva - Transmission Control Centres and Human Interface Design: The need to support situational awareness in the face of increasing complexity 202
Pavle Mijovic, Ivan Gligorijevic and Bogdan Mijovic - Electroencephalography Studies of Mental Workload A Neuroergonomics Approach 218

Audrey Reinert - Detecting Human Machine Interaction Fingerprints in Continuous Event Data 227

Catherine Burns - The new workload of new technologies 229

Lisa Jeanson, Christian Bastien, Alexandre Morais and Javier Barcenilla - Tell me how you kit, I'll tell you how you think 232
Estimating the Impact of Work Related Stress on Pilot Wellbeing and Flight Safety

Joan Cahill1[0000-0001-6944-744X], Paul Cullen1 & Keith Gaynor2

1 School of Psychology, Trinity College Dublin, Dublin 2, Ireland
2 School of Psychology, University College Dublin, Co Dublin, Ireland
cahilljo@tcd.ie

Abstract. This paper presents a preliminary account of the relationship between pilot wellbeing, pilot performance and flight safety. Drawing on prior research relating to the biopsychological model of pilot lived experience, three workshops involving the participation of thirty-three commercial pilots were undertaken. Pilot wellness is a significant performance shaping factor in terms of ensuring optimum performance. Overall, pilots are managing wellbeing issues. Pilots try to normalize/adapt to the job. However, there is much variation in relation to coping ability. This variation needs to be considered in relation to modelling performance and safety impact. Six scenarios were identified. Of these, participants suggested that the primary focus should be on the prevention of Scenario 3 (i.e. pilot not coping on the day - impacting on flight safety) and Scenario 5 (i.e. pilot suffering which ends in harm to the person). Overall, pilots need to be trained in relation to (1) coping strategies and (2) risk identifying behavior.

Keywords: Wellness, Work Related Stress, Pilot Performance, Flight Safety

1 Introduction

The local actions of the flight crew are affected by a variety of immediate influences or performance shaping factors (PSF) [1]. This includes external PSFs such as environmental conditions and procedures, and internal PSF’s such as emotional state, physical condition and stress levels. PSFs are considered to substantially increase the likelihood of human error [2]. Work Related Stress (WRS) is defined as the response people may have when presented with work demands and pressures that are not matched to their knowledge and abilities, and which challenge their ability to cope [3]. Things outside the workplace, like family problems, or debt can be responsible for stress (personal stressors). A person experiencing stressful life events may find
that he/she is less able to cope with the demands of work, even though work is not the cause and/or may not have been a problem before. Pilots can be considered as both “shift-workers” and “remote-workers”. Numerous studies indicate that these types of work can be detrimental to one’s wellbeing [4], [5], [6]. Duties consisting of long work hours have also been examined [7-12], and such duties are shown to increase the risk of:

- Anxiety, depression, increased neuroticism and impaired cognitive function
- Reduction in quality and quantity of sleep
- Widespread complaints of fatigue
- Increased risk of adverse cardiovascular effects
- Increased risk of type 2 diabetes
- Possible increase in gastrointestinal effects
- Marital strain, family dysfunction and social marginalization

Psychological problems amongst aircrew present a threat to flight safety, given the ensuing impairments to task performance [13]. Stress (arising from stressors both inside and outside work) impacts on the socio-cognitive dimensions of performance including decision making, teamwork and communication [14]. Recent research has attempted to understand the impact of WRS on pilot wellbeing (including mental health). This research suggests that aspects of the job are impacting on pilot’s physical, social, and emotional/psychological health [15], [16], [17]. It is estimated, that 80% of accidents have human error as a causal factor [18]. Since 1982, approximately 1% of fatal accidents (>1,400) have resulted from the deliberate actions of the pilot [19]. Following the Germanwings 9525 accident (2015) [20], the issue of pilot suicide and detecting/managing mental health issues amongst pilots has been gaining increased attention. Surprisingly, there has been less of a focus on (1) understanding/measuring routine suffering amongst pilots, (2) understanding/measuring the relationship between work related stress, pilot wellbeing and safety, (3) understanding how pilots adapt to WRS and identifying pilot coping/self-management techniques and (4) validating existing safety performance indicators in relation to WRS and wellbeing/MH. Potentially, pilot wellbeing can be considered a significant performance shaping factor (PSF) in ensuring safe performance. If (a) like the general population, pilots are at risk of developing wellbeing/MH issues, and (b) wellbeing/MH is a PSF, then wellbeing/MH potentially is a contributory factor to some of the 80% of the 99% of fatal accidents deliberately caused by pilots. To this end, this paper presents the findings of recent action research with commercial pilots, which attempts to unpack these issues, to estimate the impact of WRS and associated wellbeing issues, on both pilot performance and flight safety. As part of this, it introduces six wellbeing and safety scenarios which form the basis of a preliminary account of the relationship
between work related stress, pilot wellbeing, pilot performance and flight safety. First, the background to this research is introduced. An overview of the research methodology is then provided. Following this, the main workshop findings are presented and discussed. Areas for further research and next steps are then outlined. Lastly, some preliminary conclusions are drawn.

2 Background

2.1 Wellbeing

The term "wellbeing" includes various aspects of the way people feel about their lives, including their jobs, and their relationships with the people around them [21]. According to biopsychosocial models of health and wellbeing [22], [23], [24], the cause, manifestation and outcome of wellness and disease are determined by a dynamic interaction between biological, psychological and social factors. None of these factors in isolation are sufficient to lead definitively to wellness or illness. Instead, it is the interrelationships between all three pillars that leads to a given outcome. It is estimated that 33% of the population experience mental health issues at some stage in their life [25] while 16% will experience mental health issues at any given time [26].

2.2 Pilot Job, Types of Operations & Lived Experience

In many ways, pilots are in a unique occupational group. The job of a commercial pilot is challenging, both physically, mentally and socially. Working hours in a typical week can vary greatly from week to week and are regulated in accordance with several parameters, such as Duty Time and Block Time, as defined by the European Aviation Safety Agency (EASA). Duty Time is the time the pilot spends at the disposal of the airline, whereas Block Time is the time spent when the aircraft is moving. Duty Time restrictions tend to apply more on a daily basis. European pilots are generally restricted to a 13-hour duty limit, which can be reduced depending on how many flights are flown and by how much the duty overlaps the Window of Circadian Low (WOCL). This limit can also be increased if an additional pilot is carried, thus allowing in-flight rest. Block Time limits generally are longer term limits, with pilots limited to 100 hours in a rolling 28-day period, 900 hours in a rolling 365 days and 1,000 hours in a calendar year. The overall intensity of the operation can also vary greatly, typically with busier Summers and quieter periods during the Winter. The particular working routines of pilots vary according to the type of operations they fly. Three types of operations can be distinguished - namely short, medium and long range. These
different types of options pose diverse wellbeing challenges. For example, pilots working long range are more likely to spend periods of time away from home – impacting on the home/work interface and their ability to maintain social routines. Pilots operating short range tend to experience intense working days – potentially, involving three to four takeoffs and landings. This type of operation, despite accruing relatively low Block Hours, involves high workload. In medium and long range, the Block Hours may be higher, with longer periods of rest/down time while on duty (i.e. cruise periods), and typically longer duty periods. Pilots experience much disruption to their sleeping and eating patterns. Specific patterns may also vary according to the operations flown. For example, if flying short range, duty might involve a week of ‘earlies’ (i.e. starting at 5am), followed by a week of ‘lates’ (finishing at 2am) – resulting in disruption of the circadian rhythm associated with sleeping and eating patterns. As reported by Wright et al [27], mental fatigue and sleepiness may rise to unacceptably high levels during civil air operations given relatively long duty periods that may coincide with disruption of the circadian rhythm due to time zone shifts.

2.3 Mental Workload, Performance & Wellbeing

Mental workload refers to the amount of information processing capacity necessary for a person to complete a task during a specified time-period [28]. The relationship between mental workload and fatigue is well documented. Research indicates that high levels of mental workload can lead to mental fatigue [29]. Equally, mental fatigue can impact on workload/task management, along with other cognitive dimensions of the pilot performance – such as attention and decision making [14]. High workload is also associated with stress [30] – which in turn has an impact on wellbeing. However, the specific relationship between wellbeing and mental workload is under-researched.

2.4 Fatigue, Performance & Mental Health

The relationship between fatigue and performance is well documented. As reported by Caldwell (1997), shift work and long hours of duty exacerbate jet lag among aircrew [31]. This can lead to pilots becoming more tired and drowsy, and it may impair their focus and attention, thereby increasing the risk of errors [31]. According to Dinges and Kribbs (1991), specific impairments associated with drowsiness include slow reaction times, reduced vigilance, and deficits in information processing [32].
Moreover, the relationship between fatigue and aspects of mental health is receiving increased attention. A recent cross-sectional survey of >700 pilots investigating self-reported anxiety and depression, reported that 54.4% suffered from feelings of being depressed and/or anxious in the previous 12 months [33]. Respondents who typically spent longer hours on duty per week (>40 hours vs. <25 hours) were three times more likely to report feeling anxious or depressed [33].

Further, in a 2016 Harvard study, >1,800 pilots responded to the Patient Health Questionnaire 9 (PHQ-9). 12.6% of participants, and 13.5% of those who reported having flown an aircraft in the previous seven days, met the threshold for clinical depression, and 4.1% reported having suicidal thoughts within the past two weeks [34]. The authors recommended that “airline organizations increase support for preventative mental health treatment”, and called for further research to evaluate additional risk factors of depression such as sleep and circadian rhythm disturbances [34].

2.5 Coping Methods of Airline Pilots

It has been demonstrated that overall mental ill health has a very close association with lack of autonomy at work, fatigue, the inability to relax, and the lack of sufficient social support [21]. In an airline marriage, the spouse can function as a very helpful social support system, thus aiding the pilot in dealing effectively with psychosocial stressors [35]. Pilots suffering marital distress are less able to concentrate effectively on their piloting duties and responsibilities [36]. On the other hand, the positive social support provided by romantic/spousal relationships can be undermined by antisocial work practices. The effect of life disruptions on pilots who frequently leave home to perform their flying duties is not fully appreciated. Bennett (2006) highlights the importance of social support obtained from fellow pilots, reporting that team members’ mutual support, camaraderie and cohesion enhance their resilience to internal pressures (for example, busy rosters), and external pressures (for example, adverse weather, technical faults, delays and unruly passengers)” [37].

3 Research Design & Methodology

3.1 Introduction to Research & Research Question

Overall, this research seeks to investigate the relationship between work related stress (WRS), pilot wellbeing, pilot performance and flight safety. To this end, this research poses several related questions:
1. How do pilots currently perceive wellbeing and mental health issues?
2. Does the job impact on wellbeing?
3. Are pilots suffering because of the job (sources of WRS)?
4. What are the sources of WRS?
5. Do wellbeing issues have an impact on performance/safety?

3.2 Introduction to Workshops & Associated Workshop Objectives

Drawing on prior research relating to the biopsychological model of pilot lived experience [15], [16], [17], three workshops were undertaken with thirty-three commercial pilots (workshop 1: N=12, workshop 2: N=10, workshop 3: N=11). Specifically, the workshops had three objectives:

- To validate prior research relating to the impact of the job/WRS on pilot wellbeing.
- To map the relationship between WRS, pilot wellbeing, pilot performance and flight safety.
- To validate preliminary workshop findings related to the relationship between WRS, pilot wellbeing, pilot performance and flight safety.

Overall, the workshop methodology integrated participatory evaluation [38] and stakeholder evaluation approaches [39].

3.3 Procedure

In each case, participants were provided with briefing information seven days in advance of the workshop. The briefing included information about the biopsychosocial model of pilot lived experience (see Appendix 2), sources of work related stress (WRS) and a preliminary safety case outlining the potential relationship between WRS, pilot wellbeing, pilot performance and flight safety (see Appendix 3). This briefing information reflected a summation of prior research undertaken by the authors with commercial pilots [8]. Prior to commencing each workshop, participants were briefed about confidentiality issues. All participants then provided written consent - agreeing to maintain confidentiality in relation to anything disclosed by workshop attendees. Participants were then provided with a short presentation pertaining to the biopsychosocial model of pilot lived experience, the preliminary safety case and associated worked examples. In workshop 1, participants were invited to review both the model, the safety case and associated worked examples. This was followed by a group discussion concerning the relationship between WRS, pilot wellbeing, pilot performance
and flight safety. In workshop 2 and 3, the findings of workshop 1 were presented to participants. This included six scenarios pertaining to the impact of WRS on wellbeing, performance and flight safety. Participants were invited to review/validate the six scenarios. Following this, there was a general discussion about the relationship between WRS, pilot wellbeing, pilot performance and flight safety. All participants were invited to complete a homework exercise after the workshop. All participants were debriefed, at the end of each workshop. The debriefing included information about follow up supports and confidentiality. For more information about workshop procedures, please see Appendix 1.

3.4 Participants

Overall, thirty-three commercial pilots (spanning three airlines) attended the workshops. Workshop participants had on average 9,178 hours of flying experience, and included 20 Captains and 13 First Officers. Of the 33 participants, 7 were female and 26 were male. 8 participants had part time work contracts, while 25 were working full-time. In terms of flight operations, this included 4 short range, 7 long range, and 22 mid-range pilots.

3.5 Ethics

This research obtained ethics approval from the School of Psychology, Trinity College Dublin (TCD), Ireland.

4 Results

4.1 Pilot Wellbeing/Mental Health

It was agreed that pilots may be reluctant to stand down or disclose mental health problems, given real concerns over the potential impact of this on their job (i.e. fears of losing their license and/or possible impact on future career progression). Participants also stated that the prevailing culture (i.e. machoism and stigma associated with mental health issues) presents significant challenges. As reported by participants, this contributes to a situation where there is a lack of awareness/openness about MH issues, MH issues are not being identified, and MH issues are not being addressed.
4.2 WRS & Unique Features of Pilot Job

In terms of the experience of WRS, pilots have much in common with other (a) shift workers, (b) remote workers, and (c) workers involved in safety critical/high stress operations (for example, Paramedics and Firefighters). This includes a high degree of responsibility, fatigue, limited breaks, working anti-social hours and shift work. Participants suggested that what is unique to pilots is (a) the combination of factors that exist and (b) the specific way in which these factors interact. Participants emphasized the ‘unnatural’ location of the work environment. It was stated that the remote and confined nature of the cockpit imposes certain physical, social and psychological constraints on pilots. As characterized by participants, ‘ambulance drivers can step out of the vehicle, stretch their legs and talk to somebody’. As stated by one participant, ‘we are five miles up in the sky with an aircraft strapped to us’. Unlike other occupations, ‘there is no getting out of the aircraft’.

4.3 Sources of Work Related Stress (WRS)

Overall, there was a consensus that “year to year, the job is getting harder”, that “it is significantly harder than ten years ago” and that “pilots are working hard consistently through the year now, and there is no let up, or opportunity to recharge the batteries”. It was agreed that both work and personal stressors either (1) acting on their own and/or (2) acting together, put pilots in a situation where they are at increased risk of developing a MH issue, and/or worsening a pre-existing MH issue. Participants highlighted the potential impact of personal stressors which can be intensified/made worse by certain features of the job (for example, time away from home and inability to contact family while in work). Participants provided feedback as to sources of WRS and the potential impact on pilot wellbeing. Participants indicated that the key sources of WRS include the following:

1. Fatigue, potentially leading to burnout
2. Unnatural workspace (5 miles up in the sky)
3. Sleep disruption
4. Lack of breaks
5. Time away from home
6. Close confines of cockpit
7. Social isolation
8. Having different goals and values to management
9. Lack of management engagement with pilots
10. Lack of support from flight operations and management
11. Imposed sedentary nature of job
A full list of sources of WRS (aggregate of 3 workshops) is provided in Appendix 4. Participants noted that certain challenges are associated with different types of operations (short, mid and long range). For example, multi-sector days can be very tiring, with little or no breaks. However, in such operations, “pilots generally end up home at end of the day”. In long range operations (Ireland to USA), pilots obtain breaks. However, much time is spent away from home and crew can feel quite isolated. Specific feedback was provided in relation to all three pillars. For more information, please see Appendix 5, 6 and 7.

4.4 Impact of wellbeing on Pilot Performance and Flight Safety

Overall, participants agreed that pilot wellbeing is a significant factor in ensuring optimum performance and flight safety. Further, it was agreed that safety is compromised, if a wellbeing issue is not addressed. Participants agreed that aspects of the job cause stress, which in turn impacts on performance and by implication, flight safety. Further, life events (i.e. personal stressors) occurring outside of work can influence performance at work. In addition, the intensity/complexity of operational situation on the day can cause stress (i.e. bad weather, difficult interactions with management, operational & ground staff, complex routing and so forth). If prior stress exists, such operational complexity can worsen any pre-existing stress. Critically, in mapping the impact of WRS on pilot performance and flight safety, three distinct strands/sources of WRS must be considered

- General features of the job that cause stress/increase risk of MH Issue (i.e. what is in the model of lived experience)
- Current personal stressors (i.e. sick parent or child, unpaid mortgage, relationship problems)
- Current operational situation

In relation to conceptualizing the impact of wellbeing on performance and flight safety, participants noted that there are many factors to consider, and the specific impact of these factors on performance and flight safety is hard to quantify. Participants remarked that both (a) the specific constellation of factors occurring at any one time (i.e. general features of job/WRS, personal stressors, the operational situation), (b) how these factors might interact and (c) how these factors might potentially impinge on wellbeing and by implication performance and flight safety, is hard to predict. This is complicated by the fact that individual differences in relation to pilot coping ability must be considered. As reported by participants, Pilots are coping all the time. As stated by one participant, ‘pilots are managing stress, adapting
to the job and its challenges, and not having safety events/accidents’. As stated by participants, ‘some pilots cope better than others’. Specifically, ‘they have developed strategies to cope with the challenges they face’. It was noted that the general estimation amongst pilots is that ‘70% cope well, while 30% find adapting more difficult’. However, participants agreed that ‘pilots show up to work and tick all the boxes’. Further, ‘things don’t give until the very end’. As observed by participants, the fact that pilots are coping, presents its own risk. Critically, this masks the suffering that is experienced by pilots, and gives the impression that safety risks are being managed. In terms of specific impacts on performance, participants highlighted issues around impact on cognition, workload management, teamwork and communication. Specifically, participants referred to the following:

1. Potential reduction in situation awareness
2. Impaired decision making
3. Inability to focus on the current task
4. Difficulties managing multiple tasks/workload
5. Task omissions
6. Reduction in quality of error identification and management behavior
7. Poor quality communications with fellow pilot
8. Withdrawal of pilot (not communicating)

4.5 Estimating Impact & Preliminary Safety Case/Worked Examples

In relation to workshop 1, the initial safety cases/worked examples were presented to the group. For an example of this (psychological pillar), please see Appendix 3. It was noted that the initial safety case seemed contrived. Participants suggested that the worked examples should (a) involve WRS issues which span the three pillars of wellbeing, (b) consider the specific operational circumstances on the day, (c) consider pilot coping ability (and variance), and (d) consider the risk mitigating role of the other pilot.

4.6 Estimating Impact & Example Scenarios

In workshop 1, participants proposed six scenarios that might be considered in relation to conceptualizing the potential impact of WRS on wellbeing, performance and flight safety (see Table 1). As pointed out by participants, not all scenarios have direct implications in relation to performance/flight safety. For example, see scenarios (4) and (5). In terms of modelling impact of WRS on wellbeing, pilot performance and flight safety, participants noted that the more typical situation is scenario (1). As reported by participants,
this scenario reflects the typical situation of most pilots. It was stated that scenario (3) is the most likely scenario in relation to specific impact on safety and this is where attention should be focused. In relation to scenario (1), participants noted that this reflects day to day operations. Pilots are suffering, but they are also adapting and coping. Periodically, this suffering has an impact on performance – but this impact is typically managed (i.e. the pilot recognizes/identifies mistake/omission, and there is minimal safety impact). Participants agreed that many pilots experience scenario (2). The specific impacts on wellbeing (i.e. GI issues and musculoskeletal problems) are quite common, but do not impact on performance and flight safety. Overall, it was suggested that research focus on scenario (3), where there is a potential for something more serious/safety event. Typically, in such a scenario, a chain of events ensues. At a certain point, the pilot stops coping and ‘goes over the edge’. Importantly, this does not always lead to a ‘crisis situation’ and/or ‘safety event’. Performance issues are managed by the other pilot. As such, the other pilot acts as a barrier in mitigating the potential safety impact. In relation to their own direct experience, workshop participants provided several examples of such cases. Scenarios (4) and (5) primarily focus on the impact of WRS on wellbeing. In relation to scenario (5), participants noted that this situation is very real and is not discussed. In relation to workshop 1, all twelve participants were familiar with such cases (i.e. aware of other pilots who are suffering or who have committed suicide). All participants noted that there should be a stronger focus on preventing suffering/harm to pilots. In relation to workshop 2, five of the participants had direct personal knowledge of such cases. In relation to workshop 3, several participants had direct personal knowledge of such cases. As stated by participants, it was noted that scenario (6) specifically pertains to a person who might have a pre-existing MH issue. Currently, such a person is not obtaining adequate supports at an airline level. Participants noted that such a scenario is comparable to the Germanwings accident. Participants noted that pilots commuting/foreign First Officers based away from home are at risk (i.e. might be suffering from fatigue and experiencing significant social isolation). Further, if there is a pre-existing MH issue, then this risk is amplified.

4.7 Scenarios & Further Elaboration of Impact

In workshops 2 and 3, participants provided feedback about the scenarios defined in workshop 1 – specifically in relation to impact on (1) wellbeing, (2) performance and (3) safety. In relation to (2), participants also provided an estimation of the frequency such a situation would arise. It should be noted that this estimation was subjective.
Table 1. Wellbeing Scenarios

<table>
<thead>
<tr>
<th>#</th>
<th>High Level Scenario</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pilots mostly coping well – pilots may make the odd mistake but will identify this and correct actions/behavior – no safety impact/negligible</td>
<td>Frequent</td>
</tr>
<tr>
<td>2</td>
<td>Pilots mostly coping well, but impacts on physical health (i.e. GI issues and musculoskeletal problems) - no safety impact/negligible</td>
<td>Frequent</td>
</tr>
<tr>
<td>3</td>
<td>Pilots experiencing difficulties but mostly coping – however, something gives on the day – potential for safety event – but managed by other pilot, who acts as barrier to safety event</td>
<td>Infrequent, but does happen</td>
</tr>
<tr>
<td>4</td>
<td>Pilots mostly coping well, but long-term impacts on health (evidence when on annual leave, or when retire) – sudden illnesses, early death</td>
<td>Frequent</td>
</tr>
<tr>
<td>5</td>
<td>Pilots not coping – impact on wellbeing – stop working, develop MH issues, worst case is self-harm and suicide – no safety impact</td>
<td>Infrequent, but does happen</td>
</tr>
<tr>
<td>6</td>
<td>Extreme cases – murder-suicide (Germanwings) – implication for safety /fatal accident</td>
<td>Exceptionally rare</td>
</tr>
</tbody>
</table>

5 Discussion

5.1 Establishing Prevalence of MH Issues & Impact on Wellbeing

Workshop feedback highlights the need to establish the prevalence/incidence of MH issues amongst pilots and the overall impact of the job on pilot wellbeing. This might take the form of an anonymous survey (i.e. non-jeopardy). Without a measure of the extent of this problem and how it is impacting on pilots, it may be hard to convince other stakeholders (for example, both pilots and airlines) of the need to address this issue.

5.2 Conceptualizing & Measuring Safety: Have we got it right?

Currently, we are measuring safety solely by the number of “aircraft crashes”. The application of this measure leads (1) to judgements that we are ultra-safe, (2) to risk/safety management approaches that focus on near misses, safety events and accidents and (3) the application of safety performance indicators that focus on the socio-cognitive dimensions of performance (i.e. situational awareness and teamwork) and associated ‘aircraft state’ outcomes (i.e. flight level busts, overspeed in approach, runway over-runs etc). Critically, this research indicates that we need to question the validity/appropriateness of such an approach. Specifically, we need to question the current avoidance/overlooking of using metrics and safety performance indicators that address certain fundamental dimensions of human performance, such as
factors associated with pilot wellbeing and WRS. Potentially, the current framework and associated metrics and KPI result in a false/incomplete picture in relation to (1) understanding routine performance (i.e. pilots adapting/safety being maintained), (2) understanding why accidents happen and (3) making flight safety estimates. Given this, it could be argued that there are vulnerabilities in the existing approach to risk/safety management (i.e. proactive techniques are not considering MH/wellbeing and dimensions linked to WRS). As indicated in this research, if we use a different evaluation metric (for example, consider metrics and KPI linked to wellbeing and WRS), we might conclude that we are far from “Ultra-Safe” and that a significant number of safety risks (i.e. wellbeing/MH) are not being managed. Moreover, we are missing important outcomes linked to pilot suffering and wellbeing (see scenario 2, 4 and 5). Crucially, this research indicates that pilots are coping with significant challenges/sources of WRS (scenario 1). If WRS leads to a potential error (scenario 3), this is typically identified and managed by the copilot. The fact that pilots are adapting/coping and working effectively as part of a team is important. But it should not be used to underestimate or mask safety issues, or wellbeing impact (scenario 2, 4 and 5).

5.3 Addressing Routine Suffering

Recent attention on the potential safety impact of MH issues (for example, Germanwings/Scenario 6), does not serve the overall needs of pilots. More attention needs to be given to issues around routine suffering and its impact on both wellbeing (i.e. scenario 4 and 5) and safety (i.e. scenarios 1, 2 and 3).

5.4 Managing Wellbeing issues

Pilot wellbeing/MH issues need to be treated as both an Occupational Safety & Health (OSH) issue and a Flight Safety issue, and roles need to be defined and funded, so that appropriate supports for pilots can be provided. As highlighted by workshop participants, much work is required in relation to promoting awareness of wellbeing/MH problems amongst pilots and normalizing this issue. In relation to both pilot culture and airline/organizational culture, MH issues for pilots need to be destigmatized. Pilots need to be encouraged to put their hands up if they are experiencing difficulties. Critically, pilots will not do this if they believe the outcome will be punitive (i.e. loss of license, impact on career progression). The current perceived punitive culture presents a clear threat to the wellbeing of pilots, and to flight safety. At an early stage (i.e. initial Human Factors training), pilots need to be trained in terms of (a) self-managing wellbeing issues and (b)
risk identifying in relation to their own wellbeing/MH (i.e. detecting potential for problem/problem in self and managing this). In relation to (a), current training does not focus on the promotion of resilience and the development of coping skills (i.e. learning how to be resilient to challenges and practice self-management techniques). In relation to (b), checklists might be developed to support pilots to identify MH risks. This might build on prior research in relation to the application of TEM concepts to the specification of an intelligent flight plan, supporting pre-flight planning and briefing [40], [41].

5.5 Fatigue, Flight Time Limitations & Pilot Wellbeing

European Airline Pilots are exempt from the European Working Time Directive 2003 [42]. Instead, duty hours and rest periods are governed by Flight Time Limitations (FTLs), the purpose of which is to ensure air safety by protecting against the effects of fatigue [43, 44]. While intended to protect against the risks posed by tired pilots, FTLs do not protect against the other issues that pose a challenge to maintaining the different biopsychosocial dimensions of pilot wellbeing (for example, time away from home, working anti-social hours etc.).

5.6 Single Crew Concepts

Given the mitigating role of the ‘other pilot’ in terms of acting as a safety barrier (see Scenario 3), this raises issues in relation to the safety implications of single crew operations. As indicated in this research, the potential for a safety event might increase, without the supporting role of the second pilot. Further, as identified in the biopsychosocial model of pilot lived experience, social isolation and lack of peer support is a risk factor in relation to the development of MH issues.

5.7 Limitations of Research, Areas for Further Research & Next Steps

The findings presented reflect preliminary research involving a small number of participants (total of 33). Participant estimations of impact of sources of WRS on (1) wellbeing, (2) performance, and (3) flight safety are subjective. The specific estimations reflect the consensus view of participants attending Workshop 2 and 3. In relation to conceptualizing and measuring the impact on wellbeing and on performance/safety, further work is required.
<table>
<thead>
<tr>
<th>#</th>
<th>High Scenario</th>
<th>Level</th>
<th>WRS</th>
<th>Wellbeing Impact</th>
<th>Impact Performance</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pilots mostly coping well</td>
<td>Any pillar – include fatigue, social isolation</td>
<td>Minor impact</td>
<td>Minor impact</td>
<td>Frequent</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Pilots mostly coping well, but impacts on physical health (GI, musculoskeletal problems)</td>
<td>Mostly biological pillar</td>
<td>Minor impact - suffering in daily life</td>
<td>Minor impact</td>
<td>Frequent</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Pilots experiencing difficulties but mostly coping – however, something gives on the day – potential for event but co-pilot acts as barrier to compensate for the other pilot</td>
<td>Complex combination of personal factors, work factors and operational situation on the day</td>
<td>Significant/considerable impact</td>
<td>Loss of attention</td>
<td>Infrequent</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Pilots mostly coping well, but long-term impacts</td>
<td>Any pillar – include fatigue, social isolation etc</td>
<td>Long term impact on health – develop illness when on annual leave or when retire</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Pilots not coping – impact on wellbeing</td>
<td>Complex combination of personal factors and work factors</td>
<td>Significant impact - suffering in daily life – stop working – potential for serious MH issues including self-harm and suicide</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Extreme cases – murder/suicide (Germanwings) – media attention - implication for safety – very rare – should not be the immediate focus</td>
<td>Potentially pre-existing MH issue</td>
<td>Major impact</td>
<td>Major impact</td>
<td>Exceptionally Rare</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Impact – no impact/neutral, negligible impact, minor impact, significant/considerable impact, major impact.

Frequency: exceptionally rare, rare, infrequent, frequent, very frequent.
Critically, the six scenarios (and associated dimensions and measures defined in Table 2) provide an initial explanation of the relationship between WRS, pilot wellbeing and flight safety. Evidently, these six scenarios require further validation with commercial pilots and other stakeholders. It is anticipated that such feedback (scenario definition, frequency, impact) will be elicited as part of an upcoming wellbeing survey.

6 Conclusions

There is evidence that pilots are under stress and experiencing wellbeing/mental health problems. Specific features of the job can result in wellbeing problems and increase a pilot’s risk in relation to developing a MH issue. Further, personal stressors (i.e. factors outside the job), can be intensified/worsened given specific features of the job. Pilot wellness is a significant performance shaping factor in terms of ensuring optimum performance. Overall, pilots are managing wellbeing issues. In general, pilots try to normalize/adapt to the job. However, there is much variation in relation to coping ability. This variation needs to be considered in relation to modelling performance/safety impact. Six scenarios were identified. Of these, participants suggested that the primary focus should be on the prevention of Scenario 3 (i.e. something giving on the day, impacting on flight safety) and Scenario 5 (i.e. pilot suffering which ends in harm to the person). The relationship between positive mental health and pilot professionalism and flight safety needs to be understood and supported at different levels. The goal is not simply to prevent catastrophes. Rather it is to deal with wellbeing issues among pilots (routine suffering/adapting) and address safety implications (Scenario 3). Not all MH issues result in a fatal accident/catastrophe (Scenario 6). Potentially, the current framework and associated metrics and KPI result in a false/incomplete picture in relation to (1) understanding routine performance (pilots adapting/safety being maintained), (2) understanding why accidents happen and (3) making flight safety estimates. Given this, it could be argued that there are vulnerabilities in the existing approach to risk/safety management (i.e. proactive techniques are not considering MH/wellbeing and dimensions linked to WRS). Overall, pilots need to be trained in relation to (1) coping strategies and (2) risk identifying behavior. Safety is enhanced when a wellbeing issue is addressed and supported, as opposed to allowing it to go undiagnosed and untreated. Pilots with wellbeing/mental health difficulties for the most part are perfectly able to continue to do their jobs, especially with support. Mental health is a normal part of health and needs to be treated and/or managed accordingly.
References


Appendix 1: Workshop Procedures

Table 3. Workshop Procedures

<table>
<thead>
<tr>
<th>#</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Welcome</td>
</tr>
<tr>
<td>2</td>
<td>Introduce team and explain objectives</td>
</tr>
<tr>
<td>3</td>
<td>Participants provide written consent and complete profile form</td>
</tr>
<tr>
<td>4</td>
<td>Introduce model – background, research overview</td>
</tr>
<tr>
<td>5</td>
<td>General discussion (feedback on model)</td>
</tr>
<tr>
<td>6</td>
<td>Validate model with group</td>
</tr>
<tr>
<td>7</td>
<td>Introduce safety case</td>
</tr>
<tr>
<td>8</td>
<td>Validate safety case with group</td>
</tr>
<tr>
<td>9</td>
<td>Overview of homework/feedback (optional)</td>
</tr>
<tr>
<td>10</td>
<td>Debriefing, thanks &amp; Wrap up</td>
</tr>
</tbody>
</table>
Appendix 2: High Level Biopsychosocial Model of Pilot Wellbeing – The Lived Experience

Fig. 1. Biopsychosocial Model of Pilot Wellbeing: The Lived Experience
Appendix 3: Preliminary Safety Case (Psychological Pillar)

Fig. 2. Worked Example: Psychological Pillar
Appendix 4: Workshop Findings: Sources of Work Related Stress

**Sources of Work Related Stress (WRS)**
Factors impinging simultaneously on more than one pillar are highlighted with an *.

<table>
<thead>
<tr>
<th>Biological</th>
<th>Psychological</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working irregular hours*</td>
<td>Working anti-social hours*</td>
<td>Divergence of values between management and pilots*</td>
</tr>
<tr>
<td>Working within the close confines of the cockpit*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unnatural location of work environment (5 miles up in the sky – no support/can’t step out)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working long duties</td>
<td>Increased responsibility with reducing authority/support*</td>
<td></td>
</tr>
<tr>
<td>Difficulties accessing fresh, healthy food</td>
<td>Lack of engagement (management and pilots)*</td>
<td></td>
</tr>
<tr>
<td>Sedentary nature of working as a pilot</td>
<td>Perception of pilots possessing &quot;The Right Stuff&quot;*</td>
<td></td>
</tr>
<tr>
<td>Cockpit environment – air quality, oxygen level, noise</td>
<td>Perception that pilots are &quot;living the dream&quot;**</td>
<td></td>
</tr>
<tr>
<td>Time away from home*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not having a sense of home/never at home*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of certainty in relation to roster (changes)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being contacted by work when off duty if staffing/roster issues*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managing and understanding cultural differences (international workforce)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commuting lifestyle*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long working day in close contact with one other person (may or may not get on with)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Captain responsibility – never switch off</td>
<td>Working inflexible/disruptive schedules</td>
<td></td>
</tr>
<tr>
<td>Changing nature of the industry</td>
<td>Inflexible annual leave allocations</td>
<td></td>
</tr>
<tr>
<td>Automation and prolonged periods of low stimulation</td>
<td>Ever-changing crew composition</td>
<td></td>
</tr>
<tr>
<td>High training costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stringent medical certification</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Sources of WRS
Appendix 5: Updated Model: Pilot Wellbeing, The Lived Experience: Biological Pillar

Fig. 4. Updated Model: Biological Pillar

Fig. 5. Updated Model: Pilot Wellbeing, The Lived Experience: Psychological Pillar

Fig. 6. Updated Model: Pilot Wellbeing, The Lived Experience: Social Pillar
Human Resource Recovery: Does Hypnosis Help Manage Fatigue When Workload is Increased?

Marielle Plat Robain

1 Airbus Operations SAS, 316 route de Bayonne, B.P. M0151/0
31060 Toulouse Cedex 9, France
marielle.plat-robain@airbus.com

Abstract. This exploratory work aims to look at new techniques, in this case hypnosis, which has already been used in very long flights by Bertrand Piccard (Solar Impulse) to manage fatigue and resting periods. We used Multi-Attribute Task Battery II (MATB-II) software to induce different workload conditions on two studies: (1) less intense scenario (10mn, 55 events) and (2) more intense scenario (20mn, 155 events). The goal was to induce fatigue from high workload. For each participant, the scenarios were repeated twice, separated by one of three break conditions of fifteen minutes: "static rest" or "exercise" or "hypnosis" breaks. The hypothesis is that: depending on the break condition used and scenario intensity level, variations on performance and fatigue recovery will occur. The evaluation method used eye-tracking data, task performance index, fatigue and workload subjective scales plus debriefing. The results show that "hypnosis" and "exercise" breaks have interesting effects on fatigue and performance but results vary upon scenario event numbers and duration.

Keywords: Aviation, hypnosis, cognitive resources, fatigue, stress, mental workload, multitask, performance.

1 Introduction: Coping with Resources and Fatigue, the Stake for Human Factors in Aeronautics

In aeronautical workplace context, while piloting, the resource management is a prevalent topic. While coping with fatigue during long flights or multiple daily rotations this may have an impact on resources availability. Resources solicitation is quite variable during a flight and depends on flight phases: Flight preparation, taxi, take off, initial climb, as well as final approach solicit more pilot resources than when automatisms are activated. Automatisms alleviate pilot workload during different flight phases such as climb, cruise, and descent. The crew have to regulate those variations of resource
management. It is more interesting to start by looking at medium to high workload conditions as those phases are more demanding and depend on human performance. Aircrew is often exposed to « fatigue factors »: irregular hours of work, long duty times, jet lag, etc. Even if preventive solutions have been developed to cope with this issue; regulations, fatigue management training, flight data monitoring, fatigue risk management systems, crew rest, it is of interest for safety and human wellbeing to explore new techniques that allow recovery or better management of the human resources and cognitive cost. That is why, considering the recent example of the solar impulse, pilots using alternative techniques like yoga or the Erickson hypnosis techniques [1][2] seems like an interesting approach to explore. They used those techniques to manage their resources and their lack of sleep during very long flight and simulation (72 hours) where they do not have a fully automatic pilot to help with the workload except for wing stabilization for twenty minutes. This small window of time allows pilots to take naps or regulate their resources using different consciousness states such as hypnosis. There are no scientific references about this work, only video testimonies and both pilots’ interviews. One of the videos has been recorded during Bertrand Piccard’s 72 hour simulation (with EEG capture). It is estimated that he slept only 2 or 3 hours in 24 hours according to the EEG measurements and that he was managing his resources in the airplane with hypnosis the rest of the time. He declares that he feels even better when using autohypnosis. This exploratory work aims to shed light on the potential cognitive resources recovery via hypnosis [3] compared to other short break types.

2 Related work

As the crew cope with variable mental workload during flight it seems relevant to look at related work. Regarding mental workload, human factors researchers didn’t converge to a unique definition. First belief was that when task difficulty increases performance decreases [4]. In this exogenous view, the mental workload is defined as a strain consecutive to task constraint but it has to be confronted to the operator capacity. Wickens capacity definition is “maximum limit of the information treatment” [5]. Then the operator is modulating the employed capacity to respond to tasks with adaptive strategies. Moray [6] is underlining that the operator dispose to the will of involving more or less capabilities in order to regulate performance. Eggemeier et. al [7] define resources as capacity proportionality. But even if this line of thought was promising it seems that the mental processing is limited and that resource increase does not equate direct performance increase. Cegarra [8] is
using the metaphor of the cognitive engine where gasoline/resources are the combustibles that cope with tasks. He precisesthat this combustible is faster used when coping also with stress. He emphasized that the task structure may induce a constrain on the consumption. This could be due to the processing structures: stages of processing dimension, codes of processing dimension and modalities dimension that use different resources such as emphasized by Wickens [9] in his 4-D multiple resource model. Even if some perceptual channels probably allow a parallel resources use, some others compete on the same resource and produce a constrained consumption speed of the resources that may be not sufficient to cope with tasks constrains. Kahneman [10] defines a model with capability and energy, proposing effort direct measurement in the sympathetic central system like: pupil dilation, cardiac measure, and electro-dermal activity. Effort definion is a way to allocate resources regarding internal versus external regulations to cope with tasks’ requirements. Moray adds a temporal axis that allows coping with anticipation (strategic control) or not. Strategic control is a requirement from the piloting tasks Jouanneaux [11] as a pilot should not be “behind his aircraft”. It is well known that pilots deal with a different temporality of their action regarding the different flight phases and automatism involvement. This temporal constraint also impacts on diagnosis depth while coping with incidental situations as shown by Plat [12]. Another axis regarding strategies is the knowledge level of operator regard to the situation that is shown by Hoc and al. model [13]. That is why pilots are really qualified people; as well as having a professional licence, type qualifications for every aircraft, they have regular training and exams. That’s why Hoc et. al [14] emphasized the capability of managing resources as a key factor of pilot expertise; the operator will change strategies thanks to goals and consecutive capability involvements. Cegarra [8] qualifies the mental workload as the variable that explains cognitive models. As in Hoc et. al’s model [13] Cegarra [8] makes the link between Rasmussen’s SRK model involving different resources level with three different related tasks’ type: mouse designation (skill), observation and comparison (rules), process management and regulation (knowledge). He shows that subjective measures like NASA TLX [15] on each separated task increases with cognitive control level requirements, and is different from the temporal dimensions variation. Operator temporal demand would correspond to a “mental effort” and “frustration” scoring, but not to the other dimensions. It seems that the “effort dimension” is a real good indicator to discriminate tasks. As we already defined the effort, it is related to the way we allocate resources. We easily understand that resources management seems to be a central concern for human dealings in tasks, especially when multitasking in a dynamic environment, such as flying an aircraft. Therefore, we ask: Do short
breaks have an impact on resources and fatigue within different workload levels? Does hypnosis used on short breaks have a better positive effect on managing workload, resource, and fatigue than light sport breaks do? Hypnosis related works show some effect on cognition. It is known to trigger many changes at the subjective level, such as motivation, relaxation, and mental absorption [16]. The link between cognitive resources and hypnosis does not seem to have been studied in experimental psychology. From the mid twentieth century, researchers tried to explore more objectively the effects of hypnosis on cognitive performance. A study conducted on highly suggestible students showed that besides effects on motivation, hypnosis suggestion improved their performance during a test related to schoolwork [17]. However, performance improvement could result from enhanced attention and motivation. Following this pioneer study, many research teams used the Stroop task (or Stroop-like paradigms) to bring to light hypnotically induced modifications of cognitive mechanisms. On one hand, during hypnotic state and without opposite instruction, Stroop effect appears to be enhanced, i.e. subjects make more mistakes at naming colour of a word when the word indicates a different colour [18], whereas on the other hand post-hypnotic suggestion can cancel Stroop effect [19] [20] [21], by inducing transient alexia (i.e. inability to recognize words, as if they were written in a foreign unknown language). Those studies show that there are two potential impacts: one coming from the hypnotic state effect itself, and the other one from the suggestions that have been induced. The effect of hypnotic suggestion on several well-documented cognitive mechanisms has been studied [20]. They successfully significantly reduced the McGurk effect [22] by suggesting to the participants to focus only on auditory channel [21]. Other types of processes can be modified by hypnosis such as Simon’s effect [23] and Flanker effect [24]. Other studies succeeded in suppressing temporarily synesthetic associations, both at behavioural and electrophysiological levels [25]. All the above cited cognitive processes are considered as automatic processing because they occur outside conscious control and they don’t need to be learned. For instance, with an appropriate post-hypnotic suggestion it turned out that it is possible to focus attention. The above studies investigate the possibility to automatize the demand on cognitive processes thanks to hypnotic suggestion [21]. They studies explored this hypothesis using two well documented paradigms: the masked diamond paradigm and the visual search [21]. In most of those studies participants were first tested for susceptibility to hypnosis using normalized scales such as Stanford Hypnotic Susceptibility Scale Form C [19] or Harvard Group Scale of Hypnotic Susceptibility Form A [21]. By modifying a wide spectrum of cognitive processes, hypnosis can thus be a great means to study and understand those processes. Moreover, sport
psychology studied the effect of hypnosis on sport performance [26][27][28], finding that participants increased their mean short-term performance and their mean flow scores from baseline to intervention with moderate exercise. Results also indicated that during the intervention phase participants had felt more relaxed, calm, determined, happy and focused when compared to the baseline phase.

2.1 “Human Recovery Resource” Project: an Exploratory Work

Erickson hypnosis is a confidential technique, mostly used and trained for in the medical field and for different goals. Doctors trained in hypnosis use it also for their own fatigue and resources regulation all day long. Moreover, in neurophysiology more people are studying different consciousness states like meditation [29] and hypnosis [30][31] in order to explore potential benefits. In the present work, we are interested in cognitive resource recovery through hypnosis. We compare hypnosis sessions to physical exercise sessions, whose effects on cognitive abilities are known to follow Yerkes and Dodson’s Law of arousal [26]. This means that cognitive performances improve following a moderate to high intensity exercise on short to medium duration (less than 1 hour), whereas they decrease for very high intensity or long lasting activity [27]. Moreover, physical exercise does not affect all cognitive abilities in the same way and it appears that working memory is the most enhanced, whereas perceptual functions are hardly affected. Problem solving and other executive functions are moderately but robustly improved following exercise [28]. Hence our question: What about hypnosis?

3 “Human Recovery Resource” Studies (HRR1/HRR2)

We worked on two studies, one after the other, and we present them here to sustain the above question: the first experiment HRR1 was an exploratory study to investigate if there was any interest in a short break hypnosis session to recover, and the second experiment HHR2 was implemented to replicate and mature this work in a higher workload exposition.

3.1 HRR Assumptions

We assume that short breaks with exercise and short breaks with hypnosis have a better impact on resources management and fatigue than short
discussion breaks (just stop working and have a discussion without moving) in different workload conditions.

3.2 Methodology:

3.2.1 Methodology Common Points for Both Experiments: HRR1 & HRR2

Apparatus:
Participants are volunteer participants from the engineering office in Airbus. Most of them are engineers in aeronautics. Instructions are: they participate in an experiment to assess the effect of breaks and that they have to perform as well as they can. If they did perform well, they were given sweets (in order to create a motivation). For the performance they were seated in front of a computer screen (19 inches and approximate distance: 60cm) in a windowless room with controlled and constant lighting for eye tracking measurements. Materials to perform tasks were: Thrustmaster T.FlightStick X joystick to perform the “TRACK” task, a keyboard or a mouse for other tasks, at operator discretion. Position of each element on the desk could be adjusted so participants feel as comfortable as possible. Eye activity was recorded through an eye-tracking bar device, (Tobii X1 Light), fixed under the screen and the Ogama software. The device was calibrated to every participant before each scenario. This device might help to measure arousal (blinks, saccades), sympathetic system changes (pupil diameter) and efficiency scanning. They had to perform a Multi-Attribute Task Battery on a computer (NASA MATBII [20]) which is a multi-task developed by NASA to assess human workload, mimicking a piloting situation. We selected this battery as current cockpit interactions and design office workplaces include multi-solicitations of operators. MATBII is composed of 4 sub-tasks (cf figure 1) that are configurable in difficulties and number of events. So we will describe how we tuned tasks’ structure in order to induce additional constraints on the resources consumption: The “System Monitoring Task” called “SYSMON” is composed of two sub tasks: warning lights and scales. These subtasks are detection and reaction time tasks. The sub tasks difficulties consist of: for warning buttons zone, one of the warning buttons is obvious to detect but the other is less obviously detectable. Concerning the cursors on the scales to be monitored, the difficulty is that they are in constant movement even when locked at the extremity of the scales, which makes them more difficult to detect. Since the scales are in the peripheral view they require regular scanning for, participants. Those sub-tasks could be considered as a rule.
cognitive control level referring to SRK model used by Cegarra [8] with a time constraint. Consequently, they require more resources than a skill based task even if using visual processing and manual spatial responses according to Wickens [9]. The "Tracking Task" called "TRACK" is a tracking and fine motor task. It has to be performed when in manual mode (no intervention is required when in automatic mode). Accuracy is measured when the cursor is kept most in the centre by using a side stick. Moreover, cursor movements are randomly tuned which reinforces the difficulty of the task as this randomization does not allow the operator to anticipate the cursor behaviour while attempting correction. This could be compared to an operational situation of turbulence flying zone which is manually managed by pilots. This task could be considered as a skill cognitive control level referring to SRK model used by Cegarra [8] and will be here considered closer to a higher cognitive control level (more resources consumption) in our experiment. This task uses visual and spatial modalities and manual spatial responses according to Wickens [9].

The "Communication Tasks" are called "COMM" and use operator audio perception, language comprehension, short term memory plus motor selection task. This task for non-native English speakers is also a demanding one in terms of resources. This task could be considered as a rule cognitive control level referring to SRK model used by Cegarra [8] and requires a quick handling or a very good working memory. This task uses auditory, visual, verbal modalities, and manual verbal (even if only numbers are being used) responses according to Wickens [9].

The "Resource Management Task" is called "RESMON" and consists in a fuel management task which evolved slowly. This task requires from operators to observe the system architecture and evolutions in order to create mental model of the system dynamics and find some strategies to manage it. It could also require some short time decisions to be taken while failures occur. This task could be considered as a knowledge cognitive control level referring to SRK model used done by Cegarra [8].

We wanted to vary modalities’ dimensions that use different resources as is emphasized by Wickens [9] in his 4-D multiple resource model. We aim then to produce different resource consumption and medium to high amount of workload; In a cockpit multi-tasks are performed by two operators, but in the current task the operator will be alone performing the task in high difficulty and time constraints. Time constraints were added to emulate the actual cockpit environment. All of those combinations aim to create resources solicitation during the activity and consecutively fatigue. Participants were all aware of the performance calculations and constraints that are described
Each subtask of the MATB is analysed separately, aiming to extract a unique indicator for the global performance. Our method of analysis is inspired from previous work [20]. The main indicator for performance calculation is the “time spent” in abnormal situation: For the “SYSMON” task if a participant misses an abnormal event, we assign the maximum amount of time (10 secs) for that response. In other cases we count the time before action. For the “TRACK” task, the performance index uses the time spent outside of the inner box but also takes into account the distance from the point (0.0) when inside the inner box. For the “COMM” Task, the performance index uses both the reaction time and the accuracy (correct radio and frequency). For the “RESMAN” task performance is calculated from target deviation using the time spent outside the acceptable range (between 1500 and 2500) defined by blue brackets. The associated index uses the time spent outside the range defined by the blue brackets.

**Workload Rating Scale (NASA-TLX):**

At the end of each MAT-B scenario, a Workload Rating Scale based on the NASA-TLX is prompted on the screen. NASA-TLX classifies workload in six categories: mental, physical, temporal, performance, effort and frustration [15]. Participants had to rate their global workload on these six scales from “low” to “high” (0/100) except for the performance category “good” to “poor”. This subjective assessment was done two times: end of scenario 1 and end of
scenario 2 in order to compare workload subjective feeling while acting on a same scenario before and after 15 minutes break. This comparison will give us a score about the workload subjective ratings regarding the break effects.

**Fatigue scale:**

We preferred to present a fatigue scale as everybody knows about fatigue but they do not necessarily know how to define resource management. Fatigue is a more common shared word and is here envisaged as a consequence of high resource involvement during activity. A subjective numeric scale [33], was used to assess participants’ fatigue four times during the experiment: (A) at arrival, (B) before break, (C) after break, (D) at the end of experiment. The participants had to rate their fatigue from 1 (No Fatigue) to 10 (Severe Fatigue), two additional indicators were presented above the scale to help participant rating, for 4 (Mild Fatigue) and 7 (Moderate Fatigue).

**Specific procedure for hypnosis break (third group):**

Before the test: a pre-test was done to check if each volunteer is suggestible enough to be able to participate in a “hypnosis break” session. This assessment was done with Standford Modified Scale of Paris Revised Form [34]. During the test: the hypnosis technique employed here is an “Erickson” practice, where subject and hypnotherapist sit on a standard chair close together [3]. The proposed exercise could also be employed with auto-hypnosis but for better reproducibility it was performed by a qualified psychologist trained at the Medical University of Toulouse (France) in Erikson Hypnosis techniques. The script (hypnotherapist verbalisations) is based on “positive energy” incomes and tensions’ outcomes related to the breath flow, plus success memory renewal and the associated sensations (positive emotions). We used the word renewal on purpose as it was demonstrated with magnetic resonance images that hypnosis induces a state close to a revival of the situation and does not just produce memory activation [35]. Positive emotions relative to the success memory were “energy”, “joy”, “pride”, “pleasure”, “strength”, “enthusiasm” and “one’s own resources mobilization”, which have an effect on cognition[36] and consequently on subjects’ motivation, engagement and effort (resources involvement) regarding to their activity[37][38]. A post-hypnotic suggestion was done while reorienting the subject “to be here in top form”. This exercise is slightly different from Piccard’s exercise, as Piccard induced deep sleep and here we aim to induce energy. We didn’t aim to reproduce Piccard’s exercise but to study the potentiality of such a technique. Both exercises could be compatible and complementary.
**Performance analysis:**
In order to focus on the effect of the break, we chose to look at the difference of performances between scenarios (same scenario) before and after break rather than the absolute performance. The difference was always calculated as: Performance after break – performance before break.

### 3.2.2 Method Specificities of the First Experimentation (HRR1):

**Participants:**
HRR1 exploratory study involved 19 volunteer participants, (10 female, 9 males between 23 and 55 years-old) from the engineering office in Airbus without any medical antecedent, or motor or visual impairment. They were randomly allocated for first two groups: “normal break” for group 1 (reference group), “fresh air break” for group 2. But for last third group “hypnosis break” participants were differently selected with a suggestibility criteria based on Standford Modified Scale of Paris Revised Form [34] pre-test.

**General procedure and scenarios workload:**
The procedure was performed between 12:30 and 15:00 just after lunch. This hour was chosen because of its higher probability of low cognitive performance and a bigger need for a break. Upon arrival, after experiment briefing, the participants had to fill a first fatigue scale (reference). Then a first MATB scenario was used for familiarization (5 minutes, 4 MATB tasks) repeatable if the participant didn’t reach a 60% performance threshold before running the test-phase (in order to avoid too much disparity in the results). The test phase was a longer and more difficult scenario (10 minutes, 55 MATB events) but less difficult than WAM workload calibration level 4, but still more difficult than their level 3 (only 3 tasks involved) [26]. We chose to increase time duration compared to WAM test to induce some fatigue. This first scenario is even more difficult than an operational nominal activity level, as the cockpit multi-task is shared between two pilots. Nevertheless, we aim here to induce more workload to test our assumption in a more competitive way. However, this may be experienced for instance when pilots do not use full automatisms in approach vectoring. Participants were given instructions to perform as well as they can.

After the first scenario of the test phase, participants were asked to rate their fatigue and were given a 15 minutes break according to their group:
- The “normal break” group (group 1) was given a break in which they couldn’t leave the room but could have a conversation.
The “fresh air” group (group 2) was given a break with moderate physical exercise in which they were instructed to climb down the stairs from the 5th floor then walk outside and climb back up the stairs to the 5th floor.

The “hypnosis” group (group 3) was given a break in which hypnotic procedure was proposed (see below) inside the test room.

After the break, participants were asked to rate their fatigue level (third time) and they were tested again with the same scenario than before the break (second scenario). After they were done they filled in the Nasa TLX and the fatigue scales again.

Variables and Statistical Analysis:
Data acquired on Fatigue evaluations were tested using a repeated-measure two-factor Anova with the first factor being “Time” with 4 modalities (arrival; before break; after break; after test) and the second factor being the “Group” (“normal break”; “fresh air break”; “hypnosis break”).

Data from Workload Rating Scales (WRS) obtained at the end of each MATB scenario were categorized for all the five dimensions: mental; physical; temporal performance; effort; frustration). Data from MATB performance was acquired for each subtask: System Monitoring; Tracking; Communication; Resources Management. Data from eye tracking was processed using Ogama software and several markers classically used as vigilance indicators were extracted. The duration and the frequency of blinks were extracted using a custom-based algorithm prepared under Matlab that explore the dataset and check for “loss of data” under a certain time in order to avoid real data loss. Duration, frequency of blinks were then calculated for the First and the Second test. Saccade velocity, fixation durations, length of path travelled and time spent in each area of interest were extracted using OGAMA software. For each of the metrics, the difference between the first and the second scenario was calculated and then tested in a repeated-measure single factor ANOVA with the main factor being the Groups (“Normal”; “FreshAir”; “Hypnosis”).

3.2.3 Method Specificities of the Second Experimentation (HRR2):
This second experimentation aims to supplement results we observed from the first experiment by adding more participants and more workload and induced fatigue (in the first experiment it did not reach moderate fatigue level) in order to reach more significant effects of breaks differences.
Participants:
45 volunteers (15 by groups) from the engineering office in Airbus took part in this experiment (mean age 40 y.o. ±9 SD, 17 females, 6 left handed). All participants gave a written informed consent and reported no physical disability. Participants were balanced in the three groups regarding their training score to have comparable group performance level.

Added apparatus (to previous HRR1 apparatus):
Electrooculography (EOG) was recorded using 6 electrodes (BioSemi) to measure vertical and lateral moves of the eyes in order to extract blinks in addition to Tobii Eye tracker measurements; arousal (blinks, saccades), sympathetic system changes (pupil diameter) and scanning efficiency.

General procedure & scenarios workload:
All participants were delivered a longer training than in HRR1 (30 minutes): briefing and training scenario of 10min with 37 events (16 events in SYSM, 10 COMM, 5 TRCK and 6 RMAN). Based on training performance, participants were assigned and balanced in the 3 groups. Participants performed at different hours of the day (from 8h to 18h) in order to balance the period within the group. To be sure of our eye tracker data comparison the exercise break for HRR2 wasn’t performed with a fresh air part (going out of the building); as it might have an impact on eye wetness that may create a difference of blinking numbers from the two other groups that didn’t move from the room. The other two groups had the same break as in HRR1. In HRR2, MATB scenario events were increased 30% more than in the first experiment with double duration (150 events in 20 minutes: 57 events SYSM, 42 COMM, 31 TRCK and 20 RMAN). Compared to WAM calibration’s [39] highest scenario, our scenario remained between a level 3 scenario and their level 4 highest scenario (judged feasible but unacceptable by pilots). Our scenario could be compared to a degraded situation in operational context (too many tasks in parallel for just one operator). As shown in figure 2, there is no moment of calm and parallel events could vary in modalities’ solicitations from one to six; taking into account that when an event is triggered it takes a few seconds by event to be managed. This induces parallel treatment or task prioritization choices in order to better manage the resources. Such a solicitation might induce some stress as it does not allow strategic management (no anticipation of the amount of events or the type of events) and may consequently increase resources consumption according to Cegarra [8].
Variables and Statistical Analysis

Analyses on quantified data were performed using matlab R2017a. and EOG data were processed through blinker toolbox of EEGLab [40]. Non parametric statistics were computed using statistical toolbox of matlab: Mann-Whitney test for two unpaired samples, Wilcoxon test for two paired samples and Kruskal-Wallis test for three or more unpaired samples, with correction for multiple comparisons when required. Some descriptive statistics are also used to discuss results as the American statistical association underlines the importance of the interpretation of results in context [41]. Finally, those statistical results will be backed up with the subjective debrief conducted with the participants.

3.3 Results:

3.3.1 First Experiment Results: HRR1

HRR1 Fatigue Self-Evaluation:

Descriptive analysis shows: fatigue rates at arrival are all around 4, corresponding to a mild fatigue. Effect of Scenario 1: all groups perceived fatigue after scenario 1. Surprisingly, fatigue was rated higher for the hypnosis break group (group 3) after Scenario1 compared for the two other groups.

Effect of Scenario 2: no group rated the fatigue level as moderate (7). Hypnosis group still felt a decrease of fatigue after scenario 2. Two other groups experienced a fatigue increase: the “fresh air” group who experienced some exercise reported a lower increase compared to the normal group who did not. This reinforces the hypothesis that the hypnosis procedure may influence the overall subjective feeling of fatigue. Nevertheless, the Anova test did not show any significant statistical results for time, (F(2,60)=1.723, p=0.187) and group (F(3,60)=1.076, p=0.366), and for interaction time*group (F(6,60)=1.034, p=0.412). The absence of significant effect may be explained by the low number (n=6) of participants per group.
Fig. 3. Fatigue Self-Evaluation in HRR1

HRR1 Workload Rating Scale:
All participants reported a decrease in general workload after Scenario 2 (delta always negative). Differences between groups is not statistically significant (Anova test F(2,15)=1.460, p=0.263).

Fig. 4. Mean workload (Sc2 - Sc1) calculated from MATB Workload Rating Scales for the 3 groups (“normal”, “fresh air”, “hypnosis”).

Descriptive analysis shows that for all groups, workload at the end of the experiment was rated as lower than before it. As illustrated in the table, the
reference group (G1=“normal break”) is the group in which the break had the less significant effect on workload.

**Table 1. Workload scale items; group ranking depending on difference mean**

<table>
<thead>
<tr>
<th>WRS difference mean &amp; ranking per group</th>
<th>G1: “Normal break”</th>
<th>G2: “Freshair break”</th>
<th>G3: “Hypnosis break”</th>
</tr>
</thead>
<tbody>
<tr>
<td>General score</td>
<td>3 (-6.3±4.2)</td>
<td>2 (-13.7±6.0)</td>
<td>1 (-18.2±4.4)</td>
</tr>
<tr>
<td>Mental workload</td>
<td>3 (-5.0±4.1)</td>
<td>2 (-13.3±9.5)</td>
<td>1 (-17.8±7.0)</td>
</tr>
</tbody>
</table>
| Physical workload in the “Hypnosis” group (G3) is the one who best scored, and especially on “mental workload”, “own performance”, “effort” and “frustration”. “Fresh air” group (G2), who also performed the exercise increased the rating on “physical workload” and “temporal demand” dimensions. Only the participants in the “Hypnosis” group perceived their “own performance” as (much) better in the second scenario, as compared to the first scenario. But Kruskal Wallis test, did not show any statistical significance (H(2,N=18) = 3, 718, p=0.156). Their representation was the closest to real performance; their performance being better for all participants in scenario 2. “Normal” group and “fresh air” group participants estimated their performance as identical or slightly better compared to the first scenario. As emphasised by Cegarra [8], the “effort” dimension is a very good indicator of the difference of activity. As effort is defined as the will for a subject to allocate resources, we can suppose that “hypnosis” group may have the best impact of the resource allocation.

**HRR1 Eye tracking results:**

Descriptive analysis reveals an increase in vigilance/alertness for “hypnosis” group compared to the other groups: “saccade length” and “saccade velocity” are reliable indices of arousal [42]. “Path traveller” and “fixation” are longer for the “Fresh air” group. The number of blinks is higher for “normal” break
group than for the two others. The “Fresh air” group is the one who had the lowest increase of blinks which is consistent with a higher level of arousal.

**Table 2.** Eye tracker data group ranking using scenario differences mean (S2-S1)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixations (second)</td>
<td>3 (-0.04/second)</td>
<td>1 (+0.45/second)</td>
<td>2 (+0.21/second)</td>
</tr>
<tr>
<td>Saccade length (pixel)</td>
<td>3 (-1.2/pixel)</td>
<td>2 (+4.1/pixel)</td>
<td>1 (+8.2/pixel)</td>
</tr>
<tr>
<td>Saccade velocity (pixel)</td>
<td>3 (+0.01/pixel)</td>
<td>2 (+0.1/pixel)</td>
<td>1 (+0.16/pixel)</td>
</tr>
<tr>
<td>Total path traveler (kilo pixel)</td>
<td>3 (+4 kilos pixel)</td>
<td>1 (+50 kilos pixel)</td>
<td>2 (+32 kilos pixel)</td>
</tr>
<tr>
<td>Path traveler velocity (pixel/second)</td>
<td>3 (-14pixels/s)</td>
<td>1 (+93 pixel/s)</td>
<td>2 (+56 pixel/s)</td>
</tr>
<tr>
<td>Computed blinks (percent)</td>
<td>1 (+2.23%)</td>
<td>3 (+0.15%)</td>
<td>2 (+0.89%)</td>
</tr>
</tbody>
</table>

**HRR1 Performances:**
As expected, the pattern of results shows an increase of performance in scenario 2 for all groups compared to scenario 1 and to training. Nevertheless, the Anova test did not confirm this tendency \(F(2,15)=0.1958, p=0.824\). This increase of performance seems slightly better for the “hypnosis” group \(+7.7\% \pm 4.7\) and “fresh air” group \(+6.8\% \pm 1.9\) compared to “normal” group \(+6.5\% \pm 3.5\).

![Fig.5. Means of performance difference by tasks (scenario 2 - scenario 1)](image-url)

Repartition by tasks reveals a difference between groups when comparing
scenario 2 with scenario 1. The “Fresh air” break group increases its performance in all tasks quite equally. The “Normal” break group increases its performance mostly in the detection/reaction task SYSMON (visual processing and manual spatial response). The “hypnosis” break group increases its performance more than the other groups but in three of the four tasks. They put less effort on the RESMAN task, which is the only one that is slowly evolving, but could be demanding in terms of understanding and anticipation.

**HRR1 Free debriefings sum up:**

Most of the participants reported that they were concentrated. Some even forgot the notion of time. Whatever the break proposed, it was generally appreciated. Most of the subjects enjoyed the experiment. They considered the MATB as a game and were surprised at the end of the second scenario due to intense concentration. None of them felt a high fatigue even if the task was judged demanding. The “hypnosis” group made specific remarks concerning the break. Indeed, they were surprised by the fact that they performed the second scenario without any stress. They expressed that they felt physical signs of relaxation: e.g. less tension of the neck muscles compared to the arrival and the first scenario.

**HRR1 results sum up:** Workload is felt as lower at the end of second scenario for all groups. Inferential statistic test did not reveal significant results. This is probably due to the low number of participants. The descriptive analysis shows that fatigue and workload self-assessment seem affected by the type of break, by order of effect: “hypnosis”, “fresh air”, “normal”. Our results sustain other studies showing the benefits of a physical activity [28] [30] on performance and fatigue. Surprisingly, the “hypnosis” group is the only one who rated fatigue, after the experiment, lower than at arrival. On top of that, during debriefings, an unexpected finding emerged: a stress reduction effect. The eye tracker data reveal an increase arousal and alertness for “hypnosis” and “fresh air” groups compared to the “normal” break group. The increase of performance slightly better of the “hypnosis” and the “fresh air” groups than for to the “normal” one, even if all groups globally performed better. The increase in performance for the “fresh air” group is more homogeneous for all tasks. But their performance was lower than the hypnosis’s one, whose subjects seem to have adopted a different strategy such as prioritization of tasks (as the “normal” break group but with a higher global result).
3.3.2 Second Experiment Results: HRR2

**HRR2 Fatigue Self-Evaluation:**
At the end of the second scenario, all participants reported to be more tired than before the test (rest and hypnosis group p<0.005; physical exercise p<0.05).

![Fatigue Self-Evaluation HRR2](image)

*Fig. 5. Fatigue Self-Evaluation HRR2 (asterisks represent significance to a paired Wilcoxon test intra group after correction for multiple comparisons; * p<0.05; **p<0.005).*

The effect of the break is statistically significant, mostly for the “hypnosis” group (Wilcoxon sign rank paired test, p=0.0024, after correction for multiple comparisons, p=0.0072). When comparing “after scenario 1” to “before scenario 2” (after break), the Kruskal & Wallis test appears to be statistically significance for all the groups (H (2,N=45) =14.41558 p=0.0007). When comparing, this time, the groups 2 by 2, only “physical exercise” break group and “hypnosis” break group were significant together and none of them were significant when compared with the “rest” break group.
**HRR2 Workload Rating Scale:**

**Table 3.** Comparison of global workload difference (Scenario2-Scenario1) depending on break groups for the two different scenarios used in HRR1 & HRR2

<table>
<thead>
<tr>
<th>Workload decrease (Scenario2 - Scenario1)</th>
<th>Rest (control) Break Group</th>
<th>Physical exercize Break Group</th>
<th>Hypnosis Break Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRR1</td>
<td>- 6.3 ± 4.2</td>
<td>- 13.7 ± 6</td>
<td>- 18.2 ± 4.4</td>
</tr>
<tr>
<td>HRR2</td>
<td>- 2.73 ± 4.05</td>
<td>- 1.11 ± 3.98</td>
<td>- 3.9 ± 4.05</td>
</tr>
</tbody>
</table>

Workload is felt of a lesser importance at the end of second scenario for all the groups, especially for the “hypnosis” group. The difference between the two experiments is reduced (less benefit of the break) due to an increase of the workload and of the duration of the scenarios. Descriptive analysis shows that general workload self-assessment was impacted by “hypnosis” break (table 3). Surprisingly, the break with “exercise” was less efficient this time. Even the normal break scored better.

**Table 4.** HRR2 Workload scale (difference mean and group ranking) for the various measures on the groups

<table>
<thead>
<tr>
<th>WRS measures mean / ranking per group</th>
<th>G1: “Rest”</th>
<th>G2: “Exercice”</th>
<th>G3: “Hypnosis”</th>
</tr>
</thead>
<tbody>
<tr>
<td>General score</td>
<td>2 (-2.7±4.5)</td>
<td>3 (-1.1±4)</td>
<td>1 (-3.9±4.1)</td>
</tr>
<tr>
<td>Mental workload</td>
<td>3 (-2.1±3.4)</td>
<td>2 (-2.9±3.1)</td>
<td>1 (-5.4±2.6)</td>
</tr>
<tr>
<td>Physical workload</td>
<td>2 (+6.4±4.2)</td>
<td>1 (+6.3±3.8)</td>
<td>3 (+8.3±4.6)</td>
</tr>
<tr>
<td>Temporal demand</td>
<td>3 (-1.9±3.5)</td>
<td>2 (-3.6±3.3)</td>
<td>1 (-7.3±4.7)</td>
</tr>
<tr>
<td>Own performance</td>
<td>3 (-5.5±4.6)</td>
<td>1 (+1.3±4.7)</td>
<td>2 (-4.3±4)</td>
</tr>
<tr>
<td>Effort dimension</td>
<td>3 (-2.3±3.5)</td>
<td>1 (-7±5.2)</td>
<td>2 (-5.1±3.8)</td>
</tr>
<tr>
<td>Frustration</td>
<td>1 (-10.9±5.2)</td>
<td>3 (-0.7±3.9)</td>
<td>2 (-9.6±4.5)</td>
</tr>
</tbody>
</table>

When looking in details at the workload scale dimensions, “hypnosis” break group better scored the “mental workload” and the “temporal demands” (see table 4). On the contrary, the “physical workload” was rated higher by hypnosis break group than by the two other groups. “Exercise” break group indicated lower levels of workload compared to the “rest” group for “mental”,
“physical”, “temporal” and “effort”. Surprisingly, this group, who rated the workload as higher in HRR1, scored here a lower workload, thanks to the “frustration” dimension. Nevertheless, the “exercise” group is better ranked (group ranking) than the others for “physical”, “effort” and “performance” dimensions. “Exercise” group is the only one to have a positive representation of its “own performance” as they were closer to reality (all participants performed better in scenario 2). As emphasised by Cegarra [8], the “effort” dimension is a very good indicator for the difference of activity. Rating confirms here that “exercise” break group has the best impact on resource allocation.

**HRR2 Eye tracking & Electrooculography results:**

*Blink* in the literature: the number of blinks and their duration decrease when cognitive workload increases [43] [44]. Those measures were retained for comparing scenarios, before and after break (Figure 6). The type of break did not revealed any statistical significant effect on blink duration in any of the three groups. Nevertheless, we can notice a tendency of blink duration increase for “hypnosis” break group, especially just after the break (Figure 6). But, this tendency decreases as the scenario is running. The difference of the number of blinks (S2-S1) is very close in the 3 groups, whatever the scenario, and even if, the total amount of blinks is greater for the “rest” group: +20 (+6.5 “hypnosis” and +3 “physical exercise”).

![Fig. 6. Mean of blink duration for the three groups (Error bars = standard error of the mean).](image-url)

*Pupil diameter*: this parameter increases with mental workload. To analyse it along the two scenarios, we averaged it over bins of ten seconds. During the first scenario, the pupil diameter tended to decrease. This could indicate a reduction of workload due to task learning. On the contrary, during the second scenario, the pupil diameter remains globally stable. This could indicate that the performance learning plateau could have been reached. This
could explain that the workload is less decreasing in the second part of the second scenario.

Saccades: their length and velocity are reliable indices of arousal [45][46]. In our experiment, we considered as saccade an eye-shift of minimum 50 pixels (≈ 2° of visual angle) in less than 50ms. We plotted the difference in saccade length and saccade velocity between the two scenarios, for each group. No statistical significant difference was found between groups. However, participants who did physical exercise during the break made significantly longer saccades after the break (Wilcoxon distribution sign rank test, p=0.0054). There is also a tendency for faster saccades (mean difference of 275px/s, p=0.07). We can reasonably assume that they had a higher arousal level after exercising.

Fixations: we defined fixations as eye-shifts an amplitude of lower than 50 pixels and with a duration of at least 80ms. Whatever the participants, no difference was noticed between the two scenarios.

Time spent in each area of interest (AOI): the time spent looking at each area was used to determine whether or not the groups adopted different strategies. Globally, all groups exhibited the same visual pattern. All participants adjusted their behaviour the same way: they spent less time looking at the “COMM” task area and more time looking at the “SYSMON” scale subtask area.

HRR2 Performances:

The best performer was the “rest” group (+7.64%). This is quite surprising compared to the first experiment where he was the worst one. Another surprising result is the “exercise” (+5.95%) group as the worst performer with the lowest percentage of increase between the two scenarios. Indeed, this group was a good performer during the first experiment. The “hypnosis” group (+7.24%) is the second one. There was quite no difference in managing the four tasks, as already described in the AOI section.

HRR2 Free debriefings sum up:

All groups reported a significant effort of concentration while performing. At the beginning, they were surprised by the density of events in the 4 tasks (lot of simultaneous interactions that required adaptation). They all reported eye and body relaxation thanks to the breaks. They also reported they found strategies for the second scenario, due to the fact that they considered it as a challenge. All subjects felt the fatigue after each scenario.

G1 “rest”: scenario 1 and scenario 2 differences

Scenario 2 was deemed easier, as break allows them to step back, to think about their own performance and also to adjust their strategy for
accomplishing the task. They expressed for the second scenario that they were already warmed up and vigilant.

G2 “exercise”: scenario 1 and scenario 2 differences
The subjects considered the scenario 2 easier as they adjusted their strategy. They felt more engaged and quicker with a higher attention. Scenario 2 was felt shorter in time for the participants who adapted themselves but long and boring for others. The break was useful but did not change anything except eye relaxation.

G3 “hypnosis”: scenario 1 and scenario 2 differences
The “hypnosis” group expressed a gain in motivation, a step back feeling, less time pressure for performing the tasks, less stress (calm and serenity). They felt that they performed the second scenario with a better fluency and self-mastery, with self confidence in their own performance and consciousness to perform better. Some subjects used the positive emotions they experienced during hypnosis to regulate their own frustration when missing solicitations.

HRR2 results sum up:

Fatigue increased after scenario 2 for all groups and all groups felt more tired at the end of the test. The effect of the break reached the statistical threshold of significance, mostly for “hypnosis” Group. When comparing, this time, the groups 2 by 2, only “physical exercise” break group and “hypnosis” break group were significant together and none of them were significant when compared with the “rest” break group. Workload is considered as lower at the end of second scenario by all the groups. Nevertheless breaks benefit scored less in HRR2 than in HRR1. General Workload self-assessment scored “hypnosis” break group as the one having a better benefit for workload decrease. Surprisingly, the two other groups scored the opposite way in HRR2 and in HRR 1. Eye tracker data reveal that the difference (S2-S1) of the numbers of blinks and their fixations are quite similar in the 3 groups. Saccades were significantly longer in the “exercise” break group after the break (Wilcoxon distribution sign rank test, p=0.0054). They also had a tendency to be faster (mean difference of 275px/s, p=0.07). We can reasonably assume that the participants had a higher arousal level after the exercise. Globally, all the groups exhibited the same visual pattern. All adjusted their behaviour the same way, spending less time looking at the “COMM” area and more time looking at the “SYSMON” sub-scale area. Results show that during the first scenario, the pupil diameter tended to decrease. This could reveal a workload reduction due to task learning. During the second experiment, the pupil diameter remains globally stable: performance learning plateau could have been reached. Global performance tendencies are in favour of the “resting”
group (+7.64%) closely followed by the “hypnosis” group (+7.24%). Surprisingly, the “exercise” group is the one (+5.95%) whose performance was the worst. All groups reported a significant effort of concentration and surprise at the beginning of the first scenario. Eye and body relaxation during the breaks were reported, with less fatigue. They also reported they found strategies for the second scenario, often considered as a challenge. All subjects felt the fatigue after each scenario. The “Rest” break and “exercise” groups made the same comments when comparing scenario 1 and scenario 2: the break allows eyes relaxation and a “warmed up attention”. The only specific comment concerned the step back effect of the “rest” break. The “hypnosis” group produced different feedbacks: breaks bring gain in motivation, less time pressure (action peacefully done), less stress, better fluency, self-mastery and self-confidence in their own performance.

4 Discussion

This work aims to explore this question: does hypnosis help manage fatigue when workload is increased? To answer this question we compared three type of break in two different workload situations. As today workplaces (cockpit, other workplaces) is increasing multi-solicitation and divided attention, we found adequate to use the multitask battery (MATB) to challenge this question. Aeronautical operational correspondence of experimented scenarios could be in HRR1 (middle workload level), in approach vectoring when not using full automatisms. Regarding HRR2 scenario, we consider that it was closer to abnormal situation in a cockpit (too-many tasks in parallel for just one human operator) and probably not reachable in standard workplace. Those scenarios do not pretend to test only the standard cockpit workload, but was our way to challenge break techniques. Our intention while increasing workload (as it impacted fatigue), was to check if we increase break effects. “normal/rest” break (G1) aim in both experiments to be our referent situation in order to compare two others breaks. We assumed a break effect on subject activity in second scenario of both experiments, which was confirmed by results. We assume a better effect with break using exercise as already emphasized [46][47]. We assume a better effect with break using hypnosis than the reference break. The statistics performed on results were not always significant, probably due to the low number of participants especially in HRR1. HRR2, had a higher number of participants (*2,4). For this last experiment the significance was reached for fatigue comparison between “exercise” and “hypnosis”, plus saccade measurement in favour of “exercise” group arousal. By another way American Statistical Association “emphasizes
principles of good study design and conduct, a variety of numerical and graphical summaries of data, understanding of the phenomenon under study, interpretation of results in context, complete reporting and proper logical and quantitative understanding of what data summaries mean’. As they also sustain that a p value does not measure the size of an effect or the importance of a result [41], we add descriptive analysis and sum up of those two pioneer studies.

Results redundancies in both studies:
• The increase of subjects’ performance and the variation of the effect of break.
• The fatigue results of the “hypnosis” group scoring before and after break while the “rest/normal” group (reference) increase.
• The workload perception at the end of experiment with a beneficial order of effect: “hypnosis”, “fresh air/exercise”, “rest/normal” breaks.
• The higher score of blinks for “rest/normal” group according to our assumption.
• The unexpected result of “hypnosis” group on stress decrease in debriefings.

In order to increase performance, as instructions requested, they all found strategies to adapt before suffering important global performance degradation. This could be explained by “economic strategy” adaptation [49]. In HRR1 we suspect subjects to build cost/benefit model during the experiment. For instance, some subjects favoured some tasks to maintain a good performance level and neglecting some others.

Differences in results due to workload and duration variations:
• General: HRR1 descriptive results were more in line with our assumption: in favour of “hypnosis” and “exercise” groups (for workload, fatigue, arousal and performance) regarding to “normal/rest” reference group (G1). HRR2 with more workload and induced fatigue, draw less evidence to decide between “rest” and “exercise” but sustain the “hypnosis” group as good.
• Performance and workload global scores: as in HRR1 “Hypnosis” and “Exercise” groups were the best, surprisingly, in HRR2, the order of effect changed; in benefit of “rest” group, closely followed by “hypnosis”, then “exercise” as the last for performance. And in HRR2 workload scoring “hypnosis” group then “rest” group and the last score for “exercise” group. HRR2 shows smaller decrease of subjective workload than in HRR1.
• Regarding HRR1, if we look at groups’ performance more in details, all groups slightly increased but the fresh air group (exercise) was the most harmonized by increasing quite equally in the 4 tasks (but no more in the
second experiment). This result might support previous works [47] [48] in middle workload scenario but no more in high workload conditions (HRR2). In HRR1 “Normal/rest” group was the last but not very far from others, increasing mostly on system monitoring task but relinquish other 3 tasks, possibly due to resource estimation. «Hypnosis» break group showed higher increase in 3 tasks but not in one (RESM). This is consistent with an “economic strategy” [49].

- Fatigue scale rating shows that the “fresh air/exercise” break brings less benefit in HRR2 than in HRR1. Concerning “Hypnosis” break effect on fatigue it brings benefit in both experiments but didn’t show persistent benefit while coping with HRR2 second scenario as it did in HRR1: subject rating less fatigue after the experiment then before or during experiment.

- Eye tracking data were supporting the results in HRR1 in favour of vigilance effect for “fresh air/exercise” break group and “hypnosis” break group. But in HRR2 data in hypnosis break group seems more related to “drowsiness” than on other groups in the first part of the second scenario. This might be due to hypnosis break immediate post effect that is particularly pronounced with some subjects just after exposure to hypnotic state. This latency effect is usually reducing when experimenting more often hypnotic states. In HRR2 arousal was most in favour of “exercise” group and so not consistent with their performance and workload low results (lower than the 2 other groups).

Regarding the relationship performance/cost, Amalberti emphasizes own “subjective satisfaction” to keep situation mastery [49]. He also underlines that “this sufficiency is compatible with a high performance and an associated high cognitive cost.” This cost is confirmed by the workload and the fatigue subjective scales scoring in both studies. The general workload mean for HRR1 1 and 2 was respectively: 55% and 42% but turns around 65% in both HRR2 scenarios. However the performance resources correlation is not linear. Having more resources available will not imply having a better performance. As human have limited mental processing and may be “satisfied” with the performance they accomplished according to the effort they invest. So this observable alone is not the best to support our assumption and must be correlated with other dependant variables (workload and fatigue ratings, eye tracking data). Regarding fatigue results evolution with higher workload, it seems that when being too much solicited, fifteen minutes break pained more to compensate 40 minutes of intense multitasking (HRR2) whatever the break is. Fatigue as workload results are globally consistent. Nevertheless, the scales rating increased only in “hypnosis” group only after the first scenario in both studies is interesting. Subjects main difference from other subjects was
that they were select by their suggestibility (hypnosis responsiveness). We can only make some assumptions about this observation: first assumption could be that this “suggestable population” may have a more sensitive consciousness of their own. But as we tested only hypnosis volunteers, this assumption is questionable as some suggestable people that were not volunteers are probably included in other groups. Second assumption could be that this is due to the first exposition to hypnosis state (during selection) that already changed subjects own fatigue perception. Those assumptions have challenged in further works. By another way, studies about multitasking and the effect on attention are growing as our today world is more and more stimulated by multiple media formats. Addiction to this multitasking is emphasized by physiologists [50]. They show same brain area activation while having behavioural compulsions. They emphasized that reward network is involved. They also discussed the sensation seeking that occurs with multitasking. It seems that our subjects were here challenged by multitasks especially in the lower workload experiment; finding the experiment as challenging as a game. As stress is a factor of workload perception variation [51] and one of our unexpected feedbacks was the “hypnosis” stress less effect (both studies) reported by subjects. As stress is defined as a result from mismatches between a perception of task demands and the perception of resources needed to meet them [51], it might be increased in multitasking context. Moreover, neurophysiologist emphasize this bad effect of multitasking on our brain functioning, as this divide attention and we are not “designed” for that [52][53]. Those multi-stimulations increase the stress hormone cortisol as adrenaline creating mental fog or scrambled thinking. This also creates a dopamine addiction feedback loop, as the brain is using a rewarding hormone system while experiencing stimulation change. This could drive people to exhaustion like Guangzhou man or Daegu man who played games until exhaustion. Some questions could be envisaged for further studies, when thinking to those people: do they have any consciousness of their body fatigue state to be able to go to such extreme behaviour? Does cortisol help to hide body consciousness feedback? In debriefings, calm and stress less effect was naturally express by “hypnosis” participants. To understand this, we searched studies that mentioned this “calm and stress less effect” with medical rationales [54] [55]. Explanation could be linked to the fact that hypnotic state helps to synthetize more oxytocin hormone [56]. This hormone produced while being under hypnotic state has an effect on stress management pattern (it may also explain the “drowsiness” after break in second experiment). In another hand, during the hypnosis session we also suggest some positive emotions. Experiment shows that with hypnotic session combined with happiness versus anger suggestions there is a decrease of plasma cortisol
related with happiness and not anger suggestions [57]. This might be a solution to compensate the induced stress by multiple tasking and workload. We sustain that positive emotions might increase motivation or effort dimensions and help subjects to hire resources all along the activity. Hypnosis might then help being more present to our own body perception and consecutively may help own resources regulation. Descriptive analysis of results shows “hypnosis” break as good as or even better as “exercise” break results but without any move. As “exercise” break is difficult in a flying aircraft such “auto-hypnosis” break might be useful (before a more intense workload flight phase for instance). It could help to cope with fatigue and stress during long flights or multiple rotations. Other type of hypnotic scripts might be tested also in that sense. It would be of interest to test such a break effect with a lower workload scenarios and longer duration to check if it would be of any help while coping with workload more in accordance with daily flight.

5 Conclusion and perspectives

Perspective of our two pioneer experiments could be to replicate with a higher number of participants. It could worth to continue investigation about the effect of hypnosis technique on fatigue and stress. It might be of interest to add physiological measures: cardiac [38], FNIRS plus hormones to monitor sympathetic system. Kahneman proposed [10] to monitor sympathetic system to observe energy dimensions of workload and resources regulations. About pilot community, investigate other workload flight phase (like monitoring activity) and long period of time, in order to see if it helps to cope with drowsiness / boredom effects (with and without incident). For pilot and employees, hypnosis could be teaches to be practiced alone: auto-hypnosis (as Piccard’s practice). Nevertheless, hypnosis as already been recognized as very efficient in a lot of medical practice (anaesthesia, psychology...) must be practiced by qualified people trained by training centre accredited by “Fédération Française d’Hypnose et de Therapies Breves”, “European Society of Hypnosis & Milton H. Erickson Foundation” or “Milton H. Foundation Phoenix USA”. Practicing with a qualified trainer allows to adapt the hypnosis practice to the subject own needs and characteristics. This practice could also be used in small group for well-being and prevention of psycho social risk as soon as the practitioner adapts the session to people characteristics. As it brings less stress and better resource management it could also be used in education situations (it is already used for exam preparation).
References

1. Solar impulse, https://www.youtube.com/watch?v=iJRJwoz-yPA
2. Solar impulse, https://www.youtube.com/watch?v=xZO35uGAXZM


42. VanOrden, K. E., Limbert, K., Makeig, K., Jung, T. P.: Eye Activity Correlates of Workload during a Visuospatial Memory Task. Human Factors. 43(1), 111–21 (2001)


On the Earth as Well as in the Sky
Facing the Risks Between Workload and Resilience:
a Psychoanalytic Point of View

Maria Gabriella Pediconi\(^1\) and Glauco Maria Genga\(^2\)
\(^1\) Research Professor of Dynamic Psychology of Urbino University (Italy)
and Psychoanalyst of Società Amici del Pensiero ‘S. Freud’ of Milan
maria.pediconi@uniurb.it
\(^2\) Psychiatrist and Psychoanalyst of Società Amici del Pensiero ‘S. Freud’ of Milan;
Consultant of Istituto di Medicina Aerospaziale of Milan (Italy)
glaucomaria.genga@fastwebnet.it

Abstract. After introducing the notion of the resilient Ego as an essential part of the Human Mental Workload study (Genga Pediconi, Dublin 2017), we have proceeded to analyse two samples, composed respectively of firefighters and military pilots. We have explored subjective perceptions related to the profession in their relationships with: 1) some structural aspects; 2) experiences related to the risky events encountered during specific tasks (rescue operations, flight missions). On the one hand, throughout the whole sample, the risk perception increases together with the perception of the workload and the sense of belonging to a specialized body. On the other hand, risk is perceived as linked to family ties and to the feeling of one’s own safety. The paper will present the main results of the study, discussing how both professions show realism and resilience, witnessed by the close link between the sense of one’s own safety, family affections and the sense of service that animates the subjective perception of rescue operations and flight performance.

Keywords: Risk Perception, Mental Workload, Resilience, Psychoanalysis.

1 Introduction. A Call for Life as an Example of Human Mental Workload

“The other man acquired the value for him of a fellow-worker, with whom it was useful to live together”. 
(Freud, 1929)
Last year we participated in the 1st International Symposium on Human Mental Workload (Dublin), presenting a paper about the Resilient Ego in emergency situations. We aimed to offer the contribution that can be offered by psychoanalysis, in the study of the error as well as of the Human Factor, delineating the Ego as a resource. For this reason, we have chosen these two sentences as our esergy. The first is taken from Civilization and its Discontents: Freud emphasizes how it is fundamental for the individual to know how to judge the work of others as far as he is concerned. In other words, any enterprise is possible if conducted in partnership with a partner. «The other man has acquired the value for him of a fellow-worker, with whom it was useful to live together». (Freud, 1929, p. 99) The same underlining was revived in a completely different context by Captain “Sully” Sullenberger, who reminded the whole world – with his acts and his words – how precious IS an individual who thinks and behaves in a correct and responsible way. This is even more true in a highly technologically advanced context such as FLYING.

In fact, as Sullenberger writes, «Technology cannot replace experience, skill and judgement.» (Sullenberger&Zaslow, 2010, p. 190) Last year we briefly introduced a pilot study we were conducting on the mental health of Italian firefighters. In the present paper we will report the first results of our research, which has examined two different samples, composed respectively of firefighters and pilots of the Italian Air Force. Before entering into the details of this research, we would like to draw attention to a recent impressive story, which was briefly addressed by the international press. On April 17th, 2018, 11am, Southwest Airlines Flight 1380 (NY La Guardia - Dallas) was flying over Pennsylvania at an altitude of 33,000 feet, when suddenly the passengers heard an explosion on the left-hand side of the aircraft, which had passed a routine check only two days before. The left engine of the plane, a Boeing 737, exploded. The pilot was Tammie Jo Shults, one of the first US Navy woman pilots. The phone call in which she, in a very calm voice, communicates the engine fire to the control tower, has been published by all the American newspapers: «They say that a hole has opened, and someone has flown out». Shortly after the explosion, in fact, a turbine blade broke off and hit a window, breaking it. Twenty minutes after the explosion, the plane smoothly landed on the Philadelphia airport runway. The control tower told Capt. Shults to stop where she could: «Thanks, we stop here near the fire trucks. Thanks for the help, guys», she replied. A round of applause accompanied Shults as she walked down the corridor to ascertain their condition before they disembarked. She received praise from all over the world. Capt. Shults, after a degree in biology, graduated in the Navy, where she entered in 1985.
because she had not been accepted by the Air Force. She became lieutenant commander and instructor of the F/A-18 Hornet fighter pilots, but in those years in the US Armed Forces it was still forbidden for women to take part in combat operations. So, she left the service in 1993, and a year later became a Southwest Airlines pilot, where she met and married a colleague. An Italian journalist (Sarcina, 2018), commenting on this news listed what were, in his judgment, the decisive factors in the dramatic moment of the explosion of the engine: the pilot’s training, her reflexes and capability. What in common sense was once called fortitude, or ‘cold blood’, is something that today has become the object of scientific studies in various fields: it is called ‘resilience’. Her behavior is clearer when more light is shed by the psychoanalytical point of view, in which we take into account an episode of her college years: «During my senior year of high school in 1979, I attended a vocational day where I heard a retired colonel give a lecture on aviation. He started the class by asking me, the only girl in attendance, if I was lost. I mustered up the courage to assure him I was not and that I was interested in flying. He allowed me to stay but assured me there were no professional women pilots. I did not say another word. In my heart, I hoped that God had given me an interest in flying for a reason. I had never touched an airplane, but I knew flying was my future. During my junior year in college, I met a girl who had just received her Air Force wings. My heart jumped. Girls did fly! I set to work trying to break into the club.»

To tell the truth, resilience can and should be exercised in different contexts:

- to face natural events, i.e. independent of the will of someone, such as the explosion of an engine in flight or the impact with a flock of wild geese as it was for Sully Sullenberger in 2009;
- to face the behavior of other individuals, who can be injurious or offensive, as in the episode of the high school student Tammie. It is evident that these are totally different stresses and «shock waves»: in the second case, the stressor is prejudice.

The calm response of Tammie Jo Shults recalls the well-known and so-called Miracle on the Hudson (2009) when an engine failure forced an aircraft to make an emergency landing on the Hudson River. Capt. Sullenberger, the pilot, has always attributed great importance to the diligent and incessant study of prior cases: accidents, disasters, terrorist attacks, etc. (Sullenberger & Zaslow, 2010) He has repeatedly stated that he is not a hero, as he did what any experienced pilot would have done in his place. This is the point that we would like to highlight: the resilient Ego never works alone, but as part of a team. Individual responsibility is an unavoidable and precious element, just as individual

---

thinking is a «resource in emergency situations». (Genga-Pediconi, 2017) Vanni (2016) recalls the importance of individual responsibility: the assessment of the risk of a behavior, in order to be correct, must always be done on a specific subject and specific context. «Sully seems to be a good example of this too. He knew very well that what he was about to do would make him face the risk of being considered imprudent and impulsive, but we like to think that he preferred to do what was best, facing the risk of responsibility.» (Vanni 2016, p. 129) An example of impeccable and quick decision making. In these two episodes we find the individual ability to manage the HMW in a positive way, and precisely an occasion to make the best choice, which allowed the rescue of hundreds of people. «Mental workload (MWL) is a fundamental design concept in Human-Computer Interaction (HCI) and Ergonomics (Human Factors) and it is sometimes referred to as Cognitive Load (CL), specifically in Cognitive Psychology. It is intrinsically complex and multifaceted. There is no widely accepted definition of MWL, however, it can be intuitively described as the total cognitive load needed to accomplish a specific task under a finite period of time.» (Moustafa K., Luz S., Longo L., 2017, p. 31). In our work «workload can be characterized as a mental construct that reflects the mental strain resulting from performing a task under specific environmental and operational conditions, coupled with the capability of the operator to respond to those demands.» (Cain, 2007, P. 4-3) But there is more. As Hancock writes, «Brains are more difficult to understand than muscles». (Hancock, 2017, p. 15). What about this phrase so concise and suggestive? We would like to take the bull by the horns and, in a certain sense, re-launch the challenge, bringing thought into play, in addition to the muscles and the brain. In this direction, the contribution of psychoanalysis could be useful to specify the subjective components of mental workload. «However, our modern world tends much more to be a cognitive enterprise (...) Most often the computer makes cognitive gratification all the more likely. What we have not done in workload assessment is to sufficiently value, nor sufficiently evaluate this hedonic dimension of the workload response. In the same way that we can ask whether beauty is a contributor to the optimization of design, so we can also ask whether satisfaction is a governor of perceived workload?» (Hancock, 2007, pp 13-14) The question of personal satisfaction is always present and inevitable, in every kind of performance. Most researchers agree that, although mental workload is a concept that has intuitive meaning, it is difficult to define. Wickens (2017) suggests that the concept of effort or mental workload can be examined from three perspectives, those of measurement, of prediction and of consequences, even in a subjective way. We have to consider, indeed, that «sometimes we do gain intrinsic pleasure and value by investing more effort: the feeling of «flow» in working hard at an
engaging interesting task». (Wickens, 2017, p. 26) In the same direction we find the dynamic description of what happens during the training of pilots in order to obtain the necessary know-how of their delicate job: the «concept of expertise links closely to the concept of the four stages of competence, with novices moving from Unconscious incompetence, conscious incompetence, conscious competence and finally unconscious competence.» (Byrne, 2017, p. 193). We are including in the necessary know-how of demanding jobs such as military pilots and firefighters also resilience as «the process of coping with disruptive, stressful, or challenging life events in a way that provides the individual with additional protective and coping skills than prior to the disruption that results from the event.» (Richardson et al., 1990, p. 34) This paper aims to show new connections among Human Workload, risk perception and Resilience. They have a common denominator, at least in part: the role of subjective contributions even in demanding situations, in which person-systems and technical-systems could be parallel. It is useful to mention Reason’s distinction between the person-approach and the system-approach in managing risky situations. «Further improvements in reliability will require more effective methods of risk management. These, in turn, depend upon acquiring a better understanding of the breakdown of the complex socio-technical system and the development of new techniques of risk assessment.» (Reason, 1990, p.28) When demands exceed capacity, skilled operators modify their strategy to compensate, in order to avoid the degradation of their performance. Demanding situations can be analysed in order to measure the adaptive competence of operators, also in terms of resilience.

2 Dangerous Professions: Pilots’ and Firefighters’ Jobs in Terms of Risk

«Risk is ubiquitous. There is no human state or action that is without risk, short of death itself, although there are clearly some states and actions that carry substantially more risk than others.» (Hunter, 2002, p. 1). According to Hunter’s conclusions that «Poor pilot decision-making has been implicated as a leading factor in fatal general aviation accidents» (Hunter, 2002, 1):

- We assume the definition of risk according to common sense: «risk is the possibility of loss of life or injury, and it encompasses both the probability of facing an encounter with a hazard and the severity of a hazard» (Hunter, 2002, p. 1).
- We investigate the relationship between risk perception and risk tolerance, and conclude that dangerous behaviors do not derive so much from risk tolerance, but rather from the erroneous perception of it: «The results of
the present study suggest that it is risk misperception, not high-risk tolerance, that is associated with exposure to hazardous aviation events.» (Hunter, 2002, p. 21)

The sample of our research was composed of a team of firefighters and a team of military pilots. They are people typically engaged in delicate operations in which individual skills and teamwork coordination are decisive in order to obtain maximum efficiency.

Here is a brief description of firefighters’ jobs. They are primarily responsible for responding to fires, accidents and other incidents where risks are posed to life and property. The main duties of a firefighter are to help protect the public in emergency situations. They respond to a wide variety of calls, such as car crashes, chemical spills, flooding, water rescue and general rescue as well as fires. Many fire crews are trained as first responders. The work can be stressful and dangerous but there is a great deal of job satisfaction to be gained from providing such a valuable service to the community. In terms of personal qualities, firefighters should be confident, resilient, flexible, strong team members, able to think quickly under pressure and able to communicate effectively.

Firefighters are exposed to some kinds of risks.

The first kind is physical dangers. One of the primary risks of being a firefighter is that of personal physical injury. Firefighters are susceptible to burns, smoke inhalation and crush injuries from collapsing structures. Due to the highly physical nature of the job, all matter of bodily harm can befall firefighters in the line of duty. They can suffer also from long-term job-related illnesses and some physical fallout from the profession may not materialize until years down the road.

The second kind of risk is related to mental anguish. In addition to the physical perils, firefighters face the potential for mental trauma, particularly in situations involving mass casualties. As first responders on the scene, firefighters and other rescue personnel are often faced with the high stress, high stakes environment of trying to attend to as many people as possible, remove them from danger, often performing their duties in tenuous or unfolding situations.

The third kind of risk concerns stress-related diseases. The elevated stress levels don’t always dissipate when the firefighters leave the rescue scene. Much like soldiers engaged in war zone conflicts, firefighters are susceptible to developing post-traumatic stress disorder (PTSD). This pathological condition occurs when a person in a rescue or service role witnesses a traumatic event that leaves an indelible mark on the psyche. It can lead to flashbacks, poor sleep habits, anxiety
and depression. Sufferers of stress-related diseases may become withdrawn and despondent or angry, hostile and destructive.

A fourth kind of risk comes from the irregular balance between work and life. Firefighters work long hours within unstructured schedules, making it a challenge to create a viable work-life balance. Conventional family life may be skewed, which can lead to familial discontent and disconnection. The need to work long hours is an especially critical factor in firefighters who are regionally deployed to combat wildfires or to assist in large natural disaster zones. The long hours further contribute to the overall stress of the position. Even if firefighters are trained and educated about the potential physical and mental stress which can arise from the job, however not all realize the full impact until they are on the job, and some not until years afterwards.

With regard to firefighters, they are well aware that their rescue work involves dangers – they know how to foresee, recognize and face them together in teamwork. We would like to point out that the kind of dangers they face is more concerned with safety (i.e. the defence of the environment, assets, people or communities) than natural events of any kind and less with security (i.e. defence against attacks directly and maliciously caused by other individuals).

As for the risks they face in their activities, the literature on this subject is VAST. Among the most recent works, we will mention Setti & Argentero (2015), who take into account different types of stress, from the adverse microclimate to the high socio-environmental conditions. «Personnel who provide assistance in emergency situations, such as ambulance operators or firefighters, as well as rescue professionals in general, are exposed to a high risk of developing occupational stress due to frequent contact with critical and potentially traumatic events.» (Setti & Argentero, 2015, p. 101)

They also mention the «so-called ‘John Wayne syndrome’, characterized by the tendency to hide their feelings and their emotions in dealing with situations perceived as psychologically difficult » (Setti & Argentero, 2015, p. 101), as well as the vicarious trauma, outcome of the stress experienced during the support provided to suffering or traumatized people. According to Blaney & Brunsden, «the FRS is a high-risk profession yet firefighters have strategies for coping with the stress.» (Blaney & Brunsden, 2015, p. 26) Here is a brief description of military pilots’ jobs. Members of the military service train for and perform a variety of tasks in order to maintain national defence. Service members work in occupations specific to the military, such as fighter pilots or infantrymen. Pilots have to participate in, or support, combat and other military operations, as well as humanitarian or disaster relief. Specific tasks can be: operate, maintain, and repair equipment; perform technical and support activities; supervise junior enlisted personnel. Transportation officers
manage and perform activities related to the safe transport of military personnel and equipment by air and sea. Pilots operate and command an aircraft in the military, fly various types of military airplanes and helicopters to carry troops and equipment.

The specific work environments and conditions for military occupations depend on occupational specialty, unit, branch of service, and other factors. Most active-duty military personnel live and work on or near military bases and facilities throughout the world.

Military members must be physically fit, mentally stable, and ready to participate in or support combat missions that may be difficult and dangerous and involve long periods of time away from family; however, some personnel are rarely deployed near combat areas. Members of the military are often placed in dangerous situations with the risk of serious injury or death. Members deployed to combat zones or those who work in dangerous areas, such as the flight deck of an aircraft carrier, face a higher rate of injury and death. In many circumstances, military personnel work standard full time. However, hours vary significantly, depending on occupational specialty, rank, branch of service, and the needs of the military. In all cases, personnel must be prepared to work long hours to fulfill missions. (Hunter, 2002, 2006; Sicard, 2003).

All jobs come with health risks, but some are very serious and even potentially fatal. The U.S. Bureau of Labor Statistics (2016) compiled data from such hazardous jobs and created a list of The Top Ten Most Dangerous Occupations. Being an aircraft pilot is the third most dangerous occupation. Bush flying, crop dusting, banner towing and piloting commercial flights are all included in this group. Crop dusting and bush flying are by far the most dangerous due to the fact that they fly in small planes, very close to the ground, and often work long hours. Also, as with all pilots, if something goes wrong there’s little option other than to come crashing down.

If we now consider military pilots, we must contextualize their relationship with risk in a different way compared to other dangerous jobs. Indeed, military pilots have to deal with risk as part of their work: «in military aviation, mission success involves a higher level of risk due to a more dangerous and unpredictable environment» (Sicard et al., 2003, p. 879), unlike civilian pilots who have lower levels of risk tolerance. We would like to recall that the whole history of flight, from the first flight of the Wright Brothers (1903), has always seen civil aviation follow and profit from much theoretical and technological progress produced in military aviation, in different countries. Not only that: until the Second World War, a singular feeling united military pilots, to the point that the pilots of the Nation of the
opposing side were called *adversaries*, and not *enemies* (Tiberi 2011). «It has been shown that those who do not have adequate air power or cannot make it superior or equal to that of the adversary cannot in fact aspire to victory.» (Licheri 1997, p. 246, our emphasis). A rather sporty language, that we would not expect from military sources.

«Military pilots are reported to have significantly more self-control, to seek higher danger levels, and to feel more energetic and more invincible than their commercial counterparts, whereas impulsiveness scored either by EVAR or Barrat scales is very similar in both groups.» (Sicard et al., 2003, p. 881).

The greatest dangers that pilots incur can be traced back to very different sources: from the sudden change in weather conditions, to the malfunctioning of some of the aircraft systems, or even to faults in the coordination between crew members (cross-coordination can fail if everyone 'thinks that the other will take care'), or finally the attitude of awe that the younger pilot can feel towards a colleague higher in hierarchy. It is very important that these situations are foreseen and included in the training of pilots, as in fact already occurs.

As for the perception of occupational risks in civilian pilots, we cite the very detailed work of Moses&Supply (1989), whose merit is to have placed the satisfaction or dissatisfaction of commercial pilots at the center of their investigation, and their perception of the effects of deregulation on flight safety: «Dissatisfaction with training and maintenance are significantly related to the financial health of the employing airline». (Moses&Savage, 1989, p. 2)

### 3 Resilience as a Psychological Resource

In examining the psychological resources proven to be most effective in high-risk professions, a growing place must be attributed to resilience, whose importance is increasingly recognized in various fields. In sociology and psychology, as well as in systemic psychiatry, this notion began to spread in the 1980s.

It has been observed that there is no univocal definition of the concept (Bhamra et al., 2007; Lisnyak, 2015; Weidlich, 2015), just as, according to Rutter (2006), it is prudent to assume that resilience does not indicate a precise and limited concept: either because it is purely descriptive (avoiding negative behaviors and adopting the positive ones in terms of survival and success), or because the term is adopted by different disciplines and therefore crosses very different factors and parameters. For example, it has been taken
as a characteristic of the individual or the environment, of a group, of a social
system. (Kendra&-Al, 2003)

If we look at what happens within the two professions we have dealt with in
our study, we notice that firefighters have strategies to cope with stress.
«Firefighters are resilient», assert Blaney and Brunsden (2015), although in
their research they indicate that stress shall and must do so to increase
resilience, both on an individual and organizational level.

With regard to the military environment, where professional risks are more
related to the dangers caused by other human beings, rather than just natural
elements, the situation appears more complex.

In the US «the army began to explore the concept of deep resilience after the
terrorist attacks of 11 September 2001.» (Weidlich, 2015). In the same article, the
author deals with the effectiveness of the programs, such as the Care Provider
Support Program (CPSP), created to improve the resilience of the military health
care providers. The article highlights how important it is for care providers to
understand the problems of the population of veterans, such as increased
incidence of PTSD, depression and suicide. And murder, we add, as shown by
the case of Chris Kyle, killed in 2013, along with his colleague Chad
Littlefield, by the twenty-five-year-old Eddie Ray Routh AT a shooting range.
Routh was a fellow soldier who had been diagnosed with PTSD and claimed
to have fired in fear that they were about to kill him or that they wanted to
«steal his soul». He also declared that he was offended because Kyle did not
shake his hand when they first met. Based on the appraisals of his mental
state, the judge sentenced him to life imprisonment without the possibility of
parole. The widow of Chris Kyle, a lawyer, has been active in facilitating
promulgation of a law in Texas to «extend the effort to help the challenges
faced by war veterans and their families every day». (Weidlich, 2015)

These are huge problems: over 1.3 million veterans in U.S. lack health
insurance, about 535,000 of whom fall below the poverty level (Weidlich,
2015)

Another direction in which the characteristics of resilience in pilots were
investigated concerns the use of the 'surprise' factor in training.
«One pilot noted that he had to think for a moment and regain control» in
the surprise condition. Another indicated that he felt «completely
unprepared», that he had a different «mental image» of the upcoming task
and «had to switch». Pilots also remarked that the distraction method (i.e.,
being asked a question and turning away from the display) was very realistic
and representative of distractions in practice. Finally, it was noted that in
particular the conviction that a new phase of the experiment had started took
them out of «performance mode», which made them feel unprepared and
taken off guard by the stall in the surprise condition. Our outcomes substantiate recommendations of using an element of surprise in the training of «upset recovery». (Landman et al. 2017)

The aim of this paper is to highlight the role of the subject in facing the human workload, intended as demanding tasks and emergency situations. It will be enlightened by psychoanalysis as an approach that can provide novel insights, not only into human errors, but also into human resilience. Psychoanalysis provides a different yet intriguing perspective in order to recognize the role of the Ego in facing the human workload.

This paper aims to show new connections among Human Factors, Human Workload and Resilience. They have a common denominator, at least in part: the role of subjective contributions even in demanding situations, in which the person-systems and the technical-systems could be parallel.

Among psychologists, resilience refers to three general meanings: good developmental outcomes despite high risk status; sustained competence under stress; recovery from trauma (Fleming and Ledogar, 2008). Luthar (2000, 2006) called resilience a construct with two distinct dimensions: significant adversity and positive adaptation despite adversity. As a consequence, researchers claim that resilience is never directly measured, but only indirectly inferred from evidence of connected dimensions. They insist that the process of resilience is related to a given context, domain and age. Context involves social/environmental conditions and culture, which will determine if a factor is protective or not. Ungar explained that resilience is «a multidimensional construct, the definition of which is negotiated between individuals and their communities, with tendencies to display both homogeneity and homogeneity across culturally diverse research settings» (Ungar, 2005, p. 219). In reference to the possibility of measuring resilience, Ungar argues that qualitative methods are especially relevant. He stated that they are: «well suited to the discovery of unnamed processes; they study the phenomenon in very specific contexts, their trustworthiness strengthened by the thickness of the description of that context; they elicit and add power to minority ‘voices’ which account for unique localized definitions of positive outcomes; they promote tolerance for these localized constructions by avoiding generalization in favour of transferability; and they require the researchers to account for the bias inherent in the social location» (Ungar, 2005, p. 86).

Among the subjective constituents of human factors in facing the human workload, resilience shows, in a very impressive way, the role of the Ego as an adaptive resource. In this regard, our paper could strengthen the concept of resilience thanks to the Freudian doctrine. Among psychologists, resilience refers to the possibility of overcoming stressful situations with the least
damage possible. Freud’s approach gives us the possibility of comparing resilience with the process managed by the subject to overcome the mourning event. *Mourning and Melancholia* (Freud, 1915) describes the subjective faculty to start investing thoughts and affections in external reality again after having experienced mourning. This kind of work does not merely aim to limit damage, but also to obtain new opportunities. On the one hand, the work of mourning is more than adaptive, it is creative. On the other hand, each mourning, however painful, is still a normal experience, not necessarily destined for melancholy, nor for what nowadays is called depression. Resilience resembles what Freud called “working through”: when the Ego comes to light after mourning, it is a resilient Ego. Regarding the importance of collaboration or partnership, in Freud we read: «After primal man discovered that it lay on his own hands, literally, to improve his lot on earth by working, it cannot have been a matter of indifference to him whether another man worked with or against him. The other man acquired the value for him of a fellow-worker, with whom it was useful to live together.» (Freud, 1929, page 99) Our psychological approach, based on psychoanalytic discoveries, suggests a more proactive analysis on the development of resilience skills, in order to discover how individuals prepare themselves to resist the strong and even negative effects of stressful events and situations and support their overall personal well-being (Grant et al., 2014; Deppa et al., 2016). We find resilience in thinking and coping skills that are employed on the job as well as at home and in other circumstances (Malaguti, 20015; Rozenfeld, 2014).

In our research we consider resilience as the subjective mental resource in facing risks and managing workload. Mental workload affects human performance in terms of the amount of mental capacity required to perform a given task. In general, mental workload is considered to be a multidimensional construct involving interactions between task and system demands, the operator (including mental and emotional skills) and the environment. Subjective mental workload is multidimensional and its consistency is the result of interactions between the human, the task, and the environment. (Estes, 2015) Recent works consider the relationship between the concepts of mental workload, situation awareness and operative performance that can support the success of operations even in the case of variations due to circumstances beyond an individual’s control. (Borghini et al., 2012) Uncertainty and risk are such a kind of situation beyond an individual’s control even if some researches explore the individual differences in hazard perception and their relationship with mental workload and risk behavior (Di Stasi et al., 2009) We have to consider the «risk as feelings», on the border
between rational–analytic reasoning and the experiential–affective system that jointly affect the individual’s risky decisions (Kahneman & Frederick, 2005).

Perceived risk may be considered a neglected factor that contributes to the mental workload in a very particular way, because of its psychological dimension of balance between the likelihood of failure and its consequence. Indeed, the mental workload can increase with the perception of risk independently of the actual task performance and difficulty (Tevell & Burns, 2000).

Exploring the perceived margin between task demands and an individual’s motivated coping capacity, researchers consider the workload as equated with job demand. The measurement and management of job demands would serve to delineate more clearly the separate effects of workers’ capacity-limited and motivation-limited factors on their performance and associated affective states. There is a need for more multidisciplinary research in which valid and sensitive measures of both physical and psychological work demands, as well as a comprehensive range of other potential risk factors, are employed. Even if the relationships between workload, performance, satisfaction, stress and health are not well understood, there is a high level of evidence that factors such as job security, level of control and various other sources of satisfaction are also important influences. It is a common belief that people who enjoy their work - for a range of possible reasons - are to some degree protected from what might otherwise be the ill effects of excessively high workloads. (MacDonald, 2003)

### 4 The resilient Ego at work on the earth as well as in the sky: an empirical study about Firefighters and Military Pilots in their managing risks and workload.

#### 4.1 From the theoretical background to the research questions

Despite the fact that the concept of resilience is rather complicated and deep in content as well as quite complex for an assessment and measurement, increasing research on resilience in extended contexts and dimensions is being carried out. (Bahmra & al., 2011; Lisniak, 2015)

We consider resilience as the subjective mental resource in facing risks and workload. Although the quantitative methods remain privileged in the field (Bakker & al., 2007; Herbst & al., 2014; Estes, 2015), a qualitative data collection involving field operators could benefit from this kind of work and
help to elicit workload aspects related to the subjective side, both in positive and problematic terms.

What we are about to present is a pilot study that aims to measure the resilience of firefighters and military pilots in terms of the subjective perception of the job and its expected tasks. In both jobs daily performances configure defined units of time and space, strongly characterized by a demand for highly specialized tasks that often are related to the personal safety of workers and/or others.

In the case of rescue operations, the time factor is decisive with respect to the success of the operation. As well as in the case of military operations the specialization of these characterizes the task as a situation in which both the high mental workload and the mental supporting of adaptive mental support are experienced. Peculiar mental contents characterize the specific operations in both jobs, on the earth as well as in the sky. Focusing on the mental content that fosters the development of increased self-efficacy, increased social support and flexible and accurate thinking habits allow us to discover which subjective factors promote success in the management of risks.

The sample of our research was composed of a group of firefighters and a group of military pilots. They are people typically engaged in delicate operations in which individual skills and teamwork coordination are decisive in order to obtain maximum efficiency. Both the jobs of firefighters and military pilots have in common the experience of risk as an indelible part of professional life. Their personal perception of their profession is composed of cognition and feelings related to the structural aspects of their jobs and the performative aspects of their specific tasks and operations. The concept of workload and mental workload were emerging in aviation industry, where the need to evaluate the performance of pilots and air traffic controllers was increasingly necessary. Mental workload reflects not only task specificities, but also performer features: individual capabilities, motivation to perform the task, as well as physical and emotional states that affect the workload experienced. High levels of mental workload occur when task demands exceed performer capacity.

Our research aims to lead a qualitative analysis of professional experience in terms of subjective perceptions. In both jobs the strength of the Ego is very important for the success of operations and for the personal safety of the single worker. In particular, we will explore which kinds of mental contents accompany specialized performances.

Our path suggests an analysis of connections among human factors, mental workload and resilience inspired by the psychoanalytic point of view. In
particular, Freudian discoveries have helped us to focalize the importance of contributions of single individuals even in harder conditions: it was with this aim that we recalled the case of *Sully*. It is not difficult to observe that both emergency teams and military crews are composed of individuals who have to manage risk conditions and have to be ready for this special aim.

In doing their jobs, firefighters as well as pilots experience very intense situations, both related to emotions and feelings and also to professional and specific tasks (Setti et al., 2015). On the one hand, the difficult situations met in their jobs challenge their behavioural and mental health, but on the other hand, their behavioural and mental health can be seen as a favourable condition and even as a resource during specific operations.

We will analyse:

- some common structural aspects of both jobs in terms of subjective perceptions: risks, team work, workload, belonging to a special department;
- some common performative aspects of expected performance of both jobs during their specific operations (hours of flight for military pilots, rescue operations for firefighters), in terms of subjective perceptions: instructions, team-colleagues, members of family, personal safety, people involved (passengers for flights and victims for firefighters).

We will take into account some aspects of the relational and educational background of our sample: age, education (what kind of school and/or university they attended), hobbies and sports, composition of their original family and current family, the family reaction to their choice of job, how the decision for this kind of job came about.

We will answer two research questions:

- **RQ1.** The subjective network of resilience. How the perceptions of risk and workload are linked with other aspects of the job and performance and what kind of personal resources are called at the stage, based also on the relational background of the subjects.
- **RQ2.** Resilience on the earth and in the sky. Which differences we can observe between the two sub-samples of Firefighters and Military Pilots in terms of subjective perception of job and performance.

### 4.2 Methods and collecting data

Subjective self-report measures consist in the application submitted to the worker, concerning perceived level of task demand. Actually «no one else is more prepared to provide an accurate judgment on workload experienced than
oneself» (Pereira da Silva, 2014, p. 314), even if it becomes difficult to discriminate between physical workload and mental workload, and the person may consider that the external task demands and the mental effort experiences as only one perception, and quantify jointly the mental effort invested.

Our sample is composed of 53 firefighters and 53 military pilots\(^2\) who answered a survey with 9 items based on the Likert scale from 1 to 10.

For each of the 106 we obtained through the survey 9 qualitative evaluations we called indicators:

- of them in terms of subjective perception of the «setting» of their job, they are risk, team-work, workload, sense of belonging to a special profession;
- of them in terms of subjective perception about the values that affect performance during operations, they are instructions, colleagues, personal safety, family, people involved in operations.

The matching of the observed indicators gives us the subjective network of resilience composed on the one hand by the perception of structural factors of one’s own job and, on the other hand, by the individual perception of expected performance even when they manage the risks.

We will match with defined indicators the aspects of the relational and educational background of our sample: age, education (what kind of school and/or university they attended), hobbies and sports, composition of their original family and current family, the family reaction to their choice of job, how the decision for this kind of job came about.

We applied some statistical operations\(^3\) (correlations, contingency tables with chi2, averages comparison with t-student or ANOVA) considering the complete sample of 106 workers and the two separate sub-samples of firefighters and military pilots.

All qualitative indicators were considered dependent variables. They were analysed also based on independent variables taken from the subjective background relating to age, educational path, sport activity during free time, family reaction to their choice of job and from how the decision for this kind of job came about.

---

\(^2\) We would like to thank Sabrina Venzi, psychologist of the Italian Air Force for her help in collecting data.

\(^3\) We would like to thank Michela Brunori for her help in statistical analysis of data.
4.3 Results about RQ1. The network of resilience

Table 1. The network of resilience (Pearson’s Correlations - 106 subjects)

<table>
<thead>
<tr>
<th></th>
<th>Risk</th>
<th>Team Work</th>
<th>Workload</th>
<th>Belonging</th>
<th>Instructions</th>
<th>Team-mates</th>
<th>Personal Safety</th>
<th>Family</th>
<th>Passengers - victims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Team Work</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workload</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belonging</td>
<td>*</td>
<td>**</td>
<td>*</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>*</td>
<td>**</td>
<td></td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team-mates</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Personal Safety</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Family</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Passengers - victims</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>

We can observe (table 1) many significant correlations among our observed indicators. All indicators are linked with one or more of the others. We can consider these conjunctions as a confirmation of resilience as a complex and many-sided subjective dimension. The perception of risk increases on the one hand with the perception of workload and the satisfaction of belonging to a special service, on the other hand risk is linked with apprehension for family relationships and the sense of personal safety. The perception of team work as a specific factor in both jobs is specular with the importance reserved to team-mates during performance. It also associates the importance of family members and other people involved during performance, confirming a sort of altruistic propensity of firefighters and pilots. Such an indicator is even linked with the sense of belonging to a special service and the importance reserved to the instructions during performance, indicating the realism of perception of the cooperative nature of both jobs. The workload goes hand in hand with the perception of risk and the satisfaction of belonging to a specialized team. The sense of belonging is the most combined indicator. It is associated not only with perception of the structural aspects of both jobs, such as the risk, the workload and the team work, but also the factors of their performance such as the importance reserved to instructions, to team-mates and family relationships. The given importance to the instructions goes hand in hand with the value reserved to team-mates: pilots and firefighters are very aware that the knowledge of their work is composed of the specific cognition of the tasks as well as by the affection for the colleagues involved within each real operation. The feelings towards team-mates during operations are also linked with the other people involved within the operations, both passengers of flights or victims of rescue operations. The sense of personal safety is linked with apprehension for family relationships and people involved within performance. We observe such a link between family and the others involved. Both jobs link the subjective attention for oneself with the care for the others.
involved. By comparing the averages, we can observe which independent variables, taken from the personal and relational background of our workers, influence the indicators of resilience.

**Table 2. Professional choice (ANOVA – 106 subjects)**

<table>
<thead>
<tr>
<th>IDEA</th>
<th>During military service</th>
<th>By chance</th>
<th>Since childhood</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>50.71</td>
<td>47.96</td>
<td>43.98</td>
<td>.003</td>
</tr>
<tr>
<td>Belonging</td>
<td>9.57</td>
<td>8.79</td>
<td>9.61</td>
<td>.006</td>
</tr>
</tbody>
</table>

We can see that the professional choice (table n. 2) is related to the age of the subjects and gives a direction to the sense of belonging to a special team. Those who have cultivated the idea since childhood are younger and perceive a stronger sense of belonging, showing the perception of vocational nature of both the jobs we are considering. By analysing the influence of family reactions to the professional choice (table n. 3) we can observe that it changes the consideration of team work as a structural indicator of both jobs. When the relatives’ reaction to the professional choice was «worried», it increased the relevance of team work in the subject i.e. as if the current team can repair the worried evaluation of the «birth team».

**Table 3. Birth family reaction to the professional choice (t-student – 106 subjects)**

<table>
<thead>
<tr>
<th>BIRTH FAMILY REACTION</th>
<th>Positive</th>
<th>Worried</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team work</td>
<td>9.61</td>
<td>9.84</td>
<td>.003</td>
</tr>
</tbody>
</table>

There is also an influence by the current family reaction to the professional choice (table n. 4) on resilience indicators. Positive reactions promote the sense of belonging and the importance reserved to team-mates during performance. On the contrary, those who have to manage worried reactions are more concentrated on personal safety during operations.

**Table 4. Current family reaction to the professional choice (t-student – 106 subjects)**

<table>
<thead>
<tr>
<th>CURRENT FAMILY REACTION</th>
<th>Positive</th>
<th>Worried</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belonging</td>
<td>9.46</td>
<td>9.11</td>
<td>.000</td>
</tr>
<tr>
<td>Team-mates</td>
<td>9.48</td>
<td>9.32</td>
<td>.001</td>
</tr>
<tr>
<td>Personal safety</td>
<td>6.52</td>
<td>6.79</td>
<td>.041</td>
</tr>
</tbody>
</table>

Another interesting observation comes from the analysis of interests of our workers during their free time. We noted that sport is the main interest among our subjects.
Here (table n. 5) we see that those who are interested in sport activities have been in the force for a shorter period of time and follow the instructions better during performance but are less worried about team-mates during operations. Team-mates become a stronger support for those who are less keen on sport.

The results presented until now allow us to answer the first research question about the configuration of resilience as a subjective network, composed of many aspects related both to the structure of the job and the nature of performance. Job and performance are perceived in a very personal way and from their conjunction we can observe the level of capability to face adversity. All the aspects are connected with others. Among them three emerge for their relevance: the first is the sense of belonging to a special service, the second is the importance of teamwork, the third is the role reserved to the thoughts for family relationships during the operations. Three aspects that confirm the perceived collaborative nature of both professions: firefighters and military pilots are supported by relational competence and tied to an affective background.

### 4.4 Results about RQ2. On the earth as well in the sky? A comparison between firefighters and military pilots.

Now we will explore the main differences between firefighters’ and military pilots’ resilience (table n. 6).

<table>
<thead>
<tr>
<th>CAREER</th>
<th>Firefighters</th>
<th>Pilots</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belonging</td>
<td>9.21</td>
<td>9.58</td>
<td>.035</td>
</tr>
<tr>
<td>Instructions</td>
<td>7.25</td>
<td>9.30</td>
<td>.000</td>
</tr>
<tr>
<td>Team-mates</td>
<td>9.17</td>
<td>9.73</td>
<td>.000</td>
</tr>
<tr>
<td>Family</td>
<td>5.81</td>
<td>5.92</td>
<td>.001</td>
</tr>
</tbody>
</table>

The sense of belonging related to the structure of the job as well as the recognized importance given to instructions, team-mates and thoughts for family relationships are perceived more by military pilots compared to firefighters. But instructions and team-mates are experienced by firefighters in a very significant different way from pilots. The value of instructions in particular is higher for pilots, showing the really specialized nature of flying, in which they have to absolutely follow the instructions and carry out the
expected behaviours during ordinary operations. The nature of the firefighters’ performance is different, because it is very dependent on the one hand on the variability of scenarios and on the other hand conditioned by unforeseen conditions while managing rescue operations. We noted also a few differences related to the subjective background of the two professions. The first is related to their educational experience. Pilots show a higher level of education (see graphs Job*Education).

The second is related to the interests followed during time after work. Pilots show more propensity to practice sport activities during free time (see graphs Job*Sport).
The third finding is the most interesting. It is related to the stronger vocational feeling of pilots towards their job (see graphs Job*Idea).

In order to understand the direction and the influence of these differences on resilience and risk perception, and further analyse our sample, it would be necessary to increase the number of observed cases.

4.5 On the earth

On the whole, in the sample all the indicators of the firefighters are linked with other aspects, often in a very significant way (table n. 7). The sense of belonging and the importance entrusted to team-mates are aspects which are more closely linked to the other indicators. The satisfaction of belonging to a special service supports the perception of risk and team work, mediating also the perceived workload. We can say that the sense of belonging holds together the structural dimensions of firefighters’ jobs. We can imagine that without the sense of belonging it would be impossible to be a firefighter and face all adversities in a positive way. The same sense of belonging supports the importance reserved to instructions and team-mates during the rescue operations, mediating the collaborative character of this very special work, so vital for the community. The perception of risk is linked with the perceived workload.
Table 7. The network of resilience of firefighters (Pearson’s Correlations - 53 subjects)

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Risk</th>
<th>Team</th>
<th>Workload</th>
<th>Belonging</th>
<th>Instructions</th>
<th>Team-mates</th>
<th>Safety</th>
<th>Family</th>
<th>Victims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workload</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belonging</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team-mates</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safety</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victims</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

During performance the workload is supported by team-mates. The team-mates also support the importance reserved to instructions during operations. Firefighters are aware that the possibility to work as a team is the fundamental condition to achieve success, supported by the performative conjunctions between observed instructions and team-mates who work together. It is interesting to observe the links of the indicator called *instructions*: the prescribed actions from the guidelines is linked with team work, as well as the team-mates and the people involved in the rescue operation. This is a very significant sign of altruistic character of the resilience of firefighters perceived as a structural condition of their daily job. The altruism is confirmed by the strict connections among personal safety, thoughts concerning family relationships and victims involved. Only age diminishes this structural altruism, as we learn from the link between age and perception of team work: as the indicator age increases, the importance of teamwork decreases.

4.6 In the sky

The resilience of military pilots (table n. 8) seems especially based on some of the perceived factors of performance. Comparing military pilots with firefighters, we have to observe that three indicators – workload, belonging and instructions – remain unlinked with other aspects. The perception of risk is linked with the sense of personal safety and with thoughts regarding family relationships. Among the indicators of the structure of their job, only team work supports the importance of team-mates, thoughts concerning family relationships and thoughts for any passengers on the flight. The indicator which is more closely connected to the other ones is the family: it supports not only the perception of risk and team work, but also the sense of personal safety and apprehension for passengers.
Table 8. The network of resilience of military pilots (Pearson’s Correlations - 53 subjects)

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Risk</th>
<th>Team work</th>
<th>Team-mates</th>
<th>Personal Safety</th>
<th>Family</th>
<th>Passengers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk</td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team work</td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team-mates</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passengers</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

The last interesting observation concerns the direct link between age and the importance of team-mates during operations: as age increases, the role of team-mates also increases.

Now we can answer the second research question, concerning which differences we can observe between the two sub-samples of firefighters and military pilots in terms of the subjective perception of the job and performance. On the earth the satisfaction of belonging to a special service supports the perception of risk and the experience of the workload. The resilience of the firefighters is based on altruistic feelings experienced as a structural condition of their daily job. In the sky, the perception of risk is linked with the sense of personal safety and with thoughts concerning family relationships. The resilience of military pilots is supported in a special way by family relationships that maintain a considerable importance during performance. If, for military pilots the cooperative exercise during performance is more important, for firefighters it is the structure of the job supported by an altruistic propensity. Our findings show the configuration of resilience as a subjective network, composed by many aspects related both to the structure of the job and to the nature of performance. According to other researches (Kowalczyk et al., 2015) we can confirm that the resilient Ego at work manages both job and performance in order to face the adversity positively.

All aspects are linked with some other indicators. Among them, three emerge for their relevance: the first is the sense of belonging to a special service, the second is the importance of team work, the third is the place reserved to family relationships during operations. These three aspects confirm the cooperative nature of both professions, even if in a different way: both firefighters and military pilots are supported by relational competence and tied to an affective background.
5 Conclusions

Many authors point out that psychological difficulties of firefighting come from the long periods of relative inactivity punctuated by highly stressful situations, such as rescue operations in which they have to face considerable responsibilities for life and property as an addition to task demands. Individual firefighters in an active fire service are under considerable selection pressure for risk and rescue. The demands of firefighting are sporadic and unpredictable, characterized by long periods of waiting between tasks of intense activity. «This irregular pattern of activity is an important feature of firefighting, adding to the component of stress likely to be due to anxiety and responses to psychogenic stress.» (Guidotti, 1992, p. 5) The energy costs of specific firefighting tasks are more specific than psychological involvement. Firefighters' willingness to face risks can be better understood by analysing the psychological dimensions of their job. Listening to their stories, we note that they «have dealt with public expectations and their own fears by overcompensating with displays of personal control. Today (...) they take much more care to protect themselves and avoid unnecessary risk.» (Guidotti, 1992, p. 9) Managing risky situations is part of the job in military aviation as well. That is because the «military operates in a high stakes environment where crews with limited resources deal routinely with complex and time critical operations. As in many areas of healthcare, personnel can be under continual constraints. Innovation, improvisation, and timely information are critical to success. Military aircrew, like clinicians, learn to deal with periods of downtime, that rapidly gear up for challenges that require full concentration often for prolonged periods.» (Clay-Williams, 2013, p 1) Limitations imposed by the environment or mission, an increased physical and cognitive load, the psychological arousal because of a demanding operation can be exacerbated by fatigue. The job structure includes on one hand operating in dynamic conditions with rapid and unpredictable changes in a restricted time period. On the other hand, pilots make decisions in order to solve problems under pressure with limited information. We can observe «the role of a pilot in combat. The environment is complex and there are a lot of things to do, such as fly, navigate and use weapons. Pilots gradually learn strategies to simplify their world, so that radar contacts are either 'friend' or 'foe' and firing weapons, at least initially, become a simple matter of eliminating a 'foe'. Pilots also learn that the majority of rapid responses cannot be thought, but have to become automated responses to specific problems. In the same way, responses cannot be theoretical notions of how to respond, but rather as highly automated psychomotor patterns learned through long practice.» (Byrne, 2017, p. 190)
We suggest extending to firefighters the following job description of military pilots: they «learn about the limits of human performance, such as managing stress and fatigue in oneself and others, and managing attention so as not to become fixated on one task or to lose awareness of what is happening in the surrounding environment. They are trained to plan for contingency, to prioritize tasks under rapidly changing conditions, and to maintain capability in deteriorating situations. (...) – they – know what to expect when operating at the limits of performance and how to re-establish normality.» (Clay-Williams, 2013, p. 1) It is a good description of the resilient Ego, that we find in activity in the brilliant, competent and prudent workers of our sample. Based on our findings, we can add some subjective psychological sources of this resilience: it can rely on the sense of belonging to a special job and on relational experience based internally, on the one hand on the support of team mates, and externally, on the other hand on family relationships as the extensive affective support for such a very demanding job.

As we said in chapter 3, the psychoanalytic approach suggests a more proactive analysis on the development of resilience skills, in order to discover how individuals can resist the negative effects of stressful events. We find resilience in thinking and coping skills that are used at work and at home and in other circumstances. (Malaguti, 2015; Rozenfeld, 2014). Resilience, supported by subjective mental work, could confirm the role of the resilient Ego as a promoter of Human Factors, even within harder missions. The role of psychoanalysis in interpreting and influencing human performance in work systems, and particularly in emergency situations, could offer a new interpretation to very important fields, such as pilots’ and firefighters’ jobs.

On the subject of flying: «why do so many people dream of being able to fly? The answer that psychoanalysis gives is that to fly or to be a bird is only a disguise for another wish, and that more than one connection involving words or things, leads us to recognize what it is. When we consider that inquisitive children are told that babies are brought by a large bird, such as the stork (…) we learn that in dreams the wish to be able to fly is to be understood as nothing else than a longing to be capable of sexual performance. This is an early infantile wish.» (Freud, 1910) And about fire control, according to Freud, «in order to gain control over fire, men had to renounce the homosexually-tinged desire to put it out with a stream of urine», and he finds confirmation of this idea in the interpretation of the Greek myth of Prometheus, the Titan who «brought fire to men, having stolen it from the gods, hidden in a hollow stick, a fennel-stalk». Freud is asking why the acquisition of fire in ancient myths is always connected with the idea of a prohibition. «Who is it that was injured or defrauded by it? (...) It is he a god, then, who were defrauded». (Freud, 1931) And the gods, as we know,
represent the satisfaction of those wishes to which we have to renounce. So, to work in the quenching of fire is a sort of working with the ancient gods. In this research we wished to approach and study the psychic reality – thought and affections – of professionals such as firefighters and military pilots who, in their work, are not only able to face events that present themselves as dangers and threats for themselves and for the whole community and get out unharmed but show that they do all this with personal satisfaction. And this is something that cannot be imposed by anyone. Therein lies an additional resource and teaching that both these professional categories can offer us.

We cannot and do not want to do without civilization and its acquisitions (from the invention of numbers, to that of the wheel, to the discovery of X-rays and their use in medical science, or to the discovery of oil, without which we would not have had aeronautics). Sigmund Freud (1929) as well as Giacomo B. Contri (2014) claim that any improvement of subjective psychic life leads to an actual solidity of acquisitions of civilization. Among them the capacity of resilience, both of individuals and their communities, emerges as a priority.

References

Automation Interaction from a Design-Thinking Perspective: The Case of Cockpit Automation in Commercial Aircrafts in Israel

Avner Y. Bendheim
Bezalel Academy of Arts and Design, Martin Buber St 1, 9124001 Jerusalem, Israel
mail@bezalel.ac.il

Abstract. This paper studies the effects of automation on human performance in high-risk environments from a design-thinking perspective. While advances in automation have partially replaced human agency, humans still play a major role in operating and monitoring systems. The issues focused on include the degradation of the operator’s skills, difficulty of understanding automation because of increasing complexity, and "automation complacency." A case study was used as a basis for evaluating the effects of automation in high-risk environments, researching the effects of automation on pilot performance in Israeli commercial airline cockpits. The evaluation concentrated on four areas: the physical and psychological surroundings, crew dynamics, pilot culture and behavior, and key industry drivers. I performed qualitative and quantitative field research that included observations, interviews and surveys. The research identified several hypotheses: 1) Due to automation, the center of operation in the cockpit has shifted, requiring changes in cockpit design. 2) Minor adjustments in physical space could significantly improve human interaction with automation. 3) Advancements in cockpit systems do not keep pace with a person's behavior resulting from interaction with the latest technology. 4) Pilots have transitioned into powerful agents of change in the aviation industry. These hypotheses led to design opportunities and proposals to optimize synergy between pilots and the automated systems.

Keywords: Human Factors and Ergonomics, Human computer interaction (HCI), Collaborative interaction, User interface design, Aerospace, Aviation Psychology, Environmental Psychology, design for behavioral change.

1 Introduction

"If a medical algorithm is proven to produce more consistently accurate diagnoses than a physician, or semi-autonomous cars become more affordable and road-legal, it would be both unethical and legally questionable to refuse to use it. Few fields of human endeavor are likely to remain untouched" (Chatfield, 2016).

Interaction between human operators and automated systems is a key issue in the development of technology. The race to fully autonomous machines is on. In industries where human error is a matter of life or death, autonomous systems bring a dual promise – minimizing casualties and maximizing efficiency. As a result, entire industries are adopting automated systems with improved capabilities, while
receiving significant support from regulatory and government authorities. That said, there are many ethical, technological, psychological and regulatory obstacles to overcome before surgeries are performed without surgeons, and commercial planes are flown without pilots.

In the meantime, human participation remains a critical factor in operating the systems. Until full autonomy is achieved, certain aspects of the relationship between highly automated systems and humans have proven to have negative effects and tragic consequences that jeopardize safety and raise questions concerning the future of automation interaction design. Whether or not fully autonomous systems will become routine, or what the consequences to the human workforce will be, are interesting questions, but are not the focus of this paper. Rather, this paper focuses on the transition period prior to full autonomy – when highly automated systems require human engagement.

Aviation has been at the forefront of using automation to perform various tasks in the cockpit. Pilots have experienced both the benefits of automation, as well as the negative effects on their performance, capabilities and behavior in the cockpit. Extensive research and development have been dedicated to automation and system design in modern aircrafts. Aircraft automation research and development has, in many cases, been the basis of automation and interaction design principles in other high risk automated environments, such as semi-autonomous cars and modern operating rooms. The design research methodology provides a unique perspective and understanding of the effects of automation on human performance. The findings can lead to a better interaction between pilots and automated aircraft systems. I recognize that this research is only relevant for as long as human beings remain part of the equation.

2 Literature Review

Commercial aviation is the safest transportation system in the world to date, and the accident rate is the lowest it has ever been, predominantly due to improvements in aircraft systems and automation that have contributed to safety efficiency and precise flight management and operation (see Fig. 1).

Fig. 1. Fatal Accident rate per year, Commercial Jet Fleet, 1959-2016. Copyright, Boeing Aerospace 2017.
Despite the improvements, pilot interaction and automated systems present new challenges, contributing to many accidents and incidents in aviation (Nakamura, 2013). This chapter provides a framework for understanding the effects of automation on human behavior and performance in high risk environments. Specifically, I review the physical and psychological challenges that pilots face in the cockpit. In addition, reference is made to the profound impact of technology on human behavior. Finally, key industry drivers and trends are analyzed to provide an overview of additional factors that may influence the interaction between automation and pilots, as well as increasing the role of pilots as agents of change in the aviation industry.

2.1 Automation

Before understanding the effects and challenges of automation, it is important to understand the characteristics and functions of automation. The term "automation" was first applied to product manufacturing post World War II and has since expanded to automatic control in many fields, including aviation, automobiles, ships, medical systems, robots, computers, home appliances and more. Automation originally referred to the replacement of, or assistance to human physical labor. The importance of automated systems today is the ability to assist or replace human mental labor as well as physical labor. Computers and sensors that interpret data and make decisions changed the boundaries of automation beyond the scope of functionality and mechanical control (Sheridan & Raja, 2006). As computers became more involved in process and data control, the functionality and nature of automation have changed due to the complexity of cooperation and sharing responsibilities between automatic systems and human operators.

2.2 Levels of automation

The changing level of automation, from fully manual to fully automated, is crucial in considering the role of human performance. The increasing amount of intermediate levels of automation has led to an increase in complexity of its use and a decrease in "situational awareness." The original levels of automation regarded functional control and operation: Manual control - all control is accomplished by humans; supervisory control - some or all of the control is accomplished by the computer, but the human supervisor can assert control; and fully automatic control - all control is automatic and the human cannot affect the process except, perhaps, to terminate it (O’Hara & Higgins, 2010). The National Highway Traffic Safety Administration of America (NHTSA), formalized levels to describe the degree of automation, from level 0 – a totally manual car, to level 5 – a fully self-driving car (see Fig. 2). As technology advanced, computer-based systems expanded the application of automation beyond levels of control.
Data acquisition and information analysis provide the ability to automate information for decision assistance and action implementation. On one hand, this new form of automation allows designers and operators to monitor and perform many functions based on computer analysis and assistance. On the other hand, the many levels of automation control and information, increase the operation complexity and reduce the ability of users to understand what mode of automation is being used (Billings, 1996) (Kaber & Endsley, 2003) (see Fig. 3).

The latest form of automation functionality focuses on the flexibility of allocation. Older systems, called "static allocation," were designed so that the level of responsibility for performing a task was allocated to the human operator or to the machine and did not change during operation. In modern automated systems, the level of responsibility can vary and the task can be performed by personnel or automatic systems based on considerations such as workload or low system reliability. This dynamic form of automation allocation allows for the design of assistive automatic functions in aircrafts and in vehicles (Casner & Hutchens, 2016). An example of a flexible automatic system in vehicles is Automatic/Autonomous Emergency Breaking Systems (AEB). This system was designed to
intervene in critical situations by automatically applying the brakes. Although flexibility in allocation allows for a "tag-team" relationship between the operator and the system, it adds yet another dimension of complexity (O’Hara & Higgins, 2010) (see Fig. 4).

The final factor affecting the level of automation is the reliability of automation. By failing to achieve their intended function, automatic systems may compromise the system's reliability, operators trust and overall safety. Less than perfect reliability can result in misses caused by "false negative" alarms – not alarming when a condition exists, and false positives – alarming when a condition does not exist. Further, a perception of low reliability could cause the operator to misuse the entire system (O’Hara & Higgins, 2010) (see Fig. 5).
2.3 Effects of automation

"Google Maps is making me stupid, about that I have no doubt. After getting out of the subway, I’ll frequently follow my phone’s blue arrow for a block in the wrong direction before the GPS catches up and turns me around" (Carr, 2014). Carr describes how the dependence on automation among pilots, drivers and surgeons degrade the skills of the human operators. The growth and complexity of automation and the loss of skill and over-reliance on automation in high risk environments lead to unsafe operation in the case of automation failure or misuse, often resulting in fatal accidents (Carr, 2014). "Automation Complacency" refers to over-reliance on and delegation of authority to automatic systems, resulting in the loss of ability and in some cases, reluctance to intervene. An FAA report from 2013, identified that in roughly one quarter of accidents involving the use of automation, pilots were overconfident in the automated systems and in some cases, were reluctant to intervene (Nakamura, 2013). Reduced human involvement in various modes of automation make it difficult to get back "in the loop" and resume control over the system. Regarding semi-autonomous vehicles, studies show that people have extreme difficulty in resuming manual control of a vehicle following a period of automated or partially automated control (Johns & Miller, The Driver Has Control: Exploring Driving Performance With Varying Automation Capabilities, 2017). Another significant effect caused by the complexity and variant levels of automation, is the difficulty of understanding automation. Due to the many components and interrelated functions, operators may find automation confusing. The National Transportation Safety Board ruled that a contributing factor in the July 6, 2013 crash of Asiana Airlines Flight 214 crash at San Francisco International Airport, was a complex user interface to the airplane’s auto-flight system that was insufficiently understood and inadequately supervised by the flight crew (NTSB, 2014). Changes in automation design have been made to address the negative consequences of complex automation (e.g. minimization of data on pilot displays). In addition, these issues are addressed in training and flight crew procedures. None the less, the effects of automation continue to contribute to pilot error and accidents in various industries (Nakamura, 2013) (Ridella, 2017).

2.4 Human Factors

Human Factors in aviation is the study of the link between human behavior and performance, and about how to improve the way pilots and systems work together. "In the old view, systems are basically safe, and it is human error that causes most accidents. The new view sees systems as not basically safe, and human errors are symptoms of contradictions, pressures and resource limitations deeper inside the system" (Dekker, 2002). In other words, many other factors contribute to pilot error, such as poor system and hardware design, ergonomics and cockpit culture. Studies regarding human error in aviation show how throughout the history of aviation, cockpit design systematically influenced pilot errors. For example, the similarity between the flap and gear handle often caused confusion amongst pilots during World War II, leading to the re-design and re-location of the flap and gear handles in aircrafts manufactured post World War II (Fitts & Jones, 1947). The analysis of the cockpit as a physical and psychological environment is paramount to
understanding the effects of automation within a contextual space. The physical
and social environment provides a setting that impacts emotions, behavior and
well-being (Namazian & Mehdipour, 2013). For example, Namazian notes the
relationship between seating configurations and the effects on social interactions
and cooperation. This understanding leads to the ability to design social space to
achieve desired behavior.

2.5 "The third pilot"

The concept of Crew Resource Management (CRM) revolutionized the aviation
industry in the late 1970s. Many fatal aviation accidents resulted from poor crew
interaction and prompted the aviation industry to focus on the importance of social
interaction within the cockpit (Helmreich, Merritt, & Wilhelm, 1996). Organiz-
tional safety and cultural behavioral are also regarded in CRM and have critical
impact on crew interaction and actions. For example, extreme hierarchy in Korean
culture, was found to be a contributing factor to aviation accidents which occurred
on Korean flights in the 1990s, where co-pilots were reluctant to correct mistakes
made by the captain, at the cost their own and the passengers safety (Engle,
Culture in the Cockpit-CRM in a Multicultural World, 2000). CRM regards
automation as a resource that is to be managed by the human crewmembers.
Advancements in automation capabilities and complexity challenge that
assumption. Studies now regard automation as an additional crew member, which,
on one hand, can perform many of the flight's functions, and, on the other hand,
can make mistakes and require adequate monitoring and intervention. "It is time to
consider cockpit automation as another pilot, one capable of routine brilliance but
occasional catastrophe" (Albrught, 2016).

2.6 Aviation industry drivers

Global aviation growth is at an all-time high, with a forecast market value of 6.1
trillion dollars and a growth rate that predicts over 41,000 new aircraft deliveries
worldwide by 2036. In addition, the regulatory environment, or deregulated
environment, has led to the independence of the airline market, allowing them to
reduce fares, improving service, and create new business models (Boeing, Current
Market Outlook 2017-2036, 2017). The predicted increase in air travel has
significant implications for pilot demand, changing current pilot requirements and
recruitment channels worldwide. Increased airline pilot demand forecast the need
for nearly 300,000 new airline pilots over the next 10 years to meet industry growth
requirements, as well as the need to transition 180,000 co-pilots to captains by
2027. Consequences of this high demand require the re-assessment of cockpit
functionality and design, as well as the current training and requirement system,
such as adjustments to the fact that first officers with half the experience will
become captains (CAE, 2017).

Lastly, de-regulation of the aviation industry introduces new competitive business
models, such as low-cost carriers (LCCs). Although beneficial to consumers,
secondary effects of increasingly competitive business models include atypical forms of employment, such as contracts via temporary work agencies or Pay-To-Fly schemes – an industry practice wherein airlines require inexperienced pilots to pay for their ‘line-training’ on board flights. Atypical forms of employment provide another example of an industry driver affecting company cultures and consequently flight safety (Jorens, Gillis, Valcke, & Coninck, 2015).

3 Research and methodology

3.1 Methodology selection

Researchers from academic and industry sectors are constantly evaluating automation interaction and interface design. The majority of research methodologies regarding cockpit design rely on empirical testing and quantitative data. This approach is necessitated due to the complex development process of cockpit design. Technical requirements, flight test certification procedures and development limitations rely on previous data provided by test pilots and simulations. Often, the operational line pilot, who in fact is the final user of the product, is not involved in the development process or able to change the design based on operational evaluation (Singer, 2002). Recurring accidents involving poor user interaction challenge previously used research and development methodologies.

3.2 Field research and data collection

The theoretical literature provided me the basic tools which served as a valuable foundation for the qualitative field research on the use of automation in high-risk environments and cockpit surroundings. The research population of the case study was limited to Israeli commercial pilots, though the field research included a combination of observations and interviews of end users and specialists regarding automation interaction and design, psychologists, aircraft and autonomous system engineers, surgeons, and pilots from private, military and commercial industries. I observed multiple flights onboard Israeli airlines and military transport sorties, and an authorized observation of a neurosurgical procedure at an Israeli hospital. These allowed for a first-hand evaluation of end-user performance and behavior while interacting with highly automated systems in high-risk environments (see Fig. 6). After collecting data from the interviews and observations, questionnaires and surveys were used to corroborate research theories with quantitative data. Those insights led to findings, which ultimately served as a basis for new design opportunities to optimize automation interaction and synergetic co-operation between the Pilots and aircraft systems.
3.3 Research limitations:

Practical limitations regarding research assumptions, methodologies are addressed in the following section, as well as conditions that may influence information bias and data validity.

**Personal background.** I am a certified commercial aircraft pilot with over 10 years of flying experience. In addition to flying multiple transport aircrafts, in 2014 I was assigned by the Air Force to manage a sub-department of aircraft system development and design, where I evaluated the pilot interface and design of numerous aircraft systems. On the one hand, my background and knowledge of aircraft systems assisted in aspects of research and provided access to a variety of interviews and observations. On the other hand, I recognize that my objectivity regarding certain results may have been affected by personal experiences and opinions. To minimize this issue, when possible, I have validated findings on issues using empirical methods of research.

**Case study limitations.** Although cockpit automation has many similar principles to automation in other fields, the ability to generalize may be questionable, due to the many variables in a given high-risk environment. As described in the theoretical review, cultural differences can influence crew interaction and behavior. Thus, by limiting the scope of the case study to Israeli airlines, the generalization of results and behavioral insights to different cultural environments could affect data reliability.

**External factors not included in the scope of the study.** Due to the limited scope of the study, I did not include the factors identified below in the analysis. However, it is clear that these areas will need to be evaluated before making any final decisions regarding the development of new products and systems.

*Communication between pilots and Air Traffic Services.* Coordination between pilots and ATC’s have significant importance to flight safety. Amended instructions can lead to misunderstandings and errors. This issue is very broad, involving external communication with control towers, ATC’s, ground service and others. Although communication with Air Traffic Services can influence pilot performance, it is not directly affected by automation, and extends beyond the confines of the cockpit, and therefore is not included in the study.

*Methods and content of training procedures.* Training procedures are clearly vital to pilot performance. Recurring accidents resulting from automation have led to
changes in flight training. Training alone cannot remedy the problems. Rather, improved design must accompany good training. Training procedures have been extensively researched and modified worldwide, and therefore not included in the study.

3.4 Evaluation and validation process

Further research and validation through prototyping and experiments will be conducted to evaluate the findings and design opportunities regarding automation interaction and synergetic co-operation between the pilots and the automated systems in the aircraft. Testing of prototypes and experiments will be conducted via flight simulators and onboard flights, depending on the nature of the experiment and authorized procedures. In addition, opportunities to publicize and present the research process and findings will be evaluated during this time.

4 Hypotheses

4.1 A new center of operation

*People think all we need to do in the flight is take-off and land; the rest is autopilot. The truth is that automation doesn't necessarily reduce our workload; rather, changes the nature of requirements and actions performed in the flight* (E, Commercial airline pilot). Automation has changed the pilot's work requirements. Once, most of the pilot's workload involved manually flying the aircraft. Today, in standard operations, most of the actions performed by the pilot are operating and monitoring the automatic systems (see Fig. 7).

![Fig. 7. 787-8 "Dreamliner" cockpit layout.](https://www.reddit.com/r/aviation/comments/4mgvt4/united_boeing_7878_dreamliner_cockpit/)
Technically, the use of automation should increase the crew's ability to monitor the aircraft's systems and flight data, but studies show differently. In fact, secondary effects of automation, such as complacency and over-reliance, reduce a pilot's ability and likelihood to identify automation malfunctions (Nakamura, 2013). Field observations identified additional effects of increased use of automation during flight. For example, throughout the flight, the crew is required to manually input data and alter flight modes. Onboard a commercial flight from Eilat to Tel Aviv, I recorded over 150 manual inputs and alterations to the automatic systems throughout the 40-minute flight, as well as 100 more manual inputs during preflight and ground operation. The manual inputs made by the crew vary in type and nature. First, during pre-flight and ground procedures, the crew input most of the planned flight data, such as the expected route, flight altitude and speeds, number of passengers and so on. While most of the data is known in advance, some data can change, depending on tower and traffic limitations, weather, fuel considerations and more. During the flight, most of the manual input involves changes in the automation level, or alteration of flight path, speed and altitude. When the crew needed to make unexpected changes, their workload greatly increased, because they had to verify and input data, alter flight trajectory, and change selected automation level or mode. The increased amount of data input not only increases the likelihood of human error, but also makes it very difficult for crew members to monitor one another and cross-check for errors. A survey of 40 airline pilots regarding the causes of pilot error showed that 32% of pilot errors are a result of insufficient cross-monitoring because of automation. Furthermore, it was noted that the pilots' interaction with automation created a new form of communication between crew members themselves. The specific method chosen, as well as the speed and accuracy of the action, can affect the quality of communication, and even the relationship between crew members. In many cases, "Standard Operational Procedures" (SOP) suggest the optimal way to perform these functions. The increased likelihood to make mistakes and the decreased ability to monitor while inputting data have additional implications to the quality of the crew interaction despite suggested standard procedures. Further, the growing complexity of automation, which can easily cause mode confusion and misuse of the system, makes it even more difficult for the pilots to communicate.

While the theoretical framework and research shows that automation presents significant challenges to a pilot's capabilities, this finding shows further implications to cross-verification and crew communication. Recurring accidents and pilot errors as a result of automation indicate that questions about cockpit systems and design need be considered.
4.2 Minor adjustments in the physical space

As in cockpit automation, neurosurgeons interact with automation during surgical procedures. Complex systems, such as brain mapping, require manual input of data, control and monitoring. During an interview, the head of the neurosurgery department in an Israeli hospital raised concerns regarding the relationship between brain surgeons and automated systems. "Surgeons must constantly evaluate the necessity and accuracy of the automation. I've seen several cases of mistakes caused from improper use of automation. For example, during a surgery performed by a colleague of mine, the crew used the brain mapping device to assist with navigation, without noticing that the image presented on the monitor was showing playback data, and not live data. By the time they noticed, it was too late." (Prof. C, Head of Neurosurgery department).

Due to many limitations and regulations, it is not always possible to improve the systems themselves; rather, the surgeons must be aware of the weaknesses of the automations, and make the necessary adjustments to avoid costly mistakes. During deep brain surgery to remove a tumor, surgeons were constantly adjusting the physical surroundings of the operating room in accordance with the automated system being used. For example, when working with a live mapping device, the nurses placed two chairs strategically at an angle, so the surgeons could comfortably view the monitor, each other and the patient. This seating arrangement allowed for better collaboration between the surgeons, as well as the ability to better operate while monitoring the navigation system at the same time (see Fig. 8).

![Surgeons in unique seating arrangement while interacting with navigation system. Sighted during Research observation.](image-url)
In high-risk environments, these kinds of changes to the physical environment are crucial because the other option, system upgrades, take too long. Ali Namazian, in his paper "Psychological Demands of the Built Environment, Privacy, Personal Space and Territory in Architecture" suggests that in order to achieve desired behavior, environments should be designed with intent. For example, reconfiguring the seating in an office, can improve the relationship between workers. Applying these theories to aircraft automation gives new perspective to cockpit design. How might we improve interaction between the pilot and the automation by making minor adjustments in the physical space? To address this question, I first needed to determine which desired, or undesired behaviors, should be acquired or eliminated. The theoretical and field research pointed out different effects of automation that require enhanced involvement and communication both with the automated systems and between crew members. Although the center of operation in the cockpit has shifted because of automation, the physical layout and pilots' seating configuration have not changed since World War II (Fig. 6 + 9).

The current seating configuration of each pilot is facing forward, in line with the primary controls and the direction of the flight. This position allows the pilot comfortably to fly the airplane while manually controlling the airplane. As discussed in the previous finding, the majority of actions performed by the pilot are operating and monitoring the automatic systems, as opposed to manually flying the airplane. Most automated systems in the cockpit are placed between the pilots, in the center console, as well as on the autopilot panel. Notwithstanding the fundamental changes in the role of the pilot as a result of increased automation, the seating configuration remains straightforward throughout the entire flight and ignores the changes to the location of critical instruments in the cockpit. This current seating configuration does not encourage collaboration between crew members, nor does it facilitate the pilots' ability to cross verify and monitor data input (excluding the primary automated flight data, which is currently located on the external displays).

Fig. 9. B-24 cockpit. USAF Museum Photo Archives.
"Periodically, I find myself performing an entire approach without even looking once at my co-pilot’s displays and data, as if we are on two separate flights" (D, Commercial airline pilot).

I suggest a new form of seating configuration for pilots, by allowing the pilot to adjust the angle of the seat slightly inward when not manually flying the aircraft. Adjusting the chairs inward toward the fellow crewmember and the center console will result in better interaction both between the crewmembers and with the automation (see Fig. 10+11).

Fig. 10. Illustration of Suggested dynamic seating configuration in automatic flight mode. All rights reserved, Avner Bendheim. Illustration by Igor Levinsky.

Fig. 11. Illustration of crew collaboration with Suggested dynamic seating configuration in automatic flight mode. All rights reserved, Avner Bendheim. Illustration by Igor Levinsky.
4.3 Keeping pace with changes in human behavior due to technology

"The constant exposure to advanced technology is changing the way we remember tasks, search for data and navigate, just some examples of behaviors and abilities that are affected by technology" (Dr. G, former director of the neuropsychology unit at an Israeli hospital).

As human behavior is changing rapidly due to technology, two trends stand out that impact pilot behavior in the cockpit. The first is our dependence on search engines to give us information instantly. With Google search, we can find everything we want to know on any given subject. This behavioral change is very evident in the cockpit. While observing a flight on a major Israeli airline, I witnessed this phenomenon firsthand. The pilots were required to perform a holding pattern (360-degree turn) at a waypoint called "ADLOD." The Air Traffic Controller (ATC) instructed the pilots to perform the hold precisely "as published" (meaning, once they reach the waypoint "ADLOD," they must turn in a certain direction and altitude). In this case, the holding pattern at "ADLOD" required a left-hand turn, because of the proximity to a restricted area (see Fig. 12).

![Published holding pattern](image)

**Fig. 12.** The published holding at "ADLOD." Jeppesen international, 2018.

The crew entered the waypoint into the automated systems, but could not find the published holding pattern loaded on the airline's iPad. Before searching in the flight manual, the crew gave up and started to execute a right-hand turn. Luckily, the ATC at the last minute allowed the pilots to proceed on their original flight route. In this situation, the pilots gave up very quickly when the answers were not at their fingertips. The second behavioral trend is our dependence on reminders to guide us through the day. We use alarms and notifications to remind us to attend meetings, pay bills, and call our grandmothers. Because we rely so heavily on reminders, we often forget to do things without them. Our dependence on these kinds of reminders is also evident in the cockpit.
"Pilot Error" could mean the lack of doing. Many mistakes happen because we forget to perform an action that is not part of our standard procedures" (Y, Commercial airline pilot). I witnessed this trend first-hand, as well. On one of my flight observations, during the pre-flight briefing, the pilots were informed of new conditions that required them to alter the flight data prior to take-off. In order not to forget to update the systems, the pilots inserted notes in a small slit between the throttles (see Fig. 13).

![Image of aircraft cockpit with notes]

Fig. 13. The crew leaves notes to remember to input data prior to take-off. As sighted during research observation.

Both issues mentioned above require a remedy. Since we cannot control the changes in societal behavior, we must make changes in the cockpit to accommodate these new trends. Unfortunately, updating aircraft systems is a lengthy and costly process (Casner & Hutchens, 2016). These challenges demand creative ideas to keep pace with the changes in human behavior without depending on the existing aircraft systems. To pinpoint when the issues identified above most often occur during flight, I conducted a survey of 35 airline pilots. The pilot's "experience map" (see Fig. 14), demonstrates when the difference between workload and automation interaction is greatest.

a) During the briefing prior to ground operation, the workload could significantly increase due to unplanned or last-minute changes to be performed during the flight. Meanwhile, there are no automatic systems to assist with the input until the crew get in the cockpit.

b) When unexpected changes occur, pilot workload increases accordingly. Unfortunately, the vast amount of data inputting can interrupt the crew during a time of high workload.
4.4 Pilots as Agents of Change

Many driving forces effect change in the cockpit. Each interested party has its own agenda, sometimes conflicting, such as efficiency versus safety, or competition versus regulation. The inner circle of influencers are the aircraft manufacturers and regulators. Due to an extremely high barrier entry, there are very few commercial aircraft manufacturers and outfitting companies, resulting in a rather conservative development process. Major regulatory agencies, such as the Federal Aviation Administration (FAA), limit changes in aircraft systems and procedures. The outer circle of influencers varies in type from the political arena, insurance and oil companies, and the airlines themselves. Conflicting agendas from either circle are constantly pulling and influencing multiple industries. For example, Boeing's agreement to sell planes to Iran may influence the United States position regarding the Iranian nuclear deal, and vice versa (Holzer, 2017). Key industry drivers, such as global industry growth and pilot shortage, also influence cockpit design and development considerations. Airline pilots used to have had limited input on cockpit design and development (Singer, 2002). Further, regulatory rules were imposed on pilots, and it was a one-sided relationship.

According to research findings, there is a strong indication that over time, airline pilots have become more empowered, influencing regulation and system development, due to the following factors:

1. The internet has made communications between pilots much easier, with the result that they have been able to share knowledge in a broad sense. That, in turn, has allowed pilots to speak as one voice when dealing with regulatory agencies, and has made it impossible for those agencies to ignore what the pilots have to say.

2. Human-centered design has been increasingly accepted as the best approach for the safe use of complex systems. By definition, that approach has elevated the importance of input from end-users such as pilots in the process of cockpit and system design.

3. Secondary effects of the worldwide pilot shortage, such as intense competition over pilot employment, have increased pilots' bargaining position with airlines and regulatory agencies.

These factors have turned pilots into agents of change, with the ability to influence multiple interested parties in the industry. An example of such an influence is the pilot "experience map", correlating between pilot workload and automation interaction.
authorization for pilots to use iPad's in the cockpit. In the private pilot sector, iPad applications have been used for several years as aviation tools for navigation charts, weather maps and manuals. Despite that trend, until recently, regulatory agencies, aircraft system manufacturers and commercial mapping services have been reluctant to introduce the iPad to commercial aviation. Growing pressure from pilots has not only contributed to the authorization to use iPads, but has also created a new market for the development of flight apps to be used by commercial pilots (see Fig. 15). This is a major development because now pilots can get things they want relatively quickly and not have to wait for systemic change.

![Figure 15. "ForeFlight." a flight application that was purchased by Boeing after being widely used among pilots. Foreflight.com. 2018.](image)

Referring to previous findings, assistive applications could help address pilot behavioral shortcomings, such as needing alerts to remind them to complete tasks, or consolidating information into a single search engine. Of all the forces effecting change in the cockpit, pilots have become very powerful influencers. The change is being driven both top-down and bottom-up. From the top, interested parties such as aircraft manufacturers, airlines and regulators are responding to pilot demands. From the bottom, pilots input has created opportunities to develop new products without the delays inherent in system-wide change.

5 Conclusion

From the introduction of basic automation, the nature of automation has become both more ambitious and complex. With the growth of automated systems, it has become increasingly clear to researchers and designers that the interplay between those systems and human activity must be the focus of their attention. This is especially true in complex settings where the safety of others is at stake.

With that backdrop in mind, the research studied the interaction between commercial airline pilots and automated systems found in most cockpits. As a result of the introduction of those new systems, the reality is that pilots spend
more of their time monitoring or operating the automated systems rather than actually flying their airplane. My goal was to develop a solution to some of the documented negative impact automation has had on pilot performance. Those include pilot complacency and over-reliance on automation. I concluded that creating a physical environment that encourages more collaboration in the cockpit will help address those concerns. Specifically, allowing the pilots’ seats to pivot towards each other will enable each pilot to better monitor the activities of their co-pilot and at the same time help each pilot focus on the automated systems.

In addition, I identified the impact that exposure to technology has on human behavior. I then demonstrated that given the amount of lag time between the design and implementation of cockpit systems, the new systems do not always address the characteristics of the pilots who will be expected to work with those systems. Finally, I have explained why pilots today increasingly act as agents of change and how in that capacity, they are in a position to facilitate solutions without the need for industry-wide change.

References


Mental workload measurement, The case of stock market traders

Héla MOURALI¹ (Corresponding author) and Lassaad LAKHAL²
¹High School of Technological studies of Sousse, TUNISIA University of Sousse
hela.mourali.khaldi@gmail.com
²University of Sousse, Faculty of Economic Sciences and Management of Sousse
lassaad.lakhal55@gmail.com

Abstract. The aim of this study is to focus on generating factors of mental workload for the case of stockbrokers. Based on a survey of a sample of 77 Tunisian stockbrokers, we found that mental workload varies according to factors related to performed task demands. As a multidimensional phenomenon, mental workload can be intensified by other psychosocial and individual factors. Even though we have tested all three groups of factors as direct antecedents to mental workload despite the assumption that psychosocial and individual factors could be operationalised as moderators (Hancock and Parasuraman, 2001, Van Acker et al., 2018). We pointed out that in a global model only the task demands were significantly related to mental workload level. As we hypothesised emotional demands and emotional dissonance could be dimensions of task demands, we also found that emotional demands are correlated to mental workload level.

Keywords: Mental workload, task demands, emotional demands, stockbrokers.

1 Introduction

New studies are focusing on work conditions regarding the professionals of financial services and pointing out that they became harder under the effect of technological and economic forces of globalization (Blair-Loy and Jacobs, 2003). The job of stockbroker is not far away of this reality. Even though, we often call the stockbrokers as golden boys reputed to have the ability to make money for themselves and for their customers, we unfortunately forget that making money goes through making good decisions in a volatile, complex and risky market. Moreover, the daily pressures experienced by stockbrokers in terms of too much...
information flows, strong effort of concentration, memorization, ingenuity to capture all instantly submerging opportunities, faster work pace and a more and more long working day, may expose them to several risks: suicide, turnover, early retirement (Wheat, 2005; Polanco-Roman et al. 2010). Those could also be likely causes of a mental overload affecting their job performance.

The aim of this paper is to detect stockbroker’s exposure to mental workload and to identify its origins. Second section will provide the relative background of our study dedicated to the job of stockbroker. A review of literature dedicated to mental workload determinants will be implemented in section three. Section four will present the proposed global model of mental workload antecedents for the case of stockbrokers as well as the partial model that tests the effects of the supposed components task demands. Section five will present results and section six will conclude.

2 Background

2.1 Mental workload

An overview of the mental workload’s current state of literature, as presented by Young et al. (2014), displays various definitions. Although the change of variables used in the conceptualizations based on different fields of application, it is assumed that mental workload has objectively to do with a combination of resource levels related to the task as well as to the operator (Emerson and MacKay, 2006). In fact, mental workload emanates from a disequilibrium in the couple task demands and individual resources (Valdehita et al., 2012; Orlandi and Brooks, 2018). So, it occurs when the amount of resources available to an operator does not fit with the amount of resources needed by the task (Loft et al., 2007). According to Luximon and Goonetilleke (2001) definition, mental workload is the subject’s direct estimate of mental or cognitive effort experienced at a given point in time. As a subjective phenomenon, it depends not only on task specificities but also on operator capabilities, characteristics as well as emotional state at a given moment (Brookhuis et al, 2009; Obinata et al., 2009; Pereira da Silva, 2014). The mental workload level can be adjusted by available resources within the person through his individual strategies and characteristics (Claxton Bommer and Fendley, 2016; Parasuraman and Hancock, 2001; DeWaard, 1996). The operator will then perform at different levels of mental workload that would be optimal or suboptimal (Van Acker et al., 2018; Young et al., 2014) influencing the effectiveness of his task performance (Wickens and Tsang, 2014; DeWaard, 1996) and even resulting in other outcomes such as
fatigue, errors, risky behaviors (Brookhuis and DeWaard, 2000; Mathiews, 2016; Young et al., 2014; Wickens, 2017). Since 1970, there has been a large body of literature dedicated to mental workload assessment which gave rise to multiple validations and applications of the three main known assessment techniques of mental workload: the simple and second task performance, subjective scales and physical measures. However, there is much less papers dealing with the modelling of the mental workload concept in relation with its consequences in terms of decision making, behavioral choices and more generally performance (Wickens, 2017). Yet, working on mental workload models could result in practical managerial implications and thus in a great economic interest consisting on cost and loss of performance avoidance (Pereira da Silva, 2014). According to Parasunaman et al. (2008) and Yorko et al (2010), mental workload is considered unanimously as an important determinant of human performance notably in high risk and high-level reliability environments. Stockbroker’s job meets indeed these two characteristics because of nowadays high market volatility and the hard responsibility at both moral and financial levels.

2.2 Stockbroker’s job load

Technically speaking, stockbroking is a typically high-skilled job that requires the mastering of knowledge and abilities in financial trading and customers relationship management to perform well. Furthermore, this job imposes a code of deontology that places the stockbroker in front of various obligations essentially of ethical type, giving his job a moral, and to certain extent, emotional dimension (S.E.C, Duties of Brokers, Dealers, and Investment Advisers, 2013).

The last few decades have been characterized by a wave of deep transformations in the job of stockbroker. It has been expected that those transformations would simplify his daily work. However, it seems that ITC and market liberalization have more and more complicated his function (Cerny, 1994; Blair-Loy and Jacobs, 2003). Not only this job requires new skills in order to deal with a larger stock market, but also it constrains the stockbroker to manage enormous information flows as well as a greater number of customers and market transactions which are nowadays carried out round the clock (Blair-Loy and Jacobs, 2003). It is therefore assumed that stockbrokers are appealed to show more mental effort in terms of data analyzing, soft skills, risk management and self-control strategies. He has to be able to take the right decision at an elementary lap of time, to control his own and that of the customers’ stress and emotions, to assume the consequences of his choice and to meet his hierarchy
expectations and set-objectives. He would appear to face considerable level of mental workload which is expected to affect individual and even team performance.

3. Leading factors of mental workload: a review of literature

3.1 Task demands

The various researches conducted on mental workload define the task demands through a number of characteristic requirements inherent in processing the task (Van Acker et al., 2018). The most frequently studied are related to the cognitive aspect of mental workload. Hence, cognitive demands are a set of stimuli that force employees to spend cognitive resources (Van Acker et al. 2018). They involve task complexity, task switching and the perceptual format of the instructions (Wickens, 2002; 2008; 2017; Layer et al., 2009; Proctor and Vu, 2012). Some other work conditions could be included depending on the nature of the task such as the work pace, the time pressure, the information overload; all that Lahlou (2000) qualifies as antecedents of the Cognitive overflow syndrome. This definition of task demands meets perfectly with the job description of a stockbroker’s typical work day. Anecdotally, it is not as glamorous as it sounds like to be. Being a stockbroker is a very demanding job as he is not supposed to leave his office until he is sure he made his day as productive as it has to be. In fact, most of a broker’s day is likely to be juggling following price evolution on different screens, building and acting selling or buying strategies, calculating costs and earnings, searching and analyzing market’s information, talking on the phone, writing emails, sending messages and pitching stock ideas to customers or explaining to them financial concepts that are often hard to understand; that is for task switching. As far as instruction format perception is concerned, visual channel as well as digital support are the most solicited and used resources. The stockbroker is not supposed to lose sight of his computer screens for any reason during the longer and longer trading sessions. The interfaces display different kinds of information with special colorful animated features, dynamic graphs and some alphanumerical codes that change instantly according to the evolution of the market transactions. Then, Michon (1985) maintains that the deepness of mental workload comes from the degree of task complexity that they classify into three different levels; strategic, intermediate also called tactical and operational or control. The first level, here, can meet with the main processing content of the task from information analysis to strategic decisions such as portfolio structure or the amount of investment. The second level, that is the middle one of complexity, concerns the extended decision autonomy. Actually,
the stockbroker makes his own decisions for the major parts of his job. He is rarely helped by his hierarchy because he is supposed to be enough skilled to perform in an autonomous way. The third level is the checking one. It is the lowest level of decision where the stockbroker is asked to control all his job requirements, all different rules and all used software platforms.

As Information and Communications Technologies can be associated with work intensity in terms of faster work, more interruptions and increased multitasking (Chesley, 2014; Stadin et al., 2016), they seem to have a real impact in complexifying the work task of the stockbroker since they multiply the different dimensions of the job cognitive demands. Finally, we suppose that even if they are less studied than cognitive demands (Van Acker et al., 2018), emotional requirements can nevertheless be determining factors of the mental workload. Yet some researchers consider that they do not represent it because they are not part of the specified cognitive load demands related (Van Acker et al., 2018). However, the fact that they draw on the same definite set of the individual resources allows us to hypothesize that the emotional demands and emotional dissonance inherent in the work (Tschan et al., 2005) of the stockbroker are part of task demands. Accordingly, they can explain the level of mental load experienced. We dare an analogy with the work done on the relationship between emotional work as defined by Horchschild (1983) and work stress and other work-related concepts (Wharton et al., 1993; Van Deweerdt, 2002). Thus, we suppose that the dissonance between the emotions felt and the emotions expressed by the stockbroker toward his clients, his hierarchy and his colleagues, associated with the frequency and duration of the emotional demands of work determine the level of mental charge.

3.2 Psychosocial or organizational related factors

As defined by Cox and Griffith (1996), psychosocial related factors are the set of the operational and management aspects of work as well as their social and organizational context that can potentially induce mental strain. They include role conflict, social support, autonomy, risk and responsibility environment. The role of psychosocial factors is fundamental in the operationalization of the stockbroker’s mental workload as these factors are characterizing his job environment and at the same time are part of his job definition. So that it is often delicate to decide whether they must be regarded more as task demands; and then positioned as direct antecedent of mental workload or be just considered as moderators in the transactional relation between task requirements and the mental workload experienced so that they reinforce or diminish resource gap (Karazek, 1979; Fox et al., 1993; Young and Stanton, 2001; Beehr et al.,
In actual facts, a stockbroker is charged to manage customers’ financial portfolios. He is called to act in the interest of his employer, his customers and of course his own interest, at a time. This kind of confused contract may expose the stockbroker to a risk of psychological and social pressures due to the possible arbitrage he might be appealed to do against those three sometimes conflictual functions of interests. Moreover, he is expected to demonstrate a certain degree of skill that is often identical to taking risks in his business, in order to achieve some return and profit objectives. But at the same time, he must be aware of the great responsibility he bears for his clients and also for his employer. All this without always being able to count on the social support neither of the hierarchy nor of colleagues who are often more in a relationship of competition than support.

In a more general field of research on the impact of organizational factors, Parkes (1999) has shown that competitive work has led to the increase of both suicides and the appearance of new professional mental troubles. In this vein, let’s talk about the consequences of this state of facts on physical and mental health in terms of fatigue and / or other risky behaviors. Several studies have been carried in order to identify the main consequences of the mental workload. They pointed out mental fatigue, stress or burnout and absenteeism (Hart and Staveland, 1988; Hancock, 1996, Mathews et al., 2013; Young et al., 2014). We assume that the job of stockbroker is not away from these risks leading some stockbrokers to create many foundations and suicide prevention centres. Besides, the recent financial and economic crises that have followed each other in recent years have increased the questioning of both laymen and specialists, particularly in sociology, about the link between these crises and suicides and other mental illnesses affecting the upper professional categories. A study conducted by Coope et al. (2015) has demonstrated that a larger proportion of those whose suicides were recession-related came from the highest socio-economic groups of managers, senior managers and professional occupations.

3.3 Individual factors

There are certain individual dimensions or characteristics particular to each person that characterize his personality and which consequently determine his way of perceiving his environment and the events that take place around him. Several studies have indicated the existence of a relationship between what are called the individual factors on the one hand, and the perception of the mental workload experienced as well as the psychological consequences it implies (stress, anxiety, mental fatigue ...), on the other hand. Indeed, Hancock and Parasuraman (2001) argue that even though the mental workload is mainly induced by the task demands imposed on the human operator, it does not fully
determine it because we must take into account the individual response to that load, his skill level, his strategy to manage the task and finally some of his personal characteristics. We also must note, at this level of operationalization, that individual factors could be either considerate moderating or mediating more than antecedents in much literature review. They are more likely to intensify or deteriorate the amount of effort an individual spends on responding to task demands and thereby affecting the mental workload (MacDonald, 2006). Two characteristics seem to us particularly interesting: the coping strategies and the locus of control also called "fatalism" (Lazarus and Folkman, 1984; Parry, 1990; Lazarus, 1993; Lu et al., 2000; NG et al., 2006; Van Acker et al., 2018).

Lazarus (1993) defines coping strategies as the set of processes that an individual interposes between himself and the event perceived as threatening to control, tolerate or diminish his impact on his well-being. The coping strategy is therefore a cognitive, emotional and behavioral response designed to manage external demands that, during the evaluation phase, were judged to exceed personal resources. Lazarus and Folkman (1984) argue that coping has a dual function: first to master or modify the problem that causes the mental load in the environment through the acquisition of information, know-how or assistance needed, i.e. situation-oriented coping, then to regulate the emotional response which is a palliative response. In fact, the individual tries to avoid negative emotions by distracting or suppressing these emotions. That is called emotional coping. Locus of Control is defined as the extent to which individuals consider themselves as controlling and responsible for the events (desirable or undesirable / happy or unhappy) that they experience (Rotter, 1966). It illustrates the tendency to attribute the experienced events to causes or sources that depend on, or are independent of, their control and their will. Thus, the locus of control was the psychological concept most often studied in the field of research on mental load and stress in the workplace (Lu, 1999, Lu et al., 2000; NG et al., 2006). It is supposed that the individual's perception of this degree of control over the situation influences his willingness to control or change demands rather than endure them. So that, locus of control plays an important role in how individuals cope with workload (Lu et al., 2000; NG et al., 2006).

4 Research methodology

The aim of this study is to detect the degree of exposure of a stockbroker to mental workload and to identify its origins in terms of Task demands, individual factors and psychosocial factors. There are several approaches to measure mental workload: self-assessment or subjective Rating scales, Performance measures (including subdivisions of primary and secondary task measures) and,
Psychophysiological measures (see Eggemeier, Wilson et al. 1991, p. 207 and Cain B (2007)).

The first technique is based on a measurement scale leading to the calculation of a mental workload level score during an employee working session. The second approach apprehends mental workload through individual performance level according to the degree of task complexity. The third one associates the individual effort level to body response measured by heart rate variability and blood pressure, eye movement and eye blink, inter-beat heart rate interval. The observation of these different approaches of mental workload measurement shows that their implementation requires a number of prerequisites. Indeed, each approach has its advantages and disadvantages, and it is often recommended that we use at least two of them to carry out the same study. Nevertheless, the high cost of implementing them, forces researchers to opt for a single approach. According to Miller (2001), the chosen approach should show a certain level of awareness in order to be able to reach discriminating measurement criteria. According to DeWaard (1996), Derrick (1988), Wierwille (1984), susceptibility reflects the ability of a technique to detect changes in the level of mental workload, due to a change in the amount of work or of one of the determinants of the mental workload. In addition, Crabtree (1984), Kantowitz (1992), Muckler and Seven (1992), maintain that validity is an important element for measuring mental workload. Validity is verified when the current measurement parameter can be used to measure mental workload in the future. Similarly, taking into account the cost of implementation and the duration of study is important for the choice of the method to adopt.

In our study, we adopt the first approach based on subjective rating scale. To answer the aforementioned questions, we decided to carry out a survey assigned to a sample of stockbrokers. This study aims to detect a possible differentiation in terms of constraints or in mental workload intensity. Indeed, some studies have been carried out in order to measure work arduousness consecutive to job category and economic sector. In France, Bué et al. (1999) have studied the origins of mental workload between 1991 and 1998. They found that several factors influence mental workload level such as responsibility, sense of urgency, noise, attention, time pressure, tension with supervisor, colleagues and customers. They have concluded that one in three French employees perceives arduousness in their job.
In the present work, we attend to study the case of stockbrokers in order to apprehend the different origins of mental workload compared to other jobs such as car and train drivers or aircraft pilots. Our study has been carried out based on a sample of Tunisian stockbrokers. The sample represents about 60% of all parent population. In fact, the Tunisian Stock exchange includes 24 stock market intermediaries employing five to six stockbrokers. We didn’t find in the current literature a study dedicated to stockbrokers’ context. Hence, we have followed the Churchill paradigm to better fit to the research context (see figure 1)
A triangulation of three techniques was adopted; observation, interviews and questionnaires. In a first step we started by observing stockbrokers during a daily working session to better apprehend job characteristics and processing. In a second step, we have carried out a number of interviews in order to identify exact items of mental workload for this kind of job, particularly those that cannot be directly observed. Data saturation has been reached after six interviews. Finally, a questionnaire gathered all identified variables operationalized by various scales such as those of Karasek (1979), Moos (1981), Rizzo, House et Lirtzman, (1970), Bué et al. (1999), Pugliesi K (1999), Schwartz et Stone (1993) Rotter (1966), Hart et Staveland (1998) and Nasa-TLX scale (see appendices 1).

5 Model and results

We consider that mental workload could be measured as follows:

\[
MWL = C + \alpha \text{IND} + \beta \text{PSYS} + \gamma \text{TASKD} + \varepsilon
\]  

(1)

Where,

MWL is mental workload;
IND is individual factor of mental workload;
PSYS is psychosocial factor;
TASKD is task demands factor;
\(\alpha, \beta, \gamma\) are model parameters that we have to estimate, \(\varepsilon\), is the residual term of the equation.

This model aims to test the following hypothesis:

H1: Mental workload depends essentially on individual factors;
H2: Mental workload depends essentially on psychosocial factors;
H3: Mental workload depends essentially on task demand factors.

To test those hypothesis, we have to operationalize all sub-mentioned variables.

5.1 Data preparation

This first step will allow us to proceed with a transformation of all ordinal variables into continuous variables. This transformation will allow us to widen
the range of analysis towards more econometric and analytical methods. On the other hand, passing through the factor analysis method will allow us to reduce the number of study variables from 191 items to only about ten variables. After studying the reliability analysis and after purification of the scales (see appendices 1), it will be possible to regroup the representative variables in a single one-dimensional factorial axis. Factorization will be based on the Principal Component Analysis (PCA) method. The idea of this test is to select the measures used in order to retain only a sample of items that is representative of the studied construct.

5.1.1 Mental workload variable

Several studies have focused on mental workload measurement without being able to describe a sufficiently rational approach to this measure. Though the critics addressed to the subjective approach (Rubio et al., 2004), we have chosen it to measure mental workload for the case of stockbrokers for at least two reasons: first, clinical approach is very difficult to be carried out for the case of stockbroker. Second, stockbrokers are bound by the professional secrecy, so it is impossible to accompany them during their work. Regarding all those restrictions, we have opted to the survey approach. In order to get a significant mental workload level, we have required that individual answers must be provided during a trading session (to test second task effect) and we have left the items measuring the level of mental workload to the end of the survey in order to ensure that it has reached a significant one.

Table 1 provides the different items used to measure mental workload for the case of stockbrokers.
Table 1. Nasa Task Load Index (Adjusted)

<table>
<thead>
<tr>
<th>Item</th>
<th>Libellee (1 very low ------------------------------------------- 5 very high</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASACPT1</td>
<td>You consider that the level of efficiency in the accomplishment of your activities is:</td>
</tr>
<tr>
<td>NASAEMO1</td>
<td>You feel that the level of insecurity, irritation, discouragement is:</td>
</tr>
<tr>
<td>NASAIN1</td>
<td>How much did you have to work hard to achieve these performances?</td>
</tr>
<tr>
<td>NASAMENT1</td>
<td>You fill that the level of your mental and perceptual activities (Thinking, deciding, calculating, remembering, searching...) is</td>
</tr>
<tr>
<td>NASAPHY1</td>
<td>You consider that the level of physical requirement of your activity is:</td>
</tr>
<tr>
<td>NASATEM1</td>
<td>You feel that the pressure related to the pace of events that occur in the course of your activity is:</td>
</tr>
</tbody>
</table>

Based on the subjective measurement of mental workload, we have calculated a $z$-score as follows:

$$ Z \text{ score} = \frac{x - \mu}{\sigma} \quad (2) $$

In order to fix extreme values of $z$-score, we have first to verify normality conditions. Results are reported in the following table:

Table 2. Descriptive statistics of $z$-score

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>St-deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z-score Individual</td>
<td>77</td>
<td>0,01</td>
<td>0,999</td>
<td>-0,062</td>
<td>0,074</td>
</tr>
<tr>
<td>Individual</td>
<td>77</td>
<td>39,00</td>
<td>22,372</td>
<td>0,000</td>
<td>-1,200</td>
</tr>
<tr>
<td>N valid (listwise)</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows that $z$-score variable is normally distributed. In fact, both Skewness and Kurtosis values are close to zero. Consequently, sample individual can be separated regarding the mean value that equals zero. Then, a negative $z$-score means that the mental workload of the corresponding individual is inferior to the mean. In another word, the subject has a week level of workload. Alternatively, a positive $z$-score reflects a high level of mental workload for the corresponding individual. Appendices 2 provides the results of this test for the case of the 77 individuals of our sample. Results show that 33 within the 77 individuals of our sample have a high level of mental workload (42%). This level of mental workload can be explained by
several factors: individual, Task demand and psychosocial. To apprehend the extent of the influence of each factor, we have to provide a synthetic measurement of each factor using the Principle Component analysis technique.

5.1.2 Psychosocial factor

The first factor of mental workload is the Psychosocial factor. Five variables are provided to build this factor. Each variable is set of several items. The five variables are: autonomy, social support, role ambiguity, risk taking and accountability. The table below illustrates the main results from the principle Components Analysis.

- The axis related to autonomy is represented by two items that restore 60% of the total variance;
- The social support axis has eight items that return 62% of the total variance;
- The role ambiguity axis is based on three items that account for 47% of the overall variance;
- Risk-taking is composed of five items returning 59% of the initial information;
- The responsibility axis is based on four items that yield 58% of the total variance.

In the following, we will focus on the relative weight of each axis in the restitution of the psychosocial dimension of the mental workload.

Table 3. Extraction of Psychosocial factor

<table>
<thead>
<tr>
<th>Variable</th>
<th>KMO index</th>
<th>Bartlett test sig</th>
<th>Total Explained Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomy (6.267)</td>
<td>0.5</td>
<td>0.067</td>
<td>60.510</td>
</tr>
<tr>
<td>Social Support (30.259)</td>
<td>0.864</td>
<td>0.000</td>
<td>62.005</td>
</tr>
<tr>
<td>Roles Ambiguity (7.156)</td>
<td>0.517</td>
<td>0.008</td>
<td>47.397</td>
</tr>
<tr>
<td>Risk (15.586)</td>
<td>0.734</td>
<td>0.000</td>
<td>59.213</td>
</tr>
<tr>
<td>Responsibility (11.604)</td>
<td>0.708</td>
<td>0.000</td>
<td>58.173</td>
</tr>
<tr>
<td>Total explained variance</td>
<td>0.707</td>
<td>0.000</td>
<td>70.871</td>
</tr>
</tbody>
</table>

Principle Component Analysis

Table 3 shows that the psychosocial factors of the mental load are mainly represented by the social support, which restores 30% of the total explained variance followed by the risk taking with 15%, the responsibility 11%, the ambiguity of roles and the autonomy with respectively 6 and 7%. This result represents a specificity for the stockbroker as compared to other jobs such as train drivers, air pilots or astronauts for whom the relative importance is somehow different.
Thus, for the case of stockbroker, three main closely related factors characterize the profession. In the first place, the social support represented mainly by the support of the chiefs and the colleagues is the fundamental element for the management of the mental workload. Moreover, a good working atmosphere reassures the stockbroker and leads him to allow himself a higher risk taking because in case of problems, he knows that he will be supported by his superiors and by his colleagues. Thus, the brake of responsibility is somewhat attenuated because of a high level of confidence of the stockbroker in the acts that he undertakes. As for the ambiguity of roles and autonomy, we note that their importance is low; this result is not surprising given the specificities of stockbroker’s business. In fact, stockbroker operates in an atmosphere of total autonomy (since he is rated according to the completion of his investment strategies. Ambiguity is not problematic because stockbroker has to follow and to respect provided instructions. Moreover, stockbroker operates through a common interface to all stockbrokers on the stock exchange. The principles and rules of operation are also standardized, which excludes any form of ambiguity in the performance of assigned tasks.

5.1.3 Task demands factor

The second test aims to extract the determinants related to task demands. Nine determinants were initially introduced: emotional demands, work cadence, sensory-type demands, versatility, multi-skills, information overload, emotional dissonance, urgency and zapping.

<table>
<thead>
<tr>
<th>Variable</th>
<th>KMO index</th>
<th>Bartlett test sig</th>
<th>Total Explained Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional demands (17.654)</td>
<td>0.760</td>
<td>0.000</td>
<td>64.588</td>
</tr>
<tr>
<td>work cadence (21.706)</td>
<td>0.753</td>
<td>0.000</td>
<td>54.943</td>
</tr>
<tr>
<td>sensory-type demands (4.215)</td>
<td>0.790</td>
<td>0.000</td>
<td>52.094</td>
</tr>
<tr>
<td>versatility (0.000)</td>
<td>0.5</td>
<td>0.005</td>
<td>65.857</td>
</tr>
<tr>
<td>multi-skills (5.615)</td>
<td>0.748</td>
<td>0.000</td>
<td>55.857</td>
</tr>
<tr>
<td>information overload (4.004)</td>
<td>0.508</td>
<td>0.000</td>
<td>54.528</td>
</tr>
<tr>
<td>emotional dissonance (8.667)</td>
<td>0.600</td>
<td>0.000</td>
<td>65.665</td>
</tr>
<tr>
<td>Urgency (5.225)</td>
<td>0.500</td>
<td>0.001</td>
<td>68.110</td>
</tr>
<tr>
<td>Zapping (0.000)</td>
<td>0.500</td>
<td>0.003</td>
<td>66.498</td>
</tr>
</tbody>
</table>

Principle Component Analysis
Table 4 shows that Task Demands factor is based on Work Cadence (21.7% of the total Variance) followed by emotional demands (17.65%), Then emotional dissonance (8.66%), Multi-skills (5.61%), urgency (5.22%), Sensory-type demands and Information overload represent each one 4% of the total variance. Those results are in accordance with our expectations in the sense that most of the stockbroker job is essentially based on a well-defined cadence, focused on automation, synchronization, teamwork, the ritual of each day (same session transaction, same time, same duration, same preparations ...). During the administration of the questionnaire, several interviewees told us of the crucial importance of respecting this pace to succeed as a stockbroker. They admit that in practice, this pace is the main source of fatigue, stress or even mental workload.

The second variable is emotional requirements or emotional control, this variable represents 17.6% of the total variance. In fact, the stockbroker is highly exposed to this source of burden regarding the nature of his work, which consists in managing the wealth of others and exposing them in certain cases to significant risks that may lead to the destruction of their assets. Similarly, during the exercise of his work, the stockbroker is forced to dissociate his private life from his professional activity. Horcschild (1983) considers that this emotional exercise requires an effort of planning and self-control for successful interpersonal transactions (with superiors, colleagues, and customers) ... all of this contributes to generate a mental burden that makes the work more and more arduous. In addition to emotional demands, the stockbroker should deal with emotional dissonance. This phenomenon consists of replacing certain negative emotions with their exact opposites; it is a good reason to replace fear with courage, pessimism with optimism, panic with calm. It is also to broadcast positive waves to customers especially during difficult times. It is a question of knowing how to control one's most unpleasant emotions and to replace them by signs of respect, alignment, and even of submission. Emotional dissonance accounted for 8% of the total variance task demands factor. Emotional dissonance is among the most significant resources of the mental workload. This variable is very decisive for the continuity of the exercise of the stockbroker job. We deduced this from the interviews that preceded the elaboration of the questionnaire where several brokers told us of their intention to leave the profession because they do not manage their emotions especially during crises and prices collapses.
5.1.4 Individual factor of mental workload

Individual determinants are divided into two categories: determinants based on coping strategies and determinants based on locus of control. The literature distinguishes three forms of strategies based on coping:

- Situation-based coping (43% of the overall variance).
- Emotion-focused coping represents 14.88% of coping information.
- Perception-based coping is the third most important determinant of coping. It makes it possible to restore 13.763% of the total variance.

<table>
<thead>
<tr>
<th>Variable</th>
<th>KMO index</th>
<th>Bartlett test sig</th>
<th>Total Explained Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation-based coping</td>
<td>0.793</td>
<td>0.000</td>
<td>57.899</td>
</tr>
<tr>
<td>(43.894)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotion-focused coping</td>
<td>0.656</td>
<td>0.000</td>
<td>51.918</td>
</tr>
<tr>
<td>(14.880)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perception-based coping</td>
<td>0.815</td>
<td>0.000</td>
<td>71.132</td>
</tr>
<tr>
<td>(13.763)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total explained variance</td>
<td>0.809</td>
<td>0.000</td>
<td>72.537</td>
</tr>
</tbody>
</table>

Principle Component Analysis

Regarding the locus of control, we distinguish mainly two profiles of locus: the internal locus and the external locus. Results of Table 6 shows that external locus of control dominates Internal locus of control. This result seems somewhat surprising since locus of control should be based on internal loci characterized by rationality, self-confidence and realism.

<table>
<thead>
<tr>
<th>Variable</th>
<th>KMO index</th>
<th>Bartlett test sig</th>
<th>Total Explained Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Locus of control</td>
<td>0.614</td>
<td>0.002</td>
<td>51.348</td>
</tr>
<tr>
<td>(22.913)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>External Locus of control</td>
<td>0.667</td>
<td>0.000</td>
<td>55.893</td>
</tr>
<tr>
<td>(34.519)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total explained variance</td>
<td>0.608</td>
<td>0.000</td>
<td>57.432</td>
</tr>
</tbody>
</table>

Principle Component Analysis
5.2 Model test

In order to study the relative importance of each of the three factors of mental workload (Individual, Task demands and psychosocial), we regress them on the level of mental workload as measured by z-score. Results are reported in the following table.

Table 7. Regression results

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>T-statistic</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>0.429***</td>
<td>8.240</td>
<td>0.000</td>
</tr>
<tr>
<td>Psychosocial</td>
<td>0.021</td>
<td>0.178</td>
<td>0.859</td>
</tr>
<tr>
<td><strong>Task demands</strong></td>
<td><strong>0.417</strong>*</td>
<td>3.609</td>
<td>0.001</td>
</tr>
<tr>
<td>Individual</td>
<td>0.105</td>
<td>0.972</td>
<td>0.334</td>
</tr>
<tr>
<td><strong>R-square</strong></td>
<td><strong>0.194</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OLS regression
***, **, * Significant at 1, 5 and 10 percent respectively

\[ MWL = 0.429^{***} + 0.105\text{IND} + 0.021\text{PSYS} + 0.417^{***}\text{TASKD} + \epsilon \]

Results show that mental workload depends essentially on task demands factor for the case of Tunisian stockbrokers. This result can be explained by the fact that stockbrokers find more difficulties to deal with work cadence, emotional exigences and emotional dissonance. They have their appropriate strategies to control the two other factors (individual and psychosocial). To confirm our results, we have to carry out a least regression of MWL regarding variables of Task demand as expressed by the following equation:

\[ MWL = C + \alpha\text{SUPSOC} + \beta\text{RISK} + \gamma\text{RESPONS} + \delta\text{CAD} + \theta\text{EMOREQ} + \omega\text{EMODISS} + \epsilon \]

Where,
SUPSOC, reflects the social support;
RISK, is related to stockbroker’s exposure to professional risks;
RESPONS, is the level of responsibility inherent to stockbroker’s activity;
CAD, reflects job cadence;
EMOREQ, is the emotional requirements of stockbroker;
EMODISS, is the emotional dissonance;
C, is the constant reflecting other variable that are not taken into account in the present model.

\( \varepsilon \), Is the error term of the equation.

Using OLS technique, results are provided by Table 8 and the model specification can be written as follows:

\[
MWL = 0 - 0.516^{***} \text{SUPSOC} + 0.238^{**} \text{RISK} + 0.463^{**} \text{RESPONS} - 0.086 \text{EMOREQ} + 0.314^{**} \text{EMODISS} + \varepsilon
\]

**Table 8.** Regression results for Task demand

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td></td>
<td>,000</td>
<td>1,000</td>
</tr>
<tr>
<td>Social support</td>
<td>-0.516***</td>
<td>-4.169</td>
<td>0.000</td>
</tr>
<tr>
<td>Risk</td>
<td>0.238**</td>
<td>2.084</td>
<td>0.041</td>
</tr>
<tr>
<td>Responsibility</td>
<td>0.463**</td>
<td>3.355</td>
<td>0.001</td>
</tr>
<tr>
<td>Cadence</td>
<td>-0.070</td>
<td>-0.565</td>
<td>0.574</td>
</tr>
<tr>
<td>Emotional requirements</td>
<td>-0.086</td>
<td>-0.673</td>
<td>0.503</td>
</tr>
<tr>
<td>Emotional dissonance</td>
<td>-0.314**</td>
<td>-2.523</td>
<td>0.014</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td>0.361</td>
<td></td>
</tr>
</tbody>
</table>

. OLS regression

***, **, * Significant at 1, 5 and 10 percent respectively

Table 8 shows that Mental workload increases when stockbroker’s responsibility increases. In fact, the trauma of Kerviel affair, and his role in a trading scandal that cost the Société Générale Bank close to €5bn ($7bn) and that sentenced him to five years in prison, including two years suspended, and to reimburse the French bank €4.9bn lost in unauthorized trades, makes sharper the responsibility of stockbroker. We could also remember the Madoff affair who was sentenced to 150 years in prison with restitution of $170 billion. This can also explain the relationship between risk and mental workload. Besides, Emotional dissonance contributes to the decrease of mental workload that stockbrokers who are able to control his emotions toward customers, hierarchy and colleagues are the less exposed to mental workload. The same explanation could be provided to explain the relationship between mental workload and Social support so stockbrokers supported by hierarchy, colleagues and even customers are the less mental loaded.
6. Conclusion

Based on a subjective workload assessment approach, we have carried out a study in order to apprehend the specificities of the job of stockbroker. We have proved that it is indeed an arduous job that exposes stockbroker to several risks, emotional troubles and even health complexities. We have also noted a very high level of turnover (About seven years). We have also noticed among the interviewed stockbrokers a very high sentiment of insecurity. Moreover, Tunisian stockbrokers have shown a sense of irritation, frustration caused by the working cadence and the obligation of having the best performance to please both superiors and customers, especially in this context of post-revolution of 2011. Besides, several bad incidents, that the Tunisian Stock Exchange has suffered such as the brutal collapse and the suspension of trading of some market securities, have negatively affect the relationship between stockbrokers and their customers. We have noticed a feeling of regret of some stockbrokers against their past decisions especially those that have led to huge losses to customers. Those findings lead us to conclude that Tunisian stockbrokers are subject to a high level of mental workload (at least about 42% of them according to our study). This high level of mental workload is due to several factors; however, the leading one is task demands in terms of job cadence, and emotional exigences. Consequently, the H3 hypothesis has been validated (Mental workload depends essentially on task demand factors) Finally, we argue that a mental workload measurement index could be constructed around six dimensions: risk apprehension, sentiment of responsibility, time pressure, mental effort, personality and social support for the case of stockbroker’s job.

References


# Appendix

## Appendices 1: Reliability Test

<table>
<thead>
<tr>
<th>Variable</th>
<th>Libellee</th>
<th>Nbr of initial items</th>
<th>Nbr of final items</th>
<th>Cronbach Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTORG</td>
<td>Auto-regulation</td>
<td>3</td>
<td>4</td>
<td>0.658</td>
</tr>
<tr>
<td>AMBROL</td>
<td>Ambiguity of roles</td>
<td>5</td>
<td>5</td>
<td>0.762</td>
</tr>
<tr>
<td>AUTDEC</td>
<td>Autonomy</td>
<td>2</td>
<td>2</td>
<td>0.750</td>
</tr>
<tr>
<td>CONFROL</td>
<td>Role Conflict</td>
<td>5</td>
<td>2</td>
<td>0.794</td>
</tr>
<tr>
<td>COPING</td>
<td>Coping Strategy</td>
<td>16</td>
<td>10</td>
<td>0.531</td>
</tr>
<tr>
<td>EMEXATT</td>
<td>Attentional Exigences</td>
<td>8</td>
<td>7</td>
<td>0.806</td>
</tr>
<tr>
<td>EMEXCPL</td>
<td>Emotional Exigences</td>
<td>2</td>
<td>2</td>
<td>0.482</td>
</tr>
<tr>
<td>EMEXCPT</td>
<td>Mental Effort exigences (Competence)</td>
<td>9</td>
<td>9</td>
<td>0.774</td>
</tr>
<tr>
<td>EMEXDEC</td>
<td>Mental Effort exigences (decisional)</td>
<td>2</td>
<td>2</td>
<td>0.758</td>
</tr>
<tr>
<td>EMEXDIV</td>
<td>Mental Effort exigences (informal)</td>
<td>3</td>
<td>3</td>
<td>0.506</td>
</tr>
<tr>
<td>EMEXGINF</td>
<td>Mental Effort exigences (informal)</td>
<td>5</td>
<td>5</td>
<td>0.562</td>
</tr>
<tr>
<td>EMEXRE</td>
<td>Mental Effort exigences (relation)</td>
<td>3</td>
<td>3</td>
<td>0.594</td>
</tr>
<tr>
<td>EMINT</td>
<td>Mental Effort (intentional)</td>
<td>6</td>
<td>6</td>
<td>0.654</td>
</tr>
<tr>
<td>EMODISS</td>
<td>Emotional dissonance</td>
<td>6</td>
<td>6</td>
<td>0.719</td>
</tr>
<tr>
<td>EMOEXG</td>
<td>Emotional Exigence</td>
<td>6</td>
<td>6</td>
<td>0.806</td>
</tr>
<tr>
<td>EMRESATT</td>
<td>mental Effort</td>
<td>8</td>
<td>8</td>
<td>0.622</td>
</tr>
<tr>
<td>MONOT</td>
<td>Monotony</td>
<td>7</td>
<td>6</td>
<td>0.820</td>
</tr>
<tr>
<td>PTCAD</td>
<td>Cadence</td>
<td>10</td>
<td>9</td>
<td>0.713</td>
</tr>
<tr>
<td>PTURG</td>
<td>Urgency</td>
<td>4</td>
<td>4</td>
<td>0.792</td>
</tr>
<tr>
<td>PTZAP</td>
<td>Zapping</td>
<td>3</td>
<td>3</td>
<td>0.774</td>
</tr>
<tr>
<td>RESPONS</td>
<td>Responsibility</td>
<td>4</td>
<td>4</td>
<td>0.782</td>
</tr>
<tr>
<td>RISQUE</td>
<td>Risk</td>
<td>8</td>
<td>8</td>
<td>0.872</td>
</tr>
</tbody>
</table>
## Appendices 2: Z-score of mental workload

<table>
<thead>
<tr>
<th>IND</th>
<th>Z-SCORE</th>
<th>MWL(High; Low)</th>
<th>IND</th>
<th>Z-SCORE</th>
<th>MWL(High; Low)</th>
<th>IN</th>
<th>Z-SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0,897095</td>
<td>1</td>
<td>27</td>
<td>-1,639</td>
<td>0</td>
<td>53</td>
<td>1,4166</td>
</tr>
<tr>
<td>2</td>
<td>0,897095</td>
<td>1</td>
<td>28</td>
<td>-1,119</td>
<td>0</td>
<td>54</td>
<td>0,4081</td>
</tr>
<tr>
<td>3</td>
<td>-0,63090</td>
<td>0</td>
<td>29</td>
<td>-1,119</td>
<td>0</td>
<td>55</td>
<td>0,8970</td>
</tr>
<tr>
<td>4</td>
<td>0,408136</td>
<td>1</td>
<td>30</td>
<td>-0,630</td>
<td>0</td>
<td>56</td>
<td>0,8970</td>
</tr>
<tr>
<td>5</td>
<td>2,425093</td>
<td>1</td>
<td>31</td>
<td>1,4166</td>
<td>1</td>
<td>57</td>
<td>0,8970</td>
</tr>
<tr>
<td>6</td>
<td>-0,111383</td>
<td>0</td>
<td>32</td>
<td>-1,119</td>
<td>0</td>
<td>58</td>
<td>1,4166</td>
</tr>
<tr>
<td>7</td>
<td>-0,111383</td>
<td>0</td>
<td>33</td>
<td>-1,639</td>
<td>0</td>
<td>59</td>
<td>0,4081</td>
</tr>
<tr>
<td>8</td>
<td>-0,63090</td>
<td>0</td>
<td>34</td>
<td>0,8970</td>
<td>1</td>
<td>60</td>
<td>-0,1113</td>
</tr>
<tr>
<td>9</td>
<td>-0,63090</td>
<td>0</td>
<td>35</td>
<td>0,8970</td>
<td>1</td>
<td>61</td>
<td>0,4081</td>
</tr>
<tr>
<td>10</td>
<td>-2,158900</td>
<td>0</td>
<td>36</td>
<td>-0,630</td>
<td>0</td>
<td>62</td>
<td>-0,1113</td>
</tr>
<tr>
<td>11</td>
<td>-0,63090</td>
<td>0</td>
<td>37</td>
<td>0,4081</td>
<td>1</td>
<td>63</td>
<td>1,9361</td>
</tr>
<tr>
<td>12</td>
<td>-2,158900</td>
<td>0</td>
<td>38</td>
<td>2,4250</td>
<td>1</td>
<td>64</td>
<td>0,4081</td>
</tr>
<tr>
<td>13</td>
<td>-0,111383</td>
<td>0</td>
<td>39</td>
<td>-0,1113</td>
<td>0</td>
<td>65</td>
<td>0,8970</td>
</tr>
<tr>
<td>14</td>
<td>-0,111383</td>
<td>0</td>
<td>40</td>
<td>-0,1113</td>
<td>0</td>
<td>66</td>
<td>0,4081</td>
</tr>
<tr>
<td>15</td>
<td>-0,63090</td>
<td>0</td>
<td>41</td>
<td>-0,630</td>
<td>0</td>
<td>67</td>
<td>0,8970</td>
</tr>
<tr>
<td>16</td>
<td>-1,119861</td>
<td>0</td>
<td>42</td>
<td>-0,630</td>
<td>0</td>
<td>68</td>
<td>0,4081</td>
</tr>
<tr>
<td>17</td>
<td>0,408136</td>
<td>1</td>
<td>43</td>
<td>-2,1589</td>
<td>0</td>
<td>69</td>
<td>0,8970</td>
</tr>
<tr>
<td>18</td>
<td>-0,63090</td>
<td>0</td>
<td>44</td>
<td>-0,630</td>
<td>0</td>
<td>70</td>
<td>-0,1113</td>
</tr>
<tr>
<td>19</td>
<td>-0,63090</td>
<td>0</td>
<td>45</td>
<td>-2,1589</td>
<td>0</td>
<td>71</td>
<td>-0,1113</td>
</tr>
<tr>
<td>20</td>
<td>-1,119861</td>
<td>0</td>
<td>46</td>
<td>-0,1113</td>
<td>0</td>
<td>72</td>
<td>-0,630</td>
</tr>
<tr>
<td>21</td>
<td>-0,63090</td>
<td>0</td>
<td>47</td>
<td>-0,1113</td>
<td>0</td>
<td>73</td>
<td>0,8970</td>
</tr>
<tr>
<td>22</td>
<td>-0,111383</td>
<td>0</td>
<td>48</td>
<td>0,8970</td>
<td>1</td>
<td>74</td>
<td>0,8970</td>
</tr>
<tr>
<td>23</td>
<td>-0,111383</td>
<td>0</td>
<td>49</td>
<td>0,8970</td>
<td>1</td>
<td>75</td>
<td>0,4081</td>
</tr>
<tr>
<td>24</td>
<td>-0,111383</td>
<td>0</td>
<td>50</td>
<td>-0,1113</td>
<td>0</td>
<td>76</td>
<td>0,8970</td>
</tr>
<tr>
<td>25</td>
<td>-1,119861</td>
<td>0</td>
<td>51</td>
<td>0,8970</td>
<td>1</td>
<td>77</td>
<td>-0,1113</td>
</tr>
<tr>
<td>26</td>
<td>-0,63090</td>
<td>0</td>
<td>52</td>
<td>1,4166</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Eye tracking overview

Rolf Zon

1 Netherlands Aerospace Centre, Anthony Fokkerweg 2, 1059 CM Amsterdam, The Netherlands

Abstract. Eye tracking is a powerful tool to record where an operator is looking to, and therefore where s/he is paying attention to. It may reveal workload related information but in fact it provides much more information about operator state than just workload. In the current paper an overview of a number of experiments in which eye tracking played a crucial role is summarised. Further differences and commonalities about eye tracking as a research tool in different domains is discussed. Also future applications of eye tracking are presented.

Keywords: Eye tracking, attention, workload, aviation, rail, future.

1 Introduction (what is eye tracking)

This paper gives the reader an overview on the commonalities between eye tracker research in different studies. It describes eye tracking as a research tools and illustrates this with examples from different studies. Also some applications outside the research domain will be mentioned. Finally it will also shine a light on how eye tracking may be applied in the future.

Eye tracking is basically measuring eye movements [1] [2]. It can be used to assess attention, mental workload, interest, fatigue, drowsiness and arousal. As such, it is a very relevant tool for any kind of human behaviour research in a variety of domains such as aviation, road, rail, maritime, and control rooms. It can also be used to enhance human computer interaction.

Eye tracking is non-intrusive and continuous. It does not interrupt the operator during task execution. Looking at particular items in a (working) environment is indicative of where this person gives attention to. It is implicitly measuring where the person is not looking at, and therefore not paying attention to [1] [2] [3].
1.1 Background

Historically different techniques were used to measure eye scanning behaviour. Initially eye trackers were quite intrusive systems with bite bars and head fixation. These systems were used halfway the twentieth century, for example, by Yarbus [2]. From his research was learned that eye scanning behaviour is not smooth but that eyes “jump” in saccades from one fixation to another. Eye trackers were either expensive or inaccurate. Newer generation eye trackers comprised of small electromagnetic coils (contact lens like sensors) which were placed on the eye. Nowadays most often eye trackers are video based. Cameras can be mounted on the subject’s head, or they can be “wall” mounted.

1.2 Advantages and disadvantages

During the past decades these systems have become lighter and more comfortable to wear. Also the calibration procedure has become simpler and faster. These two developments made eye tracking easier to apply on location.

Head mounted has the advantage that they allow the subject to move the head freely in any direction while the eye scanning behaviour can still be measured. The disadvantage is that a contraption on which the cameras were mounted needs to be worn on the head. Sometimes a head tracker system needs to be added to measure the position and orientation of the subject’s head.

Wall mounted systems have the advantage that they are the least intrusive systems. There is nothing that the subject has to wear on the head. The disadvantage is that the person has to look in the direction of the camera from a relatively fixed position, which is not the case for head mounted systems.

As such the researcher needs, to make clear choices on what the Areas of Interest (AoI) are that s/he wants to study, and what kind of eye tracker to apply.

Since eye trackers are video based, reflections of light, for example from the sun or bright lamps, might disturb the measurements. (Sun)glasses or certain kinds of contact lenses might result in missing data. Also characteristics of the subject like droopy eye lids or long eye lashes might block the view of the camera with as result loss of data.
1.3 More than a research tool

Initially eye tracking was used as a means to study eye scanning behaviour. It was primarily a research tool that would, after an experiment, reveal the eye scanning strategy of subjects. Later with increased computer processing capacity it became possible to measure and use real time information about subjects’ eye scanning behaviour.

When the comfort to “wear” an eye tracker has increased the technique became sufficiently mature to be used as computer input device. For example, lighter head mounted equipment and availability of wall mounted systems, made that nowadays eye trackers are not intrusive at all,

1.4 Measures

The output of camera based eye tracker systems is often raw video. This initial output needs to be processed further to give meaning to the data. Due to a calibration process, that is executed prior to the data recording, the video based data can be converted into a raw scan path. A raw scan path is in fact how the eye has “scanned” a particular surface. In this path certain saccades (movements) and fixations (stops) were made. A fixation is one of the most basic events related to movements of the eye and it occurs when the eye remains still over a period of time (e.g. it is fixating on a specific point in the visual field). [4]. For example, an average reader fixates after seven to ten letters, processes the information, and then makes a saccade to a new fixation point.

From fixations data are often processed further [1] [2]. Below an overview of common ways for further processing are given.

Fixation duration. Marquart et al [5], Harris et al [1] and also Nocera [6] have established that fixations last longer with increasing mental workload.

Scan path distance. This distance is indicative for the efficiency of how information is presented and for how efficient an operator uses the available information.

Number of fixations per minute and the average fixation duration. These are indicative for the amount of time needed to process individual pieces of information. They vary with the type of information and familiarity of the subject. How long a subject should be looking at a certain point before it can be called a fixation is a long-standing discussion. Values range from 50 to 150 ms and
are domain dependent. Identifying a previously known symbol takes shorter than reading a new word.

Dwell time. NEN [7]. Dwell time is also called glance or gaze, is the time spent in one series of fixations at one particular AoI. It is the time from the first moment that one looks at an AoI until the first moment that this person looks away from this AoI. From these dwells a so called dwell map can be deduced, these dwells form the base for a heat map. This heat map (Fig. 1) is a very intuitive way to present how the gaze was distributed over an area, for example a photograph of the working position of a subject.

![Example of heat map. The colored blobs indicate where the subject has looked at, and for how long.](image)

**Fig. 1.** Example of heat map. The colored blobs indicate where the subject has looked at, and for how long.

**Transition.** This is changing the gaze from one AoI to another. Transitions can be summarised in a so called ‘transition matrix’, which provides insight into how frequently a subject switches his/her gaze between different AoIs.

**Entropy.** Entropy stems from the thermodynamics and was converted to the domain of communication [8]. In eye scanning it refers to the randomness of the transitions that are made. Basically whether the person was looking everywhere around him/her, as opposed to being focussed on a limited set of AoIs. It can be used as an indicator of mental workload [9].

**Eye blinks and blink latency.** These are not related to the location where a person looks at. Which is contrary to the measures mentioned above. They happen with a certain frequency and duration. The frequency correlates with the (visual) workload that a subject experiences. When the (visual) workload increases the eye blinks less frequently.
**PERCLOS.** This is the PERcentage of eye CLOSure (more than 80%) within a certain time frame, and is an indication of fatigue level [5].

**Pupil dilation.** This is influenced by (mental) workload [1]. When mental workload increases the pupil dilates. Unfortunately, it is also influenced by many other factors such as illumination level, stress, etc. The Index of Cognitive Activity (ICA)\(^1\) [10] is a measure where the pupil dilation is corrected for the influence of changes in illumination level, and therefore a more robust indicator of pupil diameter.

To apply such processed data [1] proposed two measurement techniques for display evaluation: (a) comparing histogram plots of dwell times for different instruments that a subject is using, and (b) measuring scan pattern changes as the difficulty of a task increases. The two analytic techniques provided converging evidence that one instrument was superior to the other. This principle is still applied in a great deal of the nowadays eye tracker based research, and also in some of the examples below.

### 1.5 Eye tracking and attention

In theory people can look in one direction and attend to another (i.e., overt vs. covert attention), but in reality covert and overt attention is often aligned [3] [11]. Therefore eye tracking is considered a valid means to study where persons are paying attention to. Eye movements are for example useful for assessing visual attention during search. A pilot can focus his / her gaze on the integrated control panel while altering the settings of the autopilot, and then focus his or her attention on another AoI, the Primary Flight Display (PFD), to verify if the changes have any effect [12]. In this example, the pilot’s attentional spotlight scans the visual field for interesting objects and once such an object is found, the eyes shift to this location. That is, after setting the autopilot the pilot focuses his / her gaze on the PFD to confirm the changes. Therefore, one can be reasonably confident that when the eyes focus on a certain location, attention is also focused at this specific location (see [11] for a discussion on this topic).

\(^1\) ICA = proprietary technology (by EyeTracking Inc.) therefore not available openly to the research community.
1.6 Eye tracking and mental workload

The relationship between eye tracking and mental workload was directly or indirectly established in numerous studies, [12], [1] and [13]. Marquart et al [5] have established that increases in workload correlate with increases in blink latency, percentage eye closure (PERCLOS), fixation duration, pupil dilation, and ICA and by decreases in blink duration and gaze variability. Further Nocera et al [6] found that the increasing eye fixation duration correlates with increasing mental workload. Given the number of eye tracking derived variables that are sensitive to mental workload, and the number of studies that can be found on this topic, the relationship is not a trivial one.

2 Current research

Eye tracking can be applied in a multitude of environments and domains. This section provides an overview of studies where eye tracking was applied. What they all have in common is that, after a certain change was introduced, they evaluate the measurable impact on the behaviour of the operator. Behaviour in this context is, for example, often operationalised as:

- distraction of the operator caused by the change;
- changes in time spent on different tasks since the change was introduced;
- changes in operator fatigue levels;
- changes in experienced workload levels influenced by the change.

In all of the cases below the emphasis is on the eye tracking and workload aspects of each study. In fact, these examples are parts of larger projects. In these projects eye tracking is always combined with several other sources of information to get a complete and accurate description of the operator behaviour. This methodology is called methodological triangulation [14].

2.1 Aviation

The Human Integration into the Lifecycle of Aviation Systems (HILAS) project focused amongst others on a “Human Factors tools” measurement battery from which different tools could be selected. Selection depended on the research question to evaluate the impact of a change (new technology, procedure or uncommon situation) that was introduced into the cockpit on the human operator. One of the tools in this battery comprised eye tracking. In the project, several scenarios were presented to pilots in flight simulators.
One scenario was about an airspeed discrepancy [15]. The speed indicators on the left and right side of the cockpit gave a different value, indicating that there was a technical problem. The eye trackers were applied to study the strategy that pilots used to identify the problem. The scanning patterns that the pilots followed were studied to see how pilots divide their attention to identify that a problem exists. Looking at the display of the other pilot to verify that both pilots have the same information is called “cross checking” and is one of those strategies.

It turned out that pilots performed cross checks, but none of the crews identified the problem at an early stage. However, once the aircraft had indicated that a problem existed, pilots started to scan directly to identify what was going on. The eye tracker provided hard evidence that the information acquisition process for both pilots had changed by the warning. Even though additional information, such as the communication within the crew was important as well to understand how the failure influenced the crew, the eye tracker output was the most objective and clear indication of the impact on the crew. Inefficient or slow information acquisition strategies may lead to increased workload later during the flight and should therefore be mitigated.

Another experiment in the same project focussed on the information acquisition trajectory from the pilots after the aircraft has indicated that a system is malfunctioning [11]. In this case the malfunction was a low fuel level, caused by a leak in the fuel distribution system. The expectation was that one pilot would focus on solving this problem while the other pilot would stay focussed on flying the aircraft. However, eye tracker data revealed that often both pilots, after announcement of the problem, immediately focussed on the displays relevant to identify the problem. Here as well the strategy chosen by the pilots made that it was more difficult to divide the work in the best possible way because both crew members were in fact each by themselves focusing on the malfunction.

At Budapest airport a Remote Tower was installed [16] to replace the conventional tower. A Remote Tower is in fact an environment from which Air Traffic Controllers (ATCos) can control the traffic at and around an airport. The difference with a conventional tower is that the ATCos are not working in a real tower with windows all around them for a view over the entire airport. In a Remote Tower the ATCos are surrounded by computer displays offering them camera images of the airport together with other information such as radar displays, weather information etc. The advantage is that an
expensive tower building is no longer necessary and that several airports can be controlled from one location. Research questions that follow from such a change comprise:

- Are ATCos able to work equally efficient in both environments?
- Do ATCos use different operating strategies in the new and the old situation, and are there differences between ATCos?
- Will ATCos become more fatigued when working in a Remote Tower?

Eye tracking research was applied in the conventional tower and the remote tower with the same ATCos to compare behaviour in both environments. PERCLOS recordings revealed that the build-up of fatigue in a two-hour shift was the same in both environments.

In the way how ATCos divide their attention no significant differences between both environments were found (see Fig. 2). However, a trend was spotted that there are individual differences in how ATCos gather their information and which information sources / displays are used for that.

![Division of attention - Aerodrome control](image)

**Fig. 2.** No differences between remote- and conventional tower in where ATCos pay attention to.

There was a trend towards a lower eye blink frequency in the Remote Tower compared to the conventional tower, thereby suggesting a higher (visual) workload in the Remote Tower.

### 2.2 Rail

In the Doorstroom Station Utrecht (DSSU) study [17] the research question was: “Do less switches on the track influence the behaviour of train drivers
and will the driving speeds change?”. One of the sub-questions was: “Do workload levels change?”. This question was, amongst others, answered by analysing the eye blink frequencies (see Fig. 3) of the train drivers while driving during different simulated scenarios. No significant differences between conditions were identified. Also questionnaire data indicated that workload levels were not influenced due to the change. The scenarios either simulated the old or the new condition with varying degrees of traffic.

In the RouteLint study [18] a new tool was offered to traindrivers. This allowed them to monitor the traffic in front of and also behind them. Further, it allowed them to monitor whether trains are planned to switch tracks in front of them. The aim of the tool was to support traindrivers to plan ahead and drive more efficiently. One of the questions that was answered by means of an eye tracking study was: does the tool distract traindrivers when looking at RouteLint is unsafe for operations.

In a simulator experiment with scenarios with varying degrees of traffic and during different moments of a ride traindrivers where allowed to use RouteLint. It was proven that traindrivers do look at the RouteLint display when information is presented (see Fig. 4). But there was no significant shift of attention from another source of information. This was interpreted as: train drivers do use the information on RouteLint but at the moments when there is time available to take a quick look at the RouteLint display, therefore no significant distraction was identified.

Fig. 3. Eye blink frequencies of traindrivers under different scenario conditions. Error bars indicate the 5% confidence interval.
In the pilot study for the European Rail Traffic Management System (ERTMS) [19], eye tracking was applied in the field and in simulators to study how and when traindrivers pay attention to the European Train Control System (ECTS) displays in the trains. The ECTS displays are the displays on which ERTMS related information is presented to traindrivers in the cabin of a train. An important aspect of this study is the fact that in the introduction phase of ERTMS there will be tracks and trains, equipped with but also without the system installed. Research questions for which eye tracking was applied were: do traindrivers actually use information of the Planning Area (PA) on the ECTS display? On this part of the display the traindriver can see what is planned for his/her train. S/he can see for example: whether the track is free or whether speeds restrictions are foreseen.

The experiments in real trains provided a high level overview about how traindrivers divide their attention, especially when driving under ERTMS. It indicated how frequently and long traindrivers use the ECTS display, compared to monitoring the signals along the track. However, the bias when measuring in reality, in particular due to differences in traffic density, allowed to use these data as stepping stone for the development of simulator scenario’s, but for the in depth data analyses about how exactly ERTMS influenced traindriver behaviour, simulator studies were needed.
In the simulator, the eye tracker provided information about how attention was distributed, about whether and when the PA was used (see Fig. 5). The performance of the train(driver) was compared with the use of the PA. It was concluded that in particular during poor visibility conditions the PA is used and supports making train rides energy efficient and more comfortable for the passengers.

Fig. 5. A significant proportion of attention is paid to the planning area when relevant information is available. This is not clearly at the cost of another AoI.

3 Future application areas

The studies discussed above provide relevant examples of the usefulness of eye tracking as an instrument to assess human factors like attention, workload and fatigue. With respect to workload and fatigue it turned out that eye blink frequency and PERCLOS are indeed useful indicators. Most often this was done afterwards. However, eye tracking devices have become lighter as well as easier and faster to calibrate. These developments allow the use of eye tracking in real time, while it is still financially affordable as well. As a result a wealth of additional application areas for eye tracking, ranging from measurement in environments where this was initially impossible, to dedicated small eye trackers mounted into existing equipment, became feasible.
A feature that will have great impact in the near future is that the software for mousecursor control and for communication with others is by default installed in Windows 10. For laptops with eye tracking camera’s eye control has become a default option in Windows 10. This functionality allows working without a mouse, or to operate simulations and games much faster and much more efficient than nowadays. The interface of all kinds of software applications on such laptops will, because of that, become more intuitive with as one of the consequences workload reduction.

This same workload reduction mechanism becomes available on the newest smartphones and tablets. These will have desktop navigation and automatic page scrolling supported by build-in eye trackers. Modern smartphones and tablets [20] will be able to detect where a person is looking and predict where this person will be focusing upon next. Operator fatigue assessment can be done with discrete eye tracking equipment installed in the working environment. For example in cars [21] demonstrated that fatigue assessment can be done in natural environments. Nowadays, such systems are no longer expensive research tools but are affordable and easy to apply in many environments.

Trainers may benefit from a better real time assessment of the performance of their students. Wetzel et al [22] identified the added value of eye tracker output of F-16 pilots for trainers. The main advantages that they found are that eye tracking can be applied to develop effective scanning behaviour to enhance training. For example, the trainers could view the summarized scan path to check the students’ behaviour themselves but also for debriefing purposes. This type of knowledge can also be applied to manage student workload, and as such the training effectiveness.

New and complex environments like the F-35, that comprise a Helmet Mounted Display (HMD), offer a great deal of information sources that make it even more complicated for the instructor to check whether his/her student is scanning all relevant information at the correct moment. Eye tracking may offer the trainer additional knowledge about the student’s information scanning strategies [23] that s/he cannot even see when sitting in the vicinity of the student.

Outside the transport domain eye tracking offers possibilities for physically disabled, amyotrophic lateral sclerosis (ALS) [24] or “locked in syndrome”, patients. Eye tracking allows them to communicate with their environment. In
the past decades this was already possible at laboratory level, current affordable and accurate modern systems allow more and more people to use this technology in real life.

4 Conclusion

Eye tracking is a technique for assessment of a multitude of behavioural aspects of humans, including mental workload. It has a long standing history, but is still under constant development.

Eye tracking can be applied in numerous domains. It turns out that flying an aircraft or driving a train share the same principles for data acquisition and interpretation of eye scanning behaviour. Important is that for every experiment care needs to be taken to select the appropriate measure for that particular purpose. For example visual workload or mental workload are not exactly the same. Different eye tracking output variables such as pupil diameter or eye blink frequency, provide different insight in the exact type of workload that is applicable.

Also differences in purpose and set up of an experiment determine which eye tracker is most appropriate and how it can be applied in the most optimal way.

The author foresees that in the future more eye trackers will be mounted into computers, tools and equipment that we use. Hence, eye tracking will even more be used to predict what we are planning to do and make computers and equipment adapt better to the human operator.

References


152


Development of an Observational Protocol for Reducing and Mitigating Workload and the Risk of Retained Foreign Objects

Alison Kay1,*, Katie O‘Byrne1, Siobhan Corrigan1, Tiziana C. Callari1,2, Dubhfeasa Slattery3,4, David Smyth5, Michael Turner6, Deirdre Bennett7, Jamie Logan3, Sam Cromie1.

1 Centre for Innovative Human Systems, School of Psychology, Trinity College Dublin, Dublin, Ireland. *kayam@tcd.ie; obyrnek1@tcd.ie; siobhan.corrigan@tcd.ie; sdcromie@tcd.ie
2 School of Art and Design, Coventry University, Coventry, CV1 5FB, UK. tiziana.callari@gmail.com
3 Royal College of Surgeons in Ireland, Dublin, Ireland. dubhfeasaslattery@rcsi.ie; jamielogan@rcsi.ie
4 Medical Professionalism, RCSI and Bon Secours Health System, Dublin, Ireland.
5 Department of Surgery & Perioperative Services, University Hospital Waterford, Waterford, Ireland. david-zxc@hotmail.com
6 University College Dublin Centre for Human Reproduction, Coombe Women and Infants University Hospital, Dublin, Ireland. michael.turner@ucd.ie
7 Medical Education Unit, University College Cork, Cork, Ireland. d.bennett@ucc.ie

Abstract. Retained Foreign Objects are an uncommon but costly problem in today’s healthcare. It regards the outcome after an unintended item is left behind in a patient after an invasive procedure. This paper presents the development of an observational protocol used for surgical observations in the FOR_RaM Project. The FOR_RaM or Foreign Object Retention – Reduction and Mitigation project aims to analyze and understand the problem of retained foreign objects in surgery and maternity settings in Ireland, develop hospital specific foreign object management processes and implementation roadmaps, with a focus on reducing and mitigating the risk of foreign object retention. This paper discusses the methodology used for developing an observational protocol as part of a socio-technical multi-methods approach in order to gain a better understanding of the existing practices that take place in these settings, including workload, operational processes and collaboration. Emphasis is placed on the observational template development and design,
observational tasks, critical points, procedures and protocols followed throughout. This observation protocol has facilitated the collection of critical data and been successful in identifying good practices and potential areas for improvement.

**Keywords:** Observations, Procedure, Protocol, Template, Retained Foreign Objects, Workload, Operational Process, Collaboration.

1 **Introduction**

The Health Research Board in Ireland is committed to analysing and understanding the problem of Retained Foreign Objects (RFOs) in surgical and maternity settings. This primarily involves developing hospital specific foreign object management processes and implementation roadmaps, with a focus on reducing and mitigating the risk of retained foreign objects. A retained foreign object or RFO is not only costly financially to the Health System in Ireland, but can also have an impact on patient health and recovery, thus the impetus for the reduction of RFOs in Irish Healthcare. A RFO is when an item used during an invasive procedure is unintentionally left behind in the patient after a procedure ends [1]. RFOs are considered an uncommon but serious event, the incident rate; often estimated as a result of underreporting ranges between 1 in 1,000 to 1 in 19,000 [2]. Retained items include soft goods (e.g. swabs/sponges), sharps (e.g. needles), miscellaneous items (e.g. vessel loops or electrosurgical scratch pads) and instruments [3]. Complex environments such as surgical and maternity settings involve complicated procedures, group dynamics, technology and high risk medication use on a regular basis, these combined elements increase the potential for an adverse event to occur [4]. A RFO is a preventable adverse event, often referred to as a ‘never event’ meaning it should not happen [5]. The impact of an RFO may result in a number of negative outcomes: i) the patient may suffer from physical (pain) and psychological (depression) harm [6]; ii) damage to the reputation of the individual healthcare provider as well as the organisation; iii) financial implications can result in the case of a claim [6, 7]. The average cost of a RFO is $95,000 [8] but this can vary substantially depending on the location, outcome and impact. The cost can range from anything between $51–$3,988,829 [9].

One of the first crucial steps is to identify and understand what the key risk factors are in relation to RFOs. These include an emergency case, unexpected change to a planned procedure, no count performed or incorrect counts, multiple procedures or multiple surgical teams involved, lengthy procedures, procedures with a high blood loss and patients with a high body mass index
Many of these factors can create an impact on the task at hand making it increasingly demanding in relation to both physical and mental workloads throughout a procedure. The Joint Commission’s analysis of the most common causes of RFO events reported to them include the following:

- A lack of policies and procedures
- Incompliance with current policies and procedures
- Issues related to hierarchy and intimidation
- Failure or poor communication with physicians
- Poor or lack of communication of necessary patient information
- Poor or incomplete staff education

This level of analysis alone does not provide the necessary understanding of what actually happens in normal operational practice, particularly in understanding the complexity of healthcare systems. It can tell us when RFOs are more likely to happen, but it does not suggest how or why they are happening. For instance, it makes sense that an incorrect surgical count would be a risk factor, but how and why do incorrect surgical counts happen? What are the key differences between surgical settings and maternity delivery suite settings? This is where a Socio-Technical Systems (STS) analysis and modelling of the human and organizational factors is critical especially in relation to implementing effective change interventions.

This research was carried out as part of the Foreign Object Retention - Reduction and Mitigation (FOR_RaM) project. This 2 year project was funded by the Health Research Board in Ireland (Grant No: RCQPS-2016-2) which aims to reduce or mitigate the risk of retained foreign objects in two specific healthcare settings; surgery and maternity. Overall objectives include; 1) analysing and understanding the problem surrounding RFOs in these two settings within Ireland 2) developing hospital specific foreign object management processes and implementation roadmaps 3) provide a foreign object management toolkit for the settings involved. The FOR_RaM project is deploying an integrated evidence-based assessment methodology based on sound social-technical theoretical principles. This approach is beneficial due to the complex nature of these settings which involve high risk tasks often requiring complex interventions to overcome risk. Over time multiple interventions (counting protocols, use of radiography and technology, education and training) have been implemented to eliminate RFOs, yet this remains an existing and preventable concern. This could be due to the implementation of standardised interventions across multiple settings without considering the
differing cultures and norms or the high mental workload involved. This is referred to in the Medical Research Council guidelines where it is suggested that interventions may be more effective if they are adapted to fit the proposed setting [13]. This paper focuses on one section (unobtrusive observations) of the wider FOR_RaM project scope, placing recognition on the impact or influence mental workload can have within the bigger picture of reducing and mitigating RFOs. Mental workload can be described as the remaining cognitive capacity a person holds whilst carrying out a task [14].

Examining this concept in surgical and maternity settings may allow us to identify the common mental workloads that are experienced and assist in supporting ways to reduce demands that may impact the path to preventing RFOs.

It is argued that this overall approach can provide a better platform for understanding the current practices to prevent the risk of RFOs within these settings in which interventions can be developed and implemented successfully. First, using semi-structured research interviews we aim to provide a good baseline understanding of the existing practices, procedures and policies in place. Following this, in order to provide further information regarding: current practices and processes, team relations, dynamics, communication pathways and the specific tasks that take place in these complex settings, we will make use of unobtrusive observations. Observations involve investigation within natural settings, and it is considered central when one wants to capture “the whole social setting in which people function, by recording the context in which they work” [15, p308]. Observation is also a method for understanding how individuals construct their realities, and to better understand their experiences. It is suggested that observations support the holistic understanding of the phenomena under study [16].

This paper presents the development of an observational protocol used during surgical observations and the associated methodology to evaluate the data collected in a systematic and robust manner. The observation protocol was adapted from that used in the TAPOIA project which was reviewed and customized for the purpose of conducting observations in the healthcare setting. TAPOIA [17, 18] was an aviation project (Enterprise Ireland Commercialization fund 2012) to validate automated voice analysis technology in flight crew. The technology detected prosodic accommodation of flight crew speech as a measure of effective flight crew communication. The observation protocol was developed to measure multiple crew parameters against which to validate the technology - communication, situational awareness, decision-making, mission analysis, leadership, adaptability and assertiveness.
The TAPOIA observation protocol was considered a suitable starting point since, despite deriving from a different industry – aviation – the contexts being observed have much in common. Both require monitoring multiple actors working both individually and as a team in order to reach a safety critical operational objective within tight time constraints. In neither situation was it possible to halt proceedings and to question actors, so traditional formats and instruments (e.g. Situation Awareness Global Assessment Technique (SAGAT), Situation Awareness Rating Technique (SART), National Aeronautics and Space Administration Task Load Index (NASA – TLX)), for gathering data would not be appropriate. The rationale and the template used to derive the observation protocol will be presented below.

2 Related Work

2.1 Mental Workload in Healthcare

Human mental workload is considered the amount of cognitive effort required during task performance, therefore mental capacity is considered a person’s limitation for dealing with information received [19]. Cognitive overload is considered one of the contributory factors towards error in the airline industry [20]. Similarly to aviation, healthcare requires both high mental and physical workload to perform and complete day to day tasks. High workload in healthcare settings has been linked as a concern to patient safety [21] and can be closely linked to burnout among staff members [22] For instance, take the surgical count during a procedure. The task demands involve the scrub and circulating nurse counting instruments and items (swabs, needles etc.) added to the sterile field, a simple task when taken out of the busy context it is performed in. Introduce issues such as interruptions, distractions, hierarchy, time pressures or noise to an already repetitive task and it becomes more difficult and demanding requiring increased level of concentration, attention and memory. It is known that complex tasks are closely linked with mental workload and fatigue [23]. Similarly for surgeons and anesthetists, these issues present the requirement for higher mental workload also increases allowing for minimal capacity to successfully address or respond to certain procedural events which presents concerns [23]. In a study by [24] noise including interruptions and informal discussions were identified as influencing factors on performance from a surgeon perspective.
If high mental workload is associated with poor performance in industries such as aviation and other industries [19], considering its impact on operational processes in settings such as surgery and maternity could provide a better understanding of the current challenges associated with measures to prevent RFOs. An observational template was developed to gather information on both mental and physical workload involved in these settings. The steps involved in the development of the observational protocol are discussed below.

2.2 Development of Template

2.2.1 Stage 1: Initial Observation Template Meeting

Emphasis was placed on the need to identify tasks and critical points, potential staff and participant involvement and a description of what was going to be observed. This highlighted some important points going forward:

a) Tasks and critical points; column needs to be empty to allow for data collection, focus placed on the process as well as communications and interactions, an agreed reference list available on each template page and identification of decision points.

b) Staff and participants; reference to staff and time of entering and exiting the operating theatre, minimal detail to be included and an area for a sketch of location of personnel within the room.

c) Description; this column requires as much detail and content as possible including communication to whom, from whom.

2.2.2 Stage 2: Content Analysis

Discussions were made on how to approach data collection from a content analysis perspective. It was highlighted that the structure of the observation template needed to correlate with NVivo coding and structure as well as facilitating ease of data collection. The TAPOIA observation protocol was used as a base to build upon and adapt to the surgical setting. Figure 1 shows the initial draft of the template. Researchers involved in the FOR_RaM project reviewed this draft and the need to simplify the columns in the template was recognised. The task and subtask columns were combined as one. Description, characteristics, communication and observed behaviour columns were integrated into one single column headed ‘description’. Previous studies examining mental workload in the operating theatre setting link disruptions such as ‘tel-
ephone calls’ [25] and workflow interruptions [26] with increased mental workload. At this point the importance of monitoring and recording disruptions and interruptions was identified and a column for staff who entered or exited the area being observed was also included. This column was titled ‘staff in/out’. The ‘comments/remarks/suggestions’ column remained for the purpose of recording additional or more in-depth information on what was being observed. The simplified version of the template can be seen in Figure 2.

For ease at the analysis stages the use of syntax was applied [27]. This assisted in distinguishing between observers and participants comments, thus making the process of analysis more robust from a research perspective (less confusion with more reliability). This included:

- clear references of date and observation time frame
- quotation marks “--” around comments made by the participants
- straight brackets [---] around own ideas and comments
- plain text when describing working situations and critical episode

2.2.3 Stage 3: Observation Template Trial Run and Review

A trial run was performed using the simplified version of the template on two scenarios: 1) a clinical scenario - Implementing the WHO Surgical Safety Checklist in an operating theatre and 2) a non-clinical scenario - video of NASA employees reacting to a time and safety critical mission. The purpose of this was to determine if the template previously used in aviation was suitable for use in the operating theatre environment. The non-clinical scenario was a fast-paced NASA/Aerospace video excerpt with differing terminology to that of a healthcare setting. This gave the clinical researcher an idea of how out of depth the non-clinical researchers would be in this environment. At this stage the use of Performance Shaping Factors (PSFs) and Co-ordination Demand Analysis (CDA) [28, 29] was proposed on the basis of recognising workload and other factors that may impact the performance of tasks. The CDA was originally used in aviation research [17, 30, 31], but is highly applicable as it performs a measure of team co-ordination. The original research examined team co-ordination between air-crew on board NATO AWACHS, however, the safety critical nature of the surgical team setting is highly relevant for these measures. Following this it was discussed if PSFs and CDAs could be recorded after the observations (i.e. populated post-observation). Therefore this would be linked but become a separate document to the observation template. We reviewed the roles of the observers; Researcher ‘A’ would follow the patient (i.e. from the anaesthetic room into the operating theatre and out to the recovery room). Researcher ‘B’ (with a clinical background) would remain in the operating theatre to observe set up and clean up. This was utilised to
maximise as much of the operational process as possible. This was agreed due to the familiarity Researcher ‘B’ had with the “normal” running of an operating theatre, layout, processes and procedures and ability to identify deviations from standard operating procedures.

2.2.4 **Stage 4: Meeting to propose new draft template and CDA**

Proposal of the new template draft, PSFs, CDAs and recommendations from the observation trial run were reviewed. The PSFs and CDA methodology was approved for use but not as part of the formal process during observations but as a post-observational recall. This decision was made on the basis of insufficient time for completing these documents while observing. The proposition to keep the task/ subtask column blank for purposes of allowing the specific observation task flow dictate what is inputted was also agreed. Instead a structured table (Table 1) of reference was produced, based on a standardized/ generic procedure flow, location and personnel involved to allow those unfamiliar with the steps to easily refer to this where necessary. Abbreviations were made for both location and personnel involved in this reference list. A box below this contained more detail (Table 2).

It was agreed this would be made visible on each template page to allow for constant reference to be made. The surgical safety checklist (often referred to as ‘Time-out’) stages; Sign-in, Time-out and Sign-out (SSC 1-3) were also made available on every template page so they could be “ticked” as they were performed without having to revert back to the first page during the observation. A review of performance shaping factors from a number of safety critical human factors research projects was carried out. The researchers involved in FOR_RaM have considerable experience in carrying out ethnographic observations. A list of proposed PSFs (Table 3) was drawn up from the following projects:

- Aviation Maintenance, AITRAM project (EU funded FP5 project IST-1999-12241), [32].
- Process Industries, Virtualis project (EU funded FP6 project, Contract N. NMP-515831) [33, 34].
- Manufacturing, ManuVAR project (EU funded FP7 Contract N. CP-IP-211548)
<table>
<thead>
<tr>
<th>Task</th>
<th>Sub-task / Description</th>
<th>DP</th>
<th>Characteristics</th>
<th>Observed behaviour / Communication</th>
<th>Comments/Remark/Suggestions</th>
<th>Operation being observed:</th>
<th>Time started:</th>
<th>Time ended:</th>
<th>Quality of SSC scale:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 = No problems at all</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 = minor problem only</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 = Small problem</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3 = Major problem</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4 = Catastrophic</td>
</tr>
</tbody>
</table>

### SSC1
- Pre-op
- Anaesthetic checks
- Identifying patient
  - Identifying procedure
  - Acknowledged

### SSC2
- Reception of patient in theatre
  - Time out
  - Silence
  - Acknowledged

### SSC3
- Patient under
- Operation commences
  - Time out
  - Silence
  - Acknowledged

### Patient handover
- In recovery complete

Fig. 1. Initial draft of the observation template
<table>
<thead>
<tr>
<th>Task / sub task</th>
<th>Description</th>
<th>Staff in/out</th>
<th>Comments / Remark / Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>SSC1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Time out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Silence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acknowledged</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SSC2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Time out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Silence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acknowledged</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SSC3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Time out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Silence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acknowledged</td>
</tr>
</tbody>
</table>

Fig. 2. Simplified version of the observation template
This list was reviewed by researchers with a clinical background to ensure that the terminology was relevant and appropriate for the clinical setting and that there were no omissions. The 29 PSFs (Table 3) provide a detailed breakdown of the workload required at a task level. Tasks can be evaluated and allow for identification of the physical, mental and social workload demands observed among individuals performing tasks individually and together.

Table 1. Task and sub-task reference list

<table>
<thead>
<tr>
<th>Task / Sub-task</th>
<th>Location</th>
<th>Personnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt arrives &amp; checked in</td>
<td>HB</td>
<td>AN, A</td>
</tr>
<tr>
<td>Pre-op anesthetic</td>
<td>AR</td>
<td>AN, A</td>
</tr>
<tr>
<td>Sign-in (SSC1)</td>
<td>AR</td>
<td>AN, A</td>
</tr>
<tr>
<td>Anesthetic administered</td>
<td>AR</td>
<td>AN, A</td>
</tr>
<tr>
<td>Documentation check</td>
<td>AR</td>
<td>SN/CN</td>
</tr>
<tr>
<td>Baseline count</td>
<td>OT/ Prep room</td>
<td>SN, CN</td>
</tr>
<tr>
<td>Patient brought into OT</td>
<td>OT</td>
<td>A, AN, P</td>
</tr>
<tr>
<td>Connected/ set up/ positioned</td>
<td>OT</td>
<td>TEAM</td>
</tr>
<tr>
<td>Time-out (SSC2)</td>
<td>OT</td>
<td>A, S, SN, CN</td>
</tr>
<tr>
<td>Operation commences</td>
<td>OT</td>
<td>A, S, SN, CN</td>
</tr>
<tr>
<td>First count</td>
<td>OT</td>
<td>SN, CN</td>
</tr>
<tr>
<td>Second count</td>
<td>OT</td>
<td>SN, CN</td>
</tr>
<tr>
<td>Final count</td>
<td>OT</td>
<td>SN, CN</td>
</tr>
<tr>
<td>Sign-out (SSC3)</td>
<td>OT</td>
<td>SN, CN, S, A</td>
</tr>
<tr>
<td>Operation ends</td>
<td>OT</td>
<td></td>
</tr>
<tr>
<td>Pt transfer &amp; handover to Recovery</td>
<td>OT/ R</td>
<td>A, SN/CN, P, RN</td>
</tr>
</tbody>
</table>

Table 2. Abbreviations of location and personnel

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Anesthetist</td>
</tr>
<tr>
<td>AN</td>
<td>Anesthetic Nurse</td>
</tr>
<tr>
<td>AR</td>
<td>Anesthetic Room</td>
</tr>
<tr>
<td>CN</td>
<td>Circulating Nurse</td>
</tr>
<tr>
<td>HB</td>
<td>Holding Bay</td>
</tr>
<tr>
<td>OT</td>
<td>Operating Theatre</td>
</tr>
<tr>
<td>P</td>
<td>Porter</td>
</tr>
<tr>
<td>R</td>
<td>Recovery</td>
</tr>
<tr>
<td>RN</td>
<td>Recovery Nurse</td>
</tr>
<tr>
<td>S</td>
<td>Surgeon</td>
</tr>
<tr>
<td>SN</td>
<td>Scrub Nurse</td>
</tr>
</tbody>
</table>

2.2.5 Stage 5: Final Observation Template Review

The final template draft was reviewed by the FOR_RaM research team and the best format for printing (booklet style) was agreed. Researcher roles during observations were confirmed:
• Researcher A: to follow the patient (pre-op in the anaesthetic room and post-op to the recovery room)
• Researcher B: to remain in theatre observing the set up/clean up

Table 3. Performance Shaping Factors (PSFs)

<table>
<thead>
<tr>
<th>Performance Shaping Factors (PSFs)</th>
<th>Visual Access</th>
<th>Allocation of function to team members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength / Ability/ Body size required</td>
<td>Physical Access</td>
<td>Handover/Shift handover</td>
</tr>
<tr>
<td>Skill</td>
<td>Physical Access</td>
<td>Handover/Shift handover</td>
</tr>
<tr>
<td>Understanding of system</td>
<td>Dexterity</td>
<td>Role ambiguity</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Comfort</td>
<td>Communication difficulties</td>
</tr>
<tr>
<td>Physical effort</td>
<td>Team Culture</td>
<td>Interpersonal tensions</td>
</tr>
<tr>
<td>Experience</td>
<td>Stress</td>
<td>Team pressure</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>Repetitiveness of the job</td>
<td>Personnel resources</td>
</tr>
<tr>
<td>Fatigue</td>
<td>Memory</td>
<td>Hardware resources</td>
</tr>
<tr>
<td>Noise</td>
<td>Concentration</td>
<td>Time</td>
</tr>
<tr>
<td>Lighting</td>
<td>Interruption</td>
<td></td>
</tr>
</tbody>
</table>

2.2.6 Stage 6: Meeting with External Advisor

Input from the Nursing lead of the external advisory group for the overall project was obtained. This provided us with two additional tools which had recently been utilised in a national leading teaching hospital in the UK. Two features were added to the final template as a result of reviewing these documents and advice from the external lead; 1) Identification of the procedure as being an emergency or elective case and 2) identification if there were any changes to the procedure list. These were the two outstanding items from the template.

2.2.7 Stage 7: Further Review of the Template

The PSF updates were agreed upon and were inputted with the finalised CDAs. The updated template was reviewed and the finalised template (Figure 3) was produced and printed. The finalised PSF document and CDA document (Figure 4) were printed as separate items from the booklet of templates.

A checklist for on-site observations was agreed as outlined:
• Diagram for each observation
• Time in and out recording for people entering and leaving theatre
• Coding of paperwork (site, number and observer)
• Need to synchronise time for start/end session

165
<table>
<thead>
<tr>
<th>Observation</th>
<th>Task / sub-task</th>
<th>Description</th>
<th>SSR</th>
<th>Initial</th>
<th>Comments/Remark/Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operation(s) being observed</td>
<td>Pre-op anaesthetic checks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anaesthetic administered (pt under)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sign in (SSC1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anaesthetic administered (pt under)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connected / Set up / Positioned</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time out (SSC1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connected / Set up / Positioned</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baseline count</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Patient brought into OT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connected / Set up / Positioned</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time out (SSC2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connected / Set up / Positioned</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Operation commences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>First count</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second count</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Final count</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>acknowledged</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Operation ends</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Patient transfer &amp; Handover to Recovery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Finalized version of the observational template
A time out for researchers after each session for PSFs & CDA prior to discussing observation (to prevent contamination of results)

<table>
<thead>
<tr>
<th>Co-ordination Demand Analysis</th>
<th>Definition</th>
<th>Score (1-5)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>Includes sending, receiving, and acknowledging information among team members</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situational Awareness (SA)</td>
<td>Refers to identifying the source and nature of problems, maintaining an accurate perception of the patient location relative to the external environment, and detecting situations that require action</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Making (DM)</td>
<td>Includes identifying possible solutions to problems, evaluating the consequences of each alternative, selecting the best alternative, and gathering information needed prior to arriving at a decision.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mission analysis (MA)</td>
<td>Includes monitoring, allocating, and co-ordinating the resources of the team, prioritising tasks, setting goals and developing plans to accomplish the goals, creating contingency plans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership</td>
<td>Refers to directing activities of others, monitoring and assessing the performance of team members, motivating members, and communicating mission requirements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptability</td>
<td>Refers to the ability to alter one’s course of action as necessary, maintain constructive behaviour under pressure, and adapt to internal or external changes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assertiveness</td>
<td>Refers to the willingness to make decisions, demonstrating initiative, and maintaining one’s position until convinced otherwise by facts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Co-ordination</td>
<td>Refers to the overall need for interaction and co-ordination among teams</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 4.** Finalized Coordination Demand Analysis (CDA) template adapted from [28 in 29, p379]

It was agreed that the PSFs and CDAs would be completed as soon as possible (within minutes where possible) after each observation and before discussing any events relating to the observations. This decision was made on the basis of the high mental workload observers may experience while observing and recording critical
data. It was agreed this would have an impact on the data and critical information may be missed as a result.

3 Design and Methodology

3.1 Settings and samples

Observations were conducted in the operating theatres within an Irish Hospital with a bed capacity of over 400 and a total of 8 operating theatres covering a number of specialties. Multiple disciplines of the surgical team were observed during the observations, these included: Nurses, Anesthetists, Surgeons and Porters. The surgical specialties observed in this study included General, Orthopedics, ENT and Ophthalmology. These surgeries comprised of a combination of both open and laparoscopic procedures.

For the purposes of confidentiality no patient demographic characteristics were documented, only procedure type. Participant’s professional roles only were documented during observations to ensure anonymity. Ethical Approval was gained from the School of Psychology Ethics Committee, Trinity College Dublin and the participating Hospital’s Ethics Committee.

3.2 Observers

There were two researchers present throughout each observation. Researcher A: an experienced organizational psychology researcher, human factors trainer and consultant; who has worked on a wide range of human factors and risk management projects across a range of sectors (aviation, process, manufacturing, maritime transport as well as healthcare). Researcher B: a researcher with a clinical background, with experience in a number of roles across multiple specialties as a perioperative nurse.

3.3 Observation Protocol

A clear summary of objectives and themes prior to commencing observations were provided and agreed upon with professional leads (Table 4).
Participation was voluntary and the following observational themes were provided to participants on what areas were planned on being observed, this was to reiterate that participants were not being audited on their performance and to perform daily tasks “as normal”.

- Physical setting
- Social setting (formal & informal interactions)
- Interplay of different roles (nurses, consultants etc.)
- Social cohesion & team-work
- Information & knowledge flows (key decision points)
- Constraints & facilitators in normal operational practice

Table 4: Observation plan

<table>
<thead>
<tr>
<th>Activity</th>
<th>Objectives</th>
<th>Description</th>
</tr>
</thead>
</table>
| Theatre Observations| Formal and structured description of the surgery to be performed, including all communications and relevant performance shaping factors. In details:  
  - Detailed Task Analysis of all relevant procedures including deviations from standard practice and their consequences  
  - A breakdown of all personnel and equipment involved in the process;  
  - A map of the task according to a time, information and task flow (i.e. logic of process from start to finish) | Hierarchical break down of a task with main sub-tasks according to a nominal path for every stakeholder.  
Identification of exchange of information and communication among stakeholders.  
Identification of consequences of deviation on the task sequence/outcome of the process.  
Interactions of note  
  - Surgical Safety Checklist SSC  
  - Recognition of all SSC  
  - Patient “going under”  
  - Challenge & responses between nursing staff and surgeons, anesthetics team (all) and surgeons,  
  - Decision Points  
  - Patients “coming around” |

3.3.1 Pre-Observation Activities

Professional leads representing the operating theatre department were liaised with prior to conducting observations. Information about the study was distributed through verbal and written communication (i.e. information on noticeboards). Patients were provided with written and verbal information about the purpose of the study on the day of the observations. Patients were invited to participate in the study, verbal consent was obtained from all patients involved. Written consent was
obtained from Staff members involved in the observations. Staff were briefed before the observations or de-briefed immediately afterwards if they were not present prior to observation commencement. All stakeholders were informed of their rights with respect to withdrawal of participation, anonymity, confidentiality, data protection and security as per the agreements set out during ethical approval with the hospital ethics committee and the School of Psychology ethics committee in Trinity College Dublin.

3.3.2 Observations

Due to limited space in the anesthetic rooms and as a courtesy to the vulnerability of patients undergoing an anesthetic, observations commenced either prior to or on arrival of the patient into the operating theatre. Therefore the Sign-in phase of the Surgical Safety Checklist was not observed as this took place in the anesthetic room. The Time-out and Sign-out phases were included in the observations. Observations finished after the patient left the operating theatre. The patient was not followed to the recovery room as to respect patients’ dignity in the “coming around” and transfer periods.

3.3.3 Post-Observation Activities

On completion of the observation the researchers completed the PSFs and CDAs after each observation on site. The researchers did not discuss any events relating to the observations prior to completion of the PSFs and CDA to ensure that there would not be any contamination of results. Within 1-2 days after the observation was conducted the researchers reviewed their notes and findings collectively and produced an agreed CDA score for each observation.

4 Results and Discussion

A total of 6 observations have been conducted on procedures in the operating theatres thus far. All procedures observed were elective or planned cases. The total observational time was 6 hours 54 minutes (414 minutes). Observational time ranged between 25 minutes and 2 hours 30 minutes. CDAs and PSFs were completed for every observation. A further 2 observations are planned at this site and a minimum of 5 further observations are planned within a maternity setting. No results have been included in this paper as we are currently in the data collection phase and more observations are planned. While the data collected as part of the observation protocol will be analyzed in its own right this data will also be cross-referenced with the other methods used in the study, including interviews, focus groups and an analysis of finalized claims pertaining to RFOs nationally.
4.1 Overcoming Challenges

It was agreed upon with the professional leads that observations would not take place as planned when the patient entered the anesthetic room, but instead on arrival into the operating theatre. This decision was made with respect for the patient during the sensitive time of receiving an anesthetic, as well as a combination of limited space in the anesthetic rooms and the nature of ‘unobtrusive’ observations.

4.2 Discussion

It is without doubt that surgical procedures require a great deal of technical and non-technical skills, which are often dependent on communication and collaboration among a number of disciplines. The high physical and mental workload required during surgery can be demanding for both the individual and the entire surgical team. These demands can often present at different levels, to different disciplines at different stages throughout the perioperative phase. Taking a systems approach and the use of this observational protocol facilitates and supports the following;

- The collection of critical data pertaining to the workload required during a surgical procedure.
- A better understanding of the existing practices that take place in the surgical setting, including workload, operational processes and collaboration in preventing RFOs in Ireland.
- Provides powerful information to assist in the development of hospital specific foreign object management processes and implementation roadmaps.
- Creates an awareness and focus on reducing and eliminating problems associated with foreign object retention, through identifying good practices and areas for improvement.

4.2.1 Use of CDA to gather data on situational awareness

Situation awareness is described as the perceived understanding an individual has of their active surroundings and their decision-making ability based on the changing environment [35]. Whilst it may be reasonable to query the lack of more objective measurement of concepts such as situational awareness and workload, the rationale behind not doing so is justified by the complex and dynamic nature of the safety critical working environment. It can be argued that self-assessments of situational awareness can often be influenced by the individual’s level of confidence [36]. It would also be folly to ask a surgical team to pause and complete instruments such as SART and SAGAT, indeed doing so would not only pose extreme risk to the patient, but would also interfere with the natural flow of the operational process at hand. It would also interrupt the co-ordination and team dynamics which is precisely what the research team is mapping and analyzing. This is why the CDA
4.2.2 Measurement of Workload

Measures of workload commonly discussed in the literature include subjective (perceptions), objective (physiological) and procedural [37]. It is argued that the concept surrounding workload measurement is to determine if certain increased workloads cause inadequate performance [38]. A decision was made not to use subjective rating instruments such as the NASA-TLX or the SURG-TLX involving pre-and post-observations for participants. This decision was made on the basis that asking a participant to complete this instrument may also interfere with the overall objective of the FOR_RaM research project as it may make staff feel that they are being monitored regarding their workload. This may alter their behavior during observations and it would not give a measure of workload throughout the entire operational process which is what the research team was attempting to elicit. Although there is evidence in the literature of workload measurement using the NASA-TLX and the SURG-TLX both in the operating theatre and in simulation settings, much of the literature on surgery focuses specifically on surgeons measures of workload, with little emphasis placed on a multi-disciplinary perspective which is of importance in the FOR_RaM project. There was also a lack of suitability to the wider FOR_RaM project to obtain objective measures of workload (i.e. linking autonomic and cognitive activities as an indirect measure), due to its invasive nature [21]. Instead it was agreed that a NASA–TLX would be performed by outside observers for every critical point in the operational process. Therefore the NASA-TLX could accommodate recognition of workload involved and allow for a structured assessment of perceived physical and cognitive demands required throughout the operational process [39]. A minimum of 3 researchers (including one with a clinical background) would individually score each critical point in the operational process to ensure inter-rater reliability. Considerable effort will be made to ensure there are no ecological fallacies resulting from the analysis [40]. To do so, each individual will be scored (individually) using the NASA TLX and the Team Workload Questionnaire [41, 42] will be used to score the team workload at that critical point. Whilst it is not ideal that subjective measures (i.e. researchers rating an individual’s workload as opposed to the participant rating the workload themselves) of workload are used, it is important that we have some measure/assessment of workload throughout the operational process. It would not be possible to ask individual participants to do this without interfering with the process flow. With the current dearth in appropriate team measures, it is important that triangulation of results is carried out. This will be done extensively as FOR_RaM will interrogate the data from multiple perspectives:

- Operational Process Maps
- Performance Shaping Factors
Co-ordination Demands Analysis
• Workload (from both team and individual levels)
• Content Analysis

The FOR_RaM research project is on-going and the results will be made available when the final phase of data collection and analysis has been completed.

6 Conclusion and Future Work

This paper has identified the observational protocol utilized during observations conducted in a surgical setting and discusses the stages involved in developing the observational template used within the operating theatre. With this protocol we were able to identify a more in-depth picture of the mental and physical demands as well as the technical and non-technical skills used during the intra-operative phases. Applying this protocol to both the surgical and maternity setting will highlight differences in; tasks, priority of tasks, demands, norms and the physical and mental workloads involved. This information will highlight similarities and differences providing progress and understanding towards implementation of effective change interventions for reducing and mitigating the retention of foreign objects. The FOR_RaM project is currently in the data collection phase, thus further data collection and analysis will be conducted in both surgical and maternity settings before interventions are implemented. It is hoped that this tool will aid analysis and the understanding of current practice, support development of foreign object management processes and implementation roadmaps whilst addressing the level of mental workload involved in these two settings.

6.1 Limitations

Due to the dynamic and complex nature of the operating theatre environment we had no control over who entered the operating theatre under observation after commencement. This resulted in some staff members who briefly entered and exited the theatre but who were not directly involved in the observation and were unidentifiable; full written consent in some instances was not achievable, despite researchers’ attempts to locate staff after the observation. This was highlighted to the clinical leads and the ethics committee at the hospital.

Acknowledgements: The authors would like to acknowledge the Health Research Board of Ireland who provided the Grant (reference: RCQPS-2016-2) for this research under the Research Collaborative in Quality and Patient Safety.

Conflicts of Interest: The authors declare no conflict of interest.
References


Interactive Head Up Display in the cockpit to reduce Crew Workload

Tanja Bos, Antoine de Reus, Marieke Suijkerbuijk, Jaap Groeneweg, Wilfred Rouwhorst

Netherlands Aerospace Centre NLR,
Anthony Fokkerweg 2, 1059 CM Amsterdam, The Netherlands
Tanja.Bos@nlr.nl

Abstract. Head-Up Displays (HUDs) are gradually finding their way into the cockpit of commercial aircraft. By presenting primary flight parameters on a HUD, pilots can remain heads-up instead of switching between instrument panels inside and the outside view. Currently executed research is looking beyond the possibility of providing flight guidance to the flight crew with a HUD, by exploring the possibility to also interact with the symbology presented, through eye-gaze in combination with a manual input device. The objective of which is to reduce crew workload. Several applications of interactive HUD symbology were developed. The applications are targeted at the role of the pilot monitoring, who can change parameters to impact aircraft guidance and control. By using an eye tracker installed in a cockpit mock-up and presenting HUD symbology on the outside view, several scenarios were flown to determine the effect of the concept on workload. The results show that the HUD interaction is rather intuitive and easy to do, provided that the eye-tracker is well calibrated. Given the advanced cockpit environments of today we realised that it is difficult to reduce workload, by changing the input modality. Changing a parameter setting is not very effortful since the current cockpit interface is highly optimized and pilots are so much used to their current working method. However, providing novel functionality through the HUD that replaces a number of steps by a single selection on the HUD, such as changing the selected runway, has potential to make a difference.

Keywords: Head Up Display, Cockpit, Workload.
1 Introduction

1.1 Head Up Displays

Multiple studies have been performed to research the effect of using Head-Up Displays on pilots in certain flight phases [1], [2]. Particularly during landing in low visibility, a HUD provides essential cues for accurate alignment with the runway thereby allowing lower landing minima compared to not having such a device. Presenting primary flight parameters on a HUD, allows pilots to stay heads-up instead of switching attention between instrument panels inside the cockpit and the outside view. As such it may increase Situational Awareness and reduce workload.

HUDs have been in use in the military domain for decades, allowing the crew to stay heads up in air-to-air combat, when verifying primary flight and mission parameters. More recently, the HUD technology found its way into civil aviation. Firstly, primarily business or private jets were equipped with HUDs allowing them to land on smaller airports where little landing facilities are provided. Nowadays, large passenger aircraft such as the Airbus A350 and the Boeing 787 are being delivered with HUDs for use throughout the flight.

In the research we are looking beyond the possibility of providing guidance to the flight crew. We are exploring the possibility to also interact with the symbology presented on the HUD, through eye-gaze in combination with a manual input device. Literature reveals no application of this technology in aviation. In the automotive industry however some initiatives are ongoing to control non critical functions on the HUD through hand gestures [3]. For as far as we could find, interactive HUD systems through which to perform flight critical tasks do not exist nor is there any publicly available research on this subject.

1.2 Background

Workload is defined in this paper as the amount of cognitive load a human operator experiences. It is subjective: one operator may experience a high workload with a task, where another may experience a low workload. Personal characteristics and experience on a task play a large role. Task load on the other hand, can be considered an objectively quantifiable load, for example the number of sub-tasks to be performed. Airliners are flown by a two pilot crew. And it is typical for information systems in the cockpit to be redundantly integrated in the cockpit: both pilots have access to the same information. Typically the roles of pilot flying and pilot monitoring are distinguished. The pilot in command of the flight, generally the one with the highest rank, like a captain, designates who will be the pilot flying for a
particular flight segment. The pilot flying will be the pilot actually flying the aircraft and managing the flight path, while the pilot monitoring (alternative-ly named pilot non flying) communicates to the outside world, takes care of the navigation and works off checklists. The pilot monitoring is - as the name is implying – also cross-checking and monitoring the flight path and control activities of the pilot flying. Certain events during flight can suddenly increase the task load in the cockpit and cause one of the pilots to be head-down. This is typically the case when the flight gets a new runway at a rather late moment in the approach. In this case the pilot flying should fly towards the other runway and the pilot monitoring needs to make several changes in the flight management computer to prepare the aircraft systems related to that new runway. Allowing the pilot monitoring to make a single selection on the HUD that would replace these actions on the flight management computer could make a difference in task load and consequently in workload.

2 Concept

The concept consists of a HUD with 2D symbology of the basic parameters from the Primary Flight Display (primarily speed, altitude, heading and aircraft attitude). Also a 3D layer with a Synthetic Vision System (SVS) presenting an analogy of the real world can be visualised. This is used typically during landing, where it visualises the horizon and the runway outline in a world-conformal way. To interact with the HUD symbology the pilot looks at the symbology (for example the speed symbol) and presses a knob (a combined button and dial). When rotating the knob a new value can be selected. Pressing the device again acknowledges the newly selected value. This is how Speed and Altitude could be altered in the concept. Naturally the symbology provides feedback to indicate the interactive state and the current selection. When the runway symbol is visible in the area covered by the HUD, the pilot (flying) can change the currently selected runway by looking at the airport and pressing the knob. The HUD interaction would no longer require the pilot monitoring to look downwards but immediately send the new desired runway to the flight management computer, which will update the navigation display and go-around procedures. After having pressed the knob, initial selection starts and other runways are visualised in the HUD symbology. Rotating the knob toggles through the available runways at the destination airport. Subsequently pressing the knob will select the new runway. Also in this case the symbology provides feedback regarding the interactive state as well as the current selection.
In addition to the 3D symbology in the HUD, the currently selected runway is also indicated in the lower area in 2D symbology – irrespective whether the runway is in the HUD field-of-view or not. This indication can also be selected with eye gaze interaction in a similar way as altitude and speed selections. In this case a menu-scroll option becomes visible after selection.

In case of the interaction with the 3D layer, the outline of the active runway becomes visible in the area covered by the HUD at around 10 mile before the
runway. From that moment onwards, the pilot can change the currently selected runway using eye-gaze interaction. A cancel method is implemented in case the pilot inadvertently changes a parameter but has not yet pressed the knob to acknowledge. Cancelling is done either by looking outside the sensitive area around the selected parameter and pressing the knob or by looking at a dedicated cancel area on the lower left corner of the HUD and pressing the knob. The cancel area is only visible when a parameter is selected.

3 The experiment

3.1 Objective

The objective of the experiment with the interactive HUD system was threefold:

- To verify if the concept of an interactive HUD system with eye-gaze could work;
- To compare two ways of cancelling: look away and cancel area;
- To compare two ways of interaction: interaction with the 2D symbology and the 3D in case of the runway selection.

3.2 Participants

With the objective to evaluate the usability of the interaction, the experiment participants did not need to be airline pilots. A combination of pilots and students were invited. Four students participated, that were or are doing their thesis at NLR, were invited to participate in the experiment. Most of the students didn’t have any flying experience, while most of them had some familiarity or at least affinity with flying. Of the seven participants, three were pilots (one airline pilot, one former F16 pilot and one glider instructor). The new concept was evaluated with a single pilot setup.

3.3 Procedure

Four conditions were tested. In condition 1 and 2, speed and altitude changes were instructed by ATC. The participant was to repeat the instruction and change the setting using the interactive HUD concept. In both conditions one of the instructions was withdrawn by ATC, as to have the participant cancel the action to test this functionality. In condition 1 the pilot was to cancel by
looking away from the selected parameter in the HUD and meanwhile pressing the input device. In condition 2 a cancel area was visualised on the lower left corner of the HUD. Looking at the area and pressing the input device cancelled the action. In conditions 3 and 4, a late runway change was instructed by ATC. In conditions 3 the change was made through the 2D menu and in condition 4 through the runways presented in the SVS. After each run a questionnaire was filled in by the participant on the following topics:

- Workload (Modified Cooper Harper for controllability);
- Information density;
- The ease of performing the different tasks in each scenario/condition.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Experiment test conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>Action</td>
</tr>
<tr>
<td>1</td>
<td>Speed and altitude change</td>
</tr>
<tr>
<td></td>
<td>Cancel</td>
</tr>
<tr>
<td>2</td>
<td>Speed and altitude change</td>
</tr>
<tr>
<td></td>
<td>Cancel</td>
</tr>
<tr>
<td>3</td>
<td>Late runway change, 2D</td>
</tr>
<tr>
<td>4</td>
<td>Late runway change, 3D</td>
</tr>
</tbody>
</table>

3.4 Environment and measurement equipment

The experiment was held in NLR’s fixed-based cockpit mock-up APERO. It consists of a cockpit console with seats, representing the cockpit physical dimensions. The outside view is presented on a large display located about 1.5 meters from the Eye Reference Point (ERP).
The eye tracker (Tobii, Fig. 4) and the manual input device (Space Navigator, Fig. 5) were used to operate the interactive HUD system.

Fig. 4 Tobii bar eye-tracker

Fig. 5 Space navigator

The participants were briefed about the concept and got the opportunity to practice using the HUD interaction before executing the runs (test conditions of Table 1). The runs were performed once by each participant and the order was randomised.

4 Results

4.1 Making speed and altitude changes through the HUD

In condition 1 and 2 questions were filled in regarding the selection of speed and altitude (Fig. 6). The average ratings were taken for those instances where the selection was not cancelled.

![Average ratings of parameter selection speed and altitude (error bar presents the 95% confidence interval) N=7.](image)

**Fig. 6** Average ratings of parameter selection speed and altitude (error bar presents the 95% confidence interval) N=7.
4.2 To cancel a selection when it is not yet acknowledged

It was verified which mode of cancelling was more convenient: looking away from the parameter to be selected towards a non-sensitive area in the HUD or looking to a specific cancel area that is presented at the moment a selection is being made.

Fig. 7 Average ratings of cancelling a selection by looking away or by looking at the cancel area (error bar presents the 95% confidence interval) N=7.

4.3 To make a late runway change

Making a late runway change was evaluated using the 2D menu and the runway visualisations in 3D. The results are presented in Fig. 8. The difference in the ease of selection between the 2D and 3D symbology is significant with a two-tailed T-test ($p=0.0004$).

5 Conclusion and discussion

The objective of the experiment with the interactive HUD system was threefold:

- To verify if the concept of an interactive HUD system with eye-gaze could work;
- To compare two ways of cancelling: look away and cancel area;
- To compare two ways of interaction: interaction with the 2D symbology and the 3D in case of the runway selection.
The experiment showed that the concept could work. All participants were able to make the selections that were part of the scenarios. Making selections of speed and altitude was rated as rather easy (rated on average higher than 4.5 on a scale of 1 to 6). To cancel a selection was on average rated slightly more effortful (rated around 4 on a scale of 1 to 6). To cancel by means of looking away or by looking at a specified area did not provide significant differences. Some participants had a preference for looking away and others for the cancel area. By looking at the cancel area there was no possibility to inadvertently cancel the action. One of the participants mentioned to prefer an option where the selection status is automatically cancelled when the pilot is not manipulating the selected parameter for a certain number of seconds. To change the runway at a late moment in the approach was rated significantly easier using the 2D menu compared to the 3D interaction. In 3D the runway are visible at a rather late stage (around 10 miles from the airfield). At a larger distance the individual runways cover a too small area in the HUD. One of the pilots expressed his doubts about the concept of facilitating a late runway change. He mentioned that it should not be encouraged to change the landing runway in a late stage (10 mile in advance of the runway), and that might be a reason not to enable the pilots to do so as it increases the workload and puts pressure on regaining the mental picture. No baseline configuration was compared in the experiment. Neither was the crew concept addressed specifically since a first investigation step into the concept evaluation was executed in a single pilot experiment. Therefore no
clear-cut conclusions can be drawn on a (crew) workload reduction in comparison to the conventional cockpit yet. It can be expected that changing the speed and altitude through the HUD concept – which is normally done by means of a rotary knob on the glareshield – will not drastically impact (decrease) the individual crew member’s workload, because that is already a highly efficient, routine, and simple action. However, for the runway selection task a serious reduction in workload can be reasonably expected. The easy action through the HUD replaces several actions to be performed heads-down. Moreover, the real advantage of the interactive HUD is in that the Pilot Monitoring, who generally is executing this task, can remain heads-up, by which he/she can maintain Situational Awareness (SA). And therefore no cognitive resources are used for regaining SA. A general observation was that sometimes the eye-tracker needed to be recalibrated during the experiment. Also, with some participants the calibration was much easier than with others. It shows that some eyes are easier tracked than others. It may be concluded that the eye gaze interaction could be implemented in addition to the existing interaction means but not yet replace it.

6 Acknowledgement

This work was performed within the ANGI-HUD project, co-funded under the EU’s horizon 2020 Research and Innovation Program via Clean Sky Joint Undertaking’s Grant Agreement number 717080.

7 References


Pilot Fatigue Measurement

Henk van Dijk 1, Alfred Roelen 1
1 Netherlands Aerospace Centre, Anthony Fokkerweg 2, 1059CM Amsterdam, the Netherlands

Abstract. Pilot fatigue is a significant safety concern and has to be effectively managed. A number of procedures are in place to control the associated risks. However, these procedures are based on generalised management techniques, such as controlling hours of work rather than measuring the actual fatigue levels of pilots. However, there is no accepted, practical way to do this in an operational context and even in a laboratory setting this remains a challenging matter. This study addressed the following research question: "Can fatigue in individuals be measured with sufficient reliability in order to make a relationship with changes (deterioration) in flying performance?" The results showed that optimisation of fatigue assessment and prediction in an operational setting should consist of a multivariate assessment of physiology, behaviour, performance, environmental exposure, and sleep history. Fit-to-fly tests should consider standardization of the environment of testing to maximize its precision and sensitivity.

Keywords: Fatigue, fit-to-fly, measurement, test, pilot, flying, performance

1 Introduction

Pilot fatigue is a major safety concern and has been extensively investigated over many years. As a result a number of procedures are in place in Europe to manage the associated risks. However, these procedures are based on generalised management techniques, such as controlling hours of work (through rostering systems) rather than measuring the fatigue levels of individual pilots. Fatigue risk assessments generally assume that safety-critical staff arrive for their work without fatigue accumulated from previous activities. The fatigue situation of individual pilots is managed by self-declaration. However, this is not without problems as individuals may feel under pressure to undertake duties that actually should be declined. Being able to objectively measure the fatigue of individuals could potentially offer a
significant safety benefit. However, there is no accepted, practical way to do this in an operational context.

1.1 Research question

This study addressed one key question towards a goal of developing methodologies to better manage pilot fatigue. The key research question to be answered by this research study was the following: Can fatigue in individuals be measured with sufficient reliability in order to make a relationship with changes (deterioration) in flying performance?

2 Methods

The research question of measuring fatigue came with the added requirement that the measurement should be operationally practicable. This provided the additional challenge of allowing for only a limited range of assessments. The research aimed to select, from a very complete multivariate assessment, the most discriminative minimal dataset that was feasible in practice yet provided a robust estimate of task-relevant present and projected pilot fatigue.

‘Flying’ involves different types of activities that may require different skills that are not necessarily equally affected by fatigue. The question on the relationship between fatigue and deterioration in flying performance therefore required a systematic breakdown of flying into tasks and activities to be able to determine the effect of fatigue on performance of those tasks.

According to the research specification the following tasks were executed:

- Task 1: Available fatigue measurement techniques were considered and an appropriate number were selected for trial. A comprehensive collection of data was established, leading to a clear position on the most likely practical techniques for any operational use.
- Task 2: Significant flying activities were defined that may be sensitive to fatigue. The definitions were supported by accident and incident data.
- Task 3: Fatigue measurement was related to flying performance in a simulator. Flight performance data from the simulator was used to support analysis of performance of pilots having differing levels of fatigue in addition to commentary from the simulator observer.
2.1 Task 1: A review of fatigue measurement methods

This review considered available fatigue measurement techniques in order to establish a comprehensive collection of data leading to a clear position on the most likely practical techniques for assessing pilot fatigue. The review investigated fundamental techniques, not the effectiveness of commercial equipment which may have proprietary processes not suitable for examination. A systematic search was conducted of PubMed, using search terms such as fatigue, drowsiness, alertness, review. Additional articles were identified by manually searching bibliographies of retrieved publications. Furthermore, e-mails were sent to relevant scientists, using the authors’ network, with a request to mail applicable recent publications on measurement of fatigue and alertness to predict performance.

Criteria for selection of techniques were the following:

• The test should be fast and easy to undertake, ideally self-administered;
• The test results should not require specialist interpretation;
• The test should be robust against falsification of parameters;
• The test results should be unambiguous;
• The test should be socially acceptable and non-invasive;
• Test results should not be subject to learning effects.

These criteria applied to the final (set of) technique(s) to be proposed for application in an operational environment.

2.2 Task 2: Analysis of accidents and incidents involving flight crew fatigue

The critical factor for safety is not fatigue, or even sleepiness, but accident risk. Therefore fatigue needs to be related to safety by linking fatigue to (reduced) performance and linking performance to safety risk. The objective of Task 2 was to define significant flying activities that may be sensitive to fatigue. The result of this task was used to define the scenarios that were used in the flight simulation experiments that were carried out as part of Task 3 of the study. The approach for identifying safety-related flying activities that may be sensitive to fatigue was to compare accidents and incidents in which fatigue was a contributing factor with accidents where fatigue was not a factor. Sources of data were official accident investigation reports (mostly from the NTSB\(^1\)) and CAA’s Mandatory Occurrence Reports (MORs).

\(^1\) The NTSB was used a primary source of data because of the quality of the investigations and the relatively large number of accidents (resulting from the large volume of US air traffic).
each accident, the core pilot competencies (as listed in the ICAO manual of evidence-based training) that were affected by fatigue (in the case of fatigue related accidents) or whose substandard performance contributed to the accident (for non-fatigue related accidents) were identified.

2.3 Task 3: Fatigue in relation to flying performance

Two flight tests were used to assess performance on pilot competencies. Both flight tests were set-up to replicate a flight in a Boeing 747-400. During the experiments, several events were triggered to test specific pilot competencies. All participants in the study population (8 active duty pilots and 24 non-active duty pilots) were subjected to a flight test in the AIRSIM desktop flight simulator immediately after completion of the lab protocol. The test started with a short familiarisation session. The familiarisation session took the participant through all elements of the controls, instruments and systems displayed on the computer screen. There was also the opportunity for the subject pilots to ask questions. In addition to the familiarisation, the participants had received a preparation briefing one week prior to the experiment date. The flight test itself took 30 to 45 minutes to complete. The test was completed by a debriefing. The 8 active duty pilots were also subjected to a flight test in the GRACE high-fidelity simulator immediately after completion of the AIRSIM flight test. The test started with a familiarisation that allowed the participant to become familiar with the controls, general behaviour of the simulator and to ask questions. In contrast to the AIRSIM test, the GRACE test was a two-pilot operation. The subject pilot was accompanied by a project pilot who was part of the research team. The task distribution between the subject pilot and the project pilot was explained during the pre-flight briefing. The test was completed by a final debriefing.

3 Results

The results were presented below per task.

3.1 Task 1: A review of fatigue measurement methods

Fatigue is a complex phenomenon with many aspects that are possibly important contributors to or expressions of fatigue. This is illustrated in Figure 1. The left side shows the contributors to fatigue; the right side
expressions of fatigue. Note that contributors can also influence behaviour and physiology directly, i.e. without fatigue as a mediator (indicated by the dashed arrow). Fatigue cannot be measured directly. Therefore the contributors to and expressions of fatigue must be measured in order to obtain information on fatigue. In the next sections, the contributors to and expressions of fatigue are described with respect to possibilities for measuring.

**Fig. 1.** Contributors to and expressions of fatigue

### Contributors to fatigue

**Sleep**

The primary physiological method to reduce fatigue is to get sleep. The amount of sleep and the quality of sleep in a relevant period before the actual fatigue measurement can be used to make predictions of the level of fatigue and to better interpret results of actual measurements. Because sleep loss can be developed cumulatively, it is important to have information on the amount of sleep and the quality in the days (typically a week) before the measurement. The most complete objective measure of sleep is polysomnography, but this has the drawback that it is practically difficult to conduct in an operational setting. Information on recent sleep can also be objectively estimated with actigraphy (sensing and recording body movements) and this is therefore most commonly used to obtain objective

---

2 Dependent on the measurement technique different terminology is used, all referring to fatigue; e.g. vigilance, sleepiness.
sleep data. Subjective sleep information can be obtained from sleep diaries for information on day-by-day sleep history and questionnaires. A standardised sleep diary has been developed [1]. Examples of rating scales for sleep quality are the Pittsburgh Sleep Quality Index [2], the Groningen Sleep Quality Scale [3] and the Insomnia Severity Index [4].

**Circadian rhythm**

Sleep and wakefulness are controlled by an interaction of an output of the circadian pacemaker and a sleep-wake dependent homeostatic process. The circadian pacemaker and the sleep homeostat contribute about equally to sleep tendency and performance [5]. The human circadian system is normally synchronised with the solar day by exposure to daylight. A remarkably tight association between circadian rhythms of sleep propensity, melatonin and core body temperature has been described in humans [6]. Melatonin, core body temperature and light exposure history can be measured to determine if there has been a shift of the circadian rhythm with respect to the normal wake-sleep pattern (e.g. due to time zone crossing or night work). Melatonin can be measured from saliva samples or blood plasma, but obtaining a blood sample is significantly invasive. Core body temperature can be measured non-invasively. Both melatonin and core body temperature, but also other factors (exposure to light, exercise). Information on the recent 24-hour activity profile can be objectively measured with actigraphy.

**Individual characteristics**

No two persons are identical and individual differences will affect the influence of fatigue on one’s performance. Inter-individual variation of test results may be caused by different fatigue levels or by these individual characteristics, which can have origins in recent history, genetics, socio-cultural background or combinations thereof. Comparison of older and young adults for instance indicates that the circadian process has a greater adverse effect (measured as subjective sleepiness, calculation test performance and attention) on younger subjects than older subjects [7] and genotype influences susceptibility to the effect of sleep loss on performance [8].

**External condition**

Any measurement may suggest intact vigilance when assessed in a brightly lit cool environment in a person that has recently been physically active and has taken an upright posture. Fatigue may immediately start to increase when a person has to be sedentary, maintaining a sitting or supine posture in a dimly lit warm environment [9]. External conditions such as ambient temperature, lighting, and noise are known to influence the results of fatigue and...
performance tests [10-13]. Environmental conditions during the test as well as the subject's recent history of exposure to environmental conditions can be objectively measured for posture, temperature, light and noise.

**Stress**
Although acute stress response increases arousal and alertness, after prolonged stress exhaustion may be near. The recent history of subjective stress can be assessed using Ecological Momentary Assessment (EMA) [14], also known as Experience Sampling (ES), using smart phones [15].

**Expressions of fatigue**

*Physiology*
The most often used physiological measurements in an operational setting are ocular and cardiographic indices. Percentage of Eyelid Closure (PERCLOS) is often mentioned as a reliable indicator of the onset of sleep, but is more suitable for real-time monitoring than for a (short duration) fit-to-fly test because of compensation effects. Saccadic velocity, pupil size and pupil constriction are not under voluntary control and are therefore robust against compensation effects. Measuring these eye metrics is minimally intrusive. Of the cardiographic indices the Heart Rate Variability (HRV) is cited as a good indicator of sleepiness but HRV is also influenced by other factors such as exercise and digestion following a large meal. Sleep deprivation disturbs the co-ordinated thermoregulatory responses between the lower and middle part of the body. Skin temperature gradient (i.e. differences between skin temperatures measured at various locations of the body) can therefore be used as an indicator of fatigue.

*Subjective fatigue*
Subjective information on fatigue can be obtained from subjective rating scales such as the Stanford Sleepiness Scale (SSS), the Karolinska Sleepiness Scale (KSS) or the Samn-Perelli Scale (SPS).

**Behaviour**

*Cognition*
There exists a large variety of tests to measure cognitive performance. As a crude categorisation of these tests a model of human-information processing developed by Wickens and Hollands [16] is applied. According to this model there are two main information processing paths: A path involving very simple cognitive operations (sensory processing - perception - response selection - response execution); and a path involving more complex cognitive
operations (sensory processing - perception - working memory/cognition - response selection - response execution).

1. Simple cognitive operations
The most commonly used test for simple cognitive operations is the Psychomotor Vigilance Task (PVT) which is a simple visual reaction time task to test sustained attention. Popularity of the PVT is largely based on its simplicity which allowed early development (i.e. before the introduction of smart phones) of hand-held PVT testers. The Mackworth Clock Vigilance test is also rather frequently used. Due to the more complex visual representation it has primarily been applied in laboratory conditions but with current technology could easily be presented on a smart phone. Motor skills are tested with tracking tasks. These can be implemented on hand held devices and are therefore popular and quite often used, especially in operational settings. Simulated driving with lane deviation as measure of performance is a similar task but may involve more sophisticated hardware that makes it only practical to perform tests in a laboratory setting. Learning effects can be expected with these types of tests.

2. Complex cognitive operations
Information processing involving more complex cognitive operations can also be tested in a sophisticated driving simulator (or flight simulator) but the measure of performance is not as straightforward as in a tracking task. The number of threats detected and successfully managed has been used but this requires careful consideration of the scenario such that performance of different individuals can be compared. A large variety of tests has been developed to assess performance on complex cognitive operations. Several test batteries have been developed that combine a number of tests. The AGARD STRES battery and the Multi Attribute Task (MAT) battery were developed specifically to test pilot performance and are therefore particularly interesting.

Sleep latency
Sleep latency is the time that it takes to fall asleep. The most complete information is obtained with the Multiple Sleep Latency Test (MSLT) or Maintenance of Wakefulness Test (MWT), which uses the same combination of techniques (Electroencephalography (EEG), Electrooculography (EOG), Electrocardiography (ECG) and Electromyography (EMG)) as polysomnography, but similar to polysomnography has the drawback of being difficult to conduct in an operational setting.
Posture

Body posture, head movement and facial expression are often used in real-time monitoring systems of driver drowsiness. Similar to PERCLOS, these indices are not considered suitable for a (short duration) fit-to-fly test because of compensation effects.

3.2 Task 2: Analysis of accidents and incidents involving flight crew fatigue

An overview of the results of the analysis of accidents and incidents is presented in Figure 2, which showed how often each competency was involved on average per accident or incident. In the total set of 30 accidents and 35 MORs involving flight crew fatigue, and 28 accidents not involving fatigue, all ICAO-defined core pilot competencies were mentioned as being substandard. The most frequently mentioned competency for accidents and incidents involving fatigue is 'problem solving and decision making'. The overall picture resulting from the analysis of MORs was remarkably similar to that resulting from the analysis of accidents.

![Figure 2. Overview of analysis of accidents and incidents.](image)

For accidents that were not fatigue related, 'application of procedures' is the most frequently mentioned competency. Comparing accidents that were fatigue related with those that were not showed that 'situation awareness', 'problem solving and decision making', 'aircraft flight path management, automation' and 'communication' contributed more frequently to fatigue related accidents than to accidents that are not associated with fatigue. It was therefore concluded that from the perspective of flight safety, the effect of fatigue on these competencies was most important.
The analysis also showed that all ICAO-defined core pilot competencies were mentioned in fatigue related accidents. Therefore, it was recommended that the scenarios that would be used in the flight simulation experiments of Task 4 represented all core pilot competencies, with emphasis on 'problem solving and decision making'.

3.3 Task 3: Fatigue in relation to flying performance

**AIRSIM**

AIRSIM is a desktop research simulator that in this experiment represents a Boeing 747-400. The simulator consists of a computer screen representing a selection of instruments (primary flight display, navigation display, clock, main engine instruments and Flight Management System (FMS), control levers (throttles, flaps, landing gear and speed brakes) and a Controller Pilot DataLink Communication (CPDLC) window for receiving Air Traffic Control (ATC) instructions. Flight control inputs are given via a joystick and a mouse is used for manipulating the buttons, dials and switches. The scenario used for this experiment represented the last phase of a flight from Rio de Janeiro to runway 06 at Amsterdam Airport Schiphol. The experiment started approximately 20 minutes prior to the expected landing. At the start of the experiment the route was already programmed in the FMS and all relevant autopilot and flight director modes were operational. The scenario started overhead of waypoint DENUT southwest of Amsterdam Airport Schiphol. The flight continued along the RIVER2A-transition towards SOKSI, after which the aircraft was guided by ATC radar vectors. Several events and flight phases in the scenario were used to measure performance on specific pilot competencies. A summary of the events triggered is provided in Figure 3.

![Fig. 3. A summary of the AIRSIM events triggered.](image-url)
GRACE
GRACE is a research simulator featuring a 6-axis moving platform, a 120
degree visual system and a fully enclosed generic cockpit representing a
modern airliner. In this experiment, GRACE was configured to represent a
Boeing 747-400. The simulated flight was executed with two pilots: the subject
pilot and a confederate pilot who was part of the research team. ATC
communication was provided from the control room by one of the researchers.
The scenario used for this experiment represented the last phase of a flight
from Los Angeles to runway 06 at Amsterdam Airport Schiphol. The
experiment started with the subject pilot as the pilot not flying and the
confederate pilot as the pilot flying. The start position of the scenario was at
Flight Level (FL) 150 about 45 NM north-east of Amsterdam Airport
Schiphol approximately halfway along the MOLIX2A Standard Terminal
Arrival Route (STAR). The flight continued toward point SUGOL, after
which the aircraft was guided by radar vectors provided by ATC. The
duration of the experiment was approximately 30 minutes.

Effect of fatigue in flight crew performance
To determine if fatigue had an effect on flight crew performance during the
AIRSIM and GRACE experiments, the PVT and KSS scores were compared
with the performance scores for the various tasks during the AIRSIM and
GRACE tests.

Table 1. Correlation (Spearman’s rho) between PVT/KSS and AIRSIM scores.

<table>
<thead>
<tr>
<th></th>
<th>PVT reaction</th>
<th>PVT lapses</th>
<th>KSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/T off</td>
<td>-0.203</td>
<td>0.453</td>
<td>-0.114</td>
</tr>
<tr>
<td>Speed deviation</td>
<td>-0.257</td>
<td>0.253</td>
<td>0.192</td>
</tr>
<tr>
<td>Y deviation</td>
<td>0.238</td>
<td>-0.072</td>
<td>-0.159</td>
</tr>
<tr>
<td>X deviation</td>
<td>-0.232</td>
<td>0.381</td>
<td>-0.0755</td>
</tr>
<tr>
<td>Wind-shear response</td>
<td>0.378</td>
<td>-0.372</td>
<td>-0.266</td>
</tr>
<tr>
<td>Wind-shear turn</td>
<td>0.320</td>
<td>-0.279</td>
<td>0.080</td>
</tr>
</tbody>
</table>

The results showed weak correlations between the fatigue measures (PVT and
KSS) and performance on the flight simulator experiments. An explanation for
this could be that there was a relatively high variation in pilot performance
that was not related to fatigue, both between different pilots and within the
same pilot. This can possibly be identified by monitoring pilot performance
over a longer period of time under varying conditions such as during a routine
operational setting rather than a laboratory experiment. It is therefore
recommended to monitor flight crew performance and fatigue over a longer period of time during routine flight operations.

Table 2. Correlation (Spearman's rho) between PVT/KSS and GRACE scores.

<table>
<thead>
<tr>
<th></th>
<th>PVT reaction</th>
<th>PVT lapses</th>
<th>KSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>0.244</td>
<td>-0.171</td>
<td>0.404</td>
</tr>
<tr>
<td>Communication, pilot not flying</td>
<td>-0.107</td>
<td>0.480</td>
<td>0.482</td>
</tr>
<tr>
<td>Communication, pilot flying</td>
<td>0.037</td>
<td>0.468</td>
<td>-0.069</td>
</tr>
<tr>
<td>Y deviation 06</td>
<td>0.179</td>
<td>0.300</td>
<td>0.826</td>
</tr>
<tr>
<td>X deviation 06</td>
<td>-0.607</td>
<td>0.060</td>
<td>-0.165</td>
</tr>
<tr>
<td>Speed deviation 06</td>
<td>-0.393</td>
<td>0.450</td>
<td>0.699</td>
</tr>
<tr>
<td>Go-around decision</td>
<td>0.154</td>
<td>0.324</td>
<td>-0.169</td>
</tr>
<tr>
<td>Birdstrike</td>
<td>0.286</td>
<td>-0.150</td>
<td>0.089</td>
</tr>
<tr>
<td>Failure detection</td>
<td>0.036</td>
<td>-0.570</td>
<td>-0.368</td>
</tr>
<tr>
<td>Y deviation 27</td>
<td>-0.214</td>
<td>-0.180</td>
<td>0.089</td>
</tr>
</tbody>
</table>

4 Conclusions

This study addressed key issues towards a goal of developing methodologies to better manage pilot fatigue. The main research question to be answered was: "Can fatigue in individuals be measured with sufficient reliability in order to make a relationship with changes (deterioration) in flying performance?"

Although fatigue has been extensively investigated over many years, there is no consensus on a golden standard for the measurement of the actual and predicted level of fatigue. Given this state of the art, it was timely to leave the univariate approach behind and pursue optimisation of fatigue assessment and prediction using multivariate assessment. The research question of measuring fatigue came with the added requirement that the measurement should be operationally practicable. This provided the additional challenge of allowing for only a limited range of assessments. The research aimed to select, from a very complete multivariate assessment, the most discriminative minimal dataset that was feasible in practice yet provided a robust estimate of task-relevant present and projected pilot fatigue. Some interesting findings that were found throughout the pilot fatigue measurement study:

- The most frequently mentioned competency for accidents and incidents involving fatigue is 'problem solving and decision making'. However, the
analysis showed that all ICAO-defined core pilot competencies were mentioned in fatigue related accidents;

- Self-directedness, the individual ability to govern behaviour according to situational demands, helps people to be resilient to the effects of restricted sleep. This ability is definitely required in aviation and should warrant future investigations in a sample of active duty airline pilots;

- Fatigue questionnaires taken only once can be indicative of average fatigue levels throughout the week and can be considered as a simple means to estimate fatigue, although limited in operational use as the assessment may be manipulated;

- Subjective sleepiness is more sensitive to sleep(-disruption) than objective measures of sleepiness. However, in an operational context subjective reporting may be sensitive to the desired outcome. In such cases subjective reports should be complemented by objective assessments such as the PVT;

- Recording sleep, time awake, posture, light exposure and skin temperature all provide relevant information for estimating real-time levels of sleepiness. Of these variables, (unsurprisingly) time awake (since waking up that day) was the most potent predictor, followed by posture, light exposure, time spent in bed the previous night and skin temperature;

- The desktop flight simulator AIRSIM turned out to be a valid way of testing flight crew performance; and

- The results of both AIRSIM and GRACE flight simulator experiments indicated weak correlations between the fatigue measures and flight performance. It is expected that this correlation will be stronger in a more routine, operational situation.

In conclusion, the results from this study clearly showed that optimisation of fatigue assessment and prediction in an operational setting should consist of a multivariate assessment of physiology, behaviour, performance, environmental exposure, and sleep history. Fit-to-fly tests should consider standardization of the recent and current environment of testing to maximize the precision and sensitivity of the measured levels of fatigue.

**Acknowledgments.** Civil Aviation Authority (UK CAA) who funded this research.
References


Transmission Control Centres and Human Interface Design: The need to support situational awareness in the face of increasing complexity

Adrian Kelly, Electric Power Research Institute, Dublin, Ireland
Maria Chiara Leva, Dublin Institute of Technology Ireland

Abstract. In recent years in the Transmission Control Centers (TCC) on Bulk Electricity Systems (BES) across the world; desk operators and operations managers are coping with an increasingly challenging environment. The range of challenges includes; generation sources diversifying and being replaced with decentralised weather based resources, radically different system operating points, new codes and regulations and reportable Key Performance Indicators (KPI), increasingly complex control systems and transmission and protection equipment, ever evolving market systems and cross border arrangements. The BES is evolving rapidly, resulting in an ever increasing cognitive complexity and the resulting mental workload for the TCC operators that have to run the system within reliable operational limits. Therefore the tools and HMI that support the operators in their tasks must also evolve to better support the operators’ situational awareness and suggest actionable mitigation measures to face complex system conditions and disturbances. This paper presents the human factors challenges facing transmission operators in the BES and a quick overview of the tools and processes being developed to meet these challenges. The paper will also identify the gaps in human factors research to address these challenges in the near future.

Keywords: HMI, Energy, Electricity, Transmission, Control Center

1 Introduction

Electricity transmission system operators’ primary role is to ensure the safe, secure and economic transmission of electric power via the High Voltage (HV) and Extra High Voltage (EHV) network from sources of energy generation, to end users of energy. Since the earliest days of electrical power systems in the 1920s the flow of energy has been unidirectional, predictable and stable. The load centers were concentrated in cities, or large industrial regions and (depending on the system and geographical location) the energy generators had traditionally been large fossil and nuclear fueled generation plants, ranging in size from 200 MW to 1,500 MW. Traditionally, there were no geographical constraints on the location of generation stations provided EHV or HV links could be built to connect the generation station into the existing EHV or HV grid. This stability and predictability of the electrical grid was
stable, allowed transmission system planners and engineers to develop the transmission grid to ensure the BES was designed and built to meet customer demand and be secure for all contingencies. All Transmission System Operators (TSO) have one or more Transmission Control Centres (TCC) staffed 24 hours per day 7 days per week. TSOs have jurisdiction by geographical area, typically but not always, one per country. It is the system operations engineer’s role to balance generation with load always on the system under their control. The measure of balance between generation and load is the system frequency. The operator must ensure this is kept stable always (50 Hz in Europe and 60 Hz in North America) and be secure for contingency losses of generation or load. For decades, the role of the system operator was relatively straightforward, when compared to the situation today, although the systems were fundamentally less stable than today. From 1920s to 2000s, a small number of large, stable and predictable generation units with predictable prices were scheduled to meet stable predictable customer load. However, in more recent times - since the deregulation of electricity systems in Europe (1990s) and North America (1980s) - the role of the TSO and the TCC system operator has become increasingly more complex. The reasons are numerous, but five specific reasons are addressed in this paper:

1. Traditional generation sources replaced with Distributed Energy Resources
2. System load patterns evolving
3. Transmission system operating points changing
4. EMS Functionality and System Operator KPIs
5. Intelligent Electronic Devices (IEDs) and the increase in alarms in TCCs

Throughout history, system blackouts have caused major societal and economic disruption. These are tabulated in Table 1. The Royal Academy of Engineering in the UK estimate that a blackout lasting at least 12 hours on a weekday would potentially cost billions of pounds [1]. These complexities will be addressed separately in this paper, followed by a suggested approach for improvements to HMI’s to improve operator situational awareness to face these challenges.

### 1.1 Traditional Generation Sources Replaced with Distributed Energy Resources

Since the 1990s, some traditional generation sources (fossil fuel powered plants such as coal, heavy oil) have increasingly been replaced by renewable energy
Gas powered generation and combined cycle gas turbines have stayed stable or grown because of economic advantages, flexibility and other factors. DER cover a broad category of generation but in the main tend to be from on-shore and off-shore wind turbines, solar PV, concentrated solar plant, battery energy storage. However, the term DER can also cover smaller, thermal generation units such as waste to energy, combined heat and power plants and biomass fueled generation units located across the BES or on distribution networks. A comparison between traditional generation sources and DER is summarized in Table 2.

**Table 1**: Table of system blackouts from the past 53 years [2].

<table>
<thead>
<tr>
<th>Year</th>
<th>Region</th>
<th>Main Cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965</td>
<td>Northeast US and Ontario</td>
<td>Protection misoperation</td>
</tr>
<tr>
<td>1977</td>
<td>New York City</td>
<td>Lightning</td>
</tr>
<tr>
<td>1982</td>
<td>West Coast USA</td>
<td>High winds</td>
</tr>
<tr>
<td>2003</td>
<td>Northeast US and Ontario</td>
<td>Trees, operator error</td>
</tr>
<tr>
<td>2003</td>
<td>London</td>
<td>Transmission cable failure</td>
</tr>
<tr>
<td>2011</td>
<td>California Arizona</td>
<td>Hot weather, operator error</td>
</tr>
<tr>
<td>2012</td>
<td>Northern India</td>
<td>Excess demand, outages</td>
</tr>
<tr>
<td>2012</td>
<td>New York City</td>
<td>Flooding</td>
</tr>
<tr>
<td>2016</td>
<td>South Australia</td>
<td>High winds</td>
</tr>
</tbody>
</table>

**Table 2**: Comparison between traditional generation sources and newer DER type generation

<table>
<thead>
<tr>
<th></th>
<th>Traditional Sources</th>
<th>DER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size</strong></td>
<td>Large</td>
<td>Smaller</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>Centralized</td>
<td>Distributed</td>
</tr>
<tr>
<td><strong>Geographical Location</strong></td>
<td>Anywhere</td>
<td>Typically, near mountains or coasts. Not near load centers.</td>
</tr>
<tr>
<td><strong>Production cost</strong></td>
<td>More expensive (fuel dependent)</td>
<td>Less expensive (free fuel but tariffs included in cost)</td>
</tr>
<tr>
<td><strong>Generation Scheduling</strong></td>
<td>Predictable</td>
<td>Variable (weather dependent)</td>
</tr>
</tbody>
</table>
The move from large, centralized, generation units to small distributed generation units has posed significant challenges to TSO planning engineers, TCC operators, and Energy Management Systems (EMS) developers that are still the subject of intense R&D today. The operator’s situational awareness has moved from monitoring a small number of variables in central locations, to monitoring many variables in often disperse locations, due to the unpredictability of weather patterns. An example of this transition can be seen in Figure 1 - the transmission system of Ireland. The red squares indicate thermal generation plants and these can be compared to the more numerous green square indicating wind generation [3]. The amount of data being available to TCC operators from each of these DER sites has also increased. For a typical wind farm substation; the active power output (MW), reactive power output (MVAR), active power availability (MW), wind speed (m/s), air pressure (pa), temperature (°C), voltage (kV), current (Amps), feeder status (ON/OFF) among many others. As the electricity grid continues to rapidly evolve with rapid changes in generation resources, the situational awareness of TCC operators to manage these systems must also evolve and increase. But, this can be challenging due to the sheer amount of DER, and other challenges the TCC operator faces as documented below.

1.2 System Load Patterns Evolving

Total system load (or demand) curves have traditionally followed a common profile. The lowest system demand for the day was traditionally at approximately 04:00 (this is country dependent). After this low point, the load begins to ramp up before 07:00 as the energy users and industrial and commercial facilities begin their day. There are two “peaks” during the day:

In colder climates, where air conditioning load is not as common, the lunchtime (or midday) peak which is the highest part of the first part of the day is proceeded by a slight dip in the demand curve as people take lunchtime breaks. The main evening peak occurs at approximately 18:00 PM as people arrive home to start oven load and lighting load increases. In colder climates the winter months typically have higher demand and higher energy usage than the summer months.
Figure 1: Transmission map of Ireland showing traditional thermal generation (red squares) and new distributed energy resources (green squares) [3].

In warmer climates where air condition is common, the summer months have higher electricity demand than the winter months and the highest peak of the day occurs at lunchtime, when the temperature is highest. In more recent years, and certainly in the future, the traditional load curve will change due to a combination of: more energy efficient technologically capable consumer electronics such as washing machines, dishwashers etc. can be programmed and controlled to operate at night when demand for electricity is less intensive. This controllability of consumer electronics right down to the customers home can also be aggregated and potentially used by TCCs as a form of demand response to frequency control. Electric vehicle load is small at present worldwide, but is projected to grow rapidly in the coming decades [4], and consumers will charge their cars mostly at night. In the western world, traditional industrial load makeup consisted of large manufacturing, mining industries, with heavy motors and large loads. As economies transition, this traditional load is being replaced by service economy loads, office load and data centers. The move from large motor based load in industrial regions to power electronic based load in cities, has caused a shift in how the network is operated, and increased the challenges that TCC operators face that will only increase in future.
1.3 Transmission System Operating Points Changing

With reference to the transmission system of Ireland in Figure 1, a distinction can be made between two distinct BES operating modes. A) A windy day and B) A calm (non-windy) day. In the windy day scenario, the wind farms (indicated by green squares) located on the western Atlantic coast will be at maximum capacity, transmitting their generation through the regional 110 kV network to the load centers on the east, Dublin and Belfast and Cork to the south. In this scenario, the 110 kV network in the west and south can be stressed and at the minimum load part of the day (04:00 AM) wind generation must be curtailed (or spilled) for frequency security reasons.

In the calm day scenario, the wind generation is no longer generating and the system load is met by traditional thermal generation, indicated by red squares in Figure 1. This means that the system returns to a traditional mode of operation, load is fed from the centrally located generators and the 220 kV and 400 kV network are utilized. There is no need to curtail wind generation for frequency control as the available generation is dispatched based on an economic merit order. Depending on the weather, the system can move into and out of these two modes of operation regularly and rapidly, especially on island systems with large amounts of variable renewable generation and small amounts of load. The system that operators are monitoring and controlling can change and contingencies and equipment overloads can unexpectedly appear, requiring immediate mitigation measures and change of operating point.

1.4 EMS Functionality and Operational KPIs

The Energy Management Systems (EMS) and other software analysis tools that are in use in TCCs have also evolved in recent times to include complex visualizations with the stated aim of improving operator situational awareness. Some examples of customized displays using 3-D contours and heat map visualizations are shown in Figure 2. The TCC EMS market is dominated by a small number of large vendors. There is currently no set standard for EMS display design for individual consoles or for larger video wall displays for TCC EMS around the world. It is common occurrence that neighboring transmission systems which are part of the same synchronous interconnection will display information to their respective operators in their TCC in a completely different manner. How data is displayed in TCCs is not strictly covered by standards such as ISA 101.01-2015 [5] for HMI design in control centers nor are displays covered by regulations or codes or reliability entities. This is perhaps understandable given that each TSO is monitoring a
slightly different system and the roles and functions can vary but it leads to inconsistency in displays and a resistance to change despite best evidence that display techniques (such as use of color, non-use of trends etc) used are sub optimal, especially for something so complex and critical to society as the control centers of the BES.

![Figure 2: Examples of EMS TCC displays showing system operating modes. Source: Control Room Visualization and Situational Awareness [6]](image)

Traditionally, the original transmission control centers were known as load dispatch centers. Load dispatch centers essentially resembled (and evolved from) large substation control rooms and the mosaic type or “map-board” displays in load dispatch centers were displayed like the layout of large substations - with interconnecting lines and shapes, instrumentation panels displayed together. In the 1980s and 1990s as EMSs began to replace the mosaic boards the design of the original EMS displays mimicked directly the displays that were used on the mosaic panels in the load dispatch centers. This allowed operators to transition as easily as possible to EMS which was a revolutionary new technology, while still retaining their mental model of the system. Use of colour was ad-hoc and system dependent and displays evolved independently, based on need at the time or operator suggestions. Likewise, as EMS software was required to be upgraded over fixed time cycles, the display designs were just replicated in the upgraded software package, improving or designing new displays is low priority relative to delivering the stable EMS upgrade on time and securely. The system operation variables in EMS (such as line flows (MW and MVAR) and voltage at busbars were traditionally displayed like the software systems used to analyse and plan the transmission system, with tables of numbers instead of using trends and charts to display context dependent information.
As the BES is evolving in more recent years the EMS functionality and display design capability has increased, but the design of displays is system dependent and it is the responsibility of EMS database administrators in the individual system operators to utilize the capabilities of the EMS to maximize operator’s situational awareness in the face of increasing system complexity.

Added to the increases in EMS functionality in recent years is a steady increase of the metrics and KPIs that operators must control their system to. As well as the information and KPIs that the transmission system operator must report to their regulator or reliability coordinators. With increasingly interconnected systems (via HVDC and HVAC interties) and electricity markets, the interdependencies between system operators and their neighbours have consequently increased. These metrics are codified in Europe in the ENTSO-E System Operation Network Codes [7] or by the North American Reliability Corporation (NERC) in North America. System operators must be increasingly cognizant of new metrics and system limits KPIs as they relate to neighbouring transmission systems and breaches of these metrics and limits may result in enforcement measures being taken against the system operator, with consequent financial and reputational harm. Responsibility for ensuring the system is operated within limits lies primarily with the TCC operators.

1.5 The issues of Human Factors in Control Room Design and the specific challenges of Transmission Control Centers

To support the challenging task of including human factors engineering considerations in Control Room Design there are number of standards and guidelines available. These international, national and industrial standards
and guidelines will provide guidance on the minimum requirements in terms of human factor issues to be considered at design. For example, the international ISO 11064 standard for the Ergonomic design of control centres (ISO, 2006) [8] is an example of such international standards. In terms of industrial guidelines, the EEMUA 191 [9] is an industrial standard developed by the Engineering Equipment and Materials Users’ Association to support the design of alarm systems taking into account the requirements of the human operator receiving and responding to those alarms, while EEMUA 201 [10] is focused on the design of HMIs and gives guidance on areas such as display hierarchies, the design of the screen format, and the attributes of the environment which may affect the use of the HMI. These standards and guidelines are covering different aspects of design with different approaches. For example, ISO standards tend to use a general and broad approach in order to maximise the use of the application of standards for a variety of industrial purposes while EEMUA uses a detailed approach towards HFE for certain branches of industry. What they are both missing is any reference to the need to assess human workload connected to alarm management or specific scenarios that are foreseeable situations that an operator on his own or with the support of a supervisor may be expected to deal with. This assessment may have specific implications in terms of manning and or the need for better decision and or situational awareness support. The problem of the gaps in standards and methodologies for integration of human factors into design procedure is not a new issue and it has been pointed out in some recent literature [11] [12].

In this context TCCs poses specific issues as it differ from the control rooms of other industries that require human control of complex processes (such as chemical processing, petrochemical or manufacturing). The main difference is in the decentralized nature of the process under control. With the aid of Supervisory Control and Data Acquisition (SCADA) and Remote Terminal Units (RTUs) a TCC can exercise monitoring and control over hundreds or thousands of geographically disperse substations of varying degrees of criticality. Another feature of TCCs that is different to other industries is the diversity in technology and age of the equipment under control. This is especially true of protection equipment on the BES. Protection has been installed and monitored on power systems since its earliest days. Traditional protection equipment is classified as “electromechanical”, composed of rotating discs and mechanical parts. There was very little alarming on electromechanical relays for technological reasons. In more recent times, electromechanical relays have evolved to transistor based devices and transistor based devices have evolved to microprocessor based devices
classified as Intelligent Electronic Devices (IEDs). Modern IEDs that protect and control HV and EHV transmission equipment on the BES have alarm and monitoring capability up to hundreds or thousands of alarms that can be digitally transmitted to TCCs. Where IEDs have replaced electromechanical devices, the quantity of alarms available or presented to TCC operators has grown exponentially. Added to this growth in alarm quantity is the move to digital substations, the increasing number of renewable DER substations, and sensor data from these new sources, and the rise in recent years of power-electronic based transmission system HVDC intertie connections, with the consequent increase in alarms being brought to TCCs. Alarm overload for control room operators has been a root cause of some major accidents [13]. Alarm overload is also a typical factor causing a spike in cognitive workload. One solution, which has been adopted by some TCCs involves an agreed Alarm Management Philosophy (AMP) across the organisation, between the TCC, EMS Team and Engineering Design Managers. The philosophy documents how alarms are categorized in the TCC EMS and maps the alarm categories to operator action and action time-frame. The philosophy should also document which alarms should be sent from substations through to the TCC and how new alarms are added or existing alarms are modified in the EMS. Research in other industries has shown that the issue of alarm overload can be solved with a rigorous, consistent, documentation and process approach to the issue as described in the Alarm Management Handbook [14]. However vast improvements to the alarm overload issue can be made by simple changes to alarm displays and HMI graphics as will be shown in section 2 and 3.

2 TCC Function and Task Analysis Display Design

There are two quite simple Human Factors Engineering methods that can be used at design stage to improve HMI and EMS display for TCCs:

a) TCC Function and Task Analysis and

b) Work Domain Analysis and Abstraction Hierarchy

The TCC function and task analysis is a simple method to decipher and display the most relevant content to desk operators. It involves the EMS display (or HMI) designer analyzing - at a high level - the role that an operator on a desk carries out in relation to the overall system operator KPIs metrics. As an example of this approach, take the role of a generation engineer in a TCC (also called balancing engineer or frequency controller). The seven main tasks that the generation engineer may perform are; 1) Frequency Control, 2) Generation Dispatch 3) Reserve Control, 4) Load Forecasting 5) Variable Generation 6) Interconnection or Tie-Line control and monitoring 7)
Generation Based Alarm Management. Each of the seven roles may have varying priority levels or be of equal importance, depending on the system. To design a display for these roles, first design a tile based wireframe plotting the tasks at a high level. This should be done in consultation with experienced generation engineers to gain feedback. Once the initial wireframe is developed, the exact details of each display tile can be specified for the display design, including variables, colors, size, and description in words. This process should again be carried out in consultation with desk operators so that the HMI designer can produce as precise a display as possible for the generation engineer. When the detailed design is developed and agreed among the design team it can be produced as a customized HMI for the generation engineer in the TCC. An example is shown in Figure 5. This display design uses best practice HMI design principles as documented in The High-Performance HMI Handbook [15], including limited use of color except where abnormal operation is detected, information presented in its context (i.e. use of trends), use different types of trends for different information such as pie, radar, arrow, chart or based trends. The display shown in Figure 5 is a dashboard, giving the generation engineer a “one-look” overview of the system under their control. Each tile can be clicked and expanded to monitor or control an element of the display in more detail and will give a marked improvement in situational awareness for the generation engineer when compared to the display in Figure 3 can also be displayed as a video wall display in the TCC in front of the operator.

![Figure 4: Example role and function analysis of a generation engineer wireframe mockup for a HMI display design](image-url)
3 TCC Work Domain Analysis and Abstraction Hierarchy

To take the TCC function and task analysis discussed in section 2 further, an Ecological Interface Design (EID) as proposed by Vicente and Rasmussen [16] is deployed in many environments for many years to simplify complex data presentation for human operators to process. The environments where it has been successfully deployed include air traffic control and military [17], which are similarly highly stressful environments with ranges of complex data sources and human interaction. The process for creating an effective EID is underpinned by Work Domain Analysis (WDA) process, whereby an abstraction hierarchy is developed linking functional purposes, abstract functions, purpose-related functions, physical functions and physical objects. The first WDA of this kind for electric power grids and transmission system operators was developed for the Independent Electricity System Operator in Ontario Canada (IESO) by Hillard, Tran and Jamieson [17]. The WDA development involved interviews with operators and engineering staff and research into the functions, KPIs and roles of IESO and the desk operators in the control centre. The abstraction hierarchy, once developed, gives clarity to the operator’s role and function in the control center in the context of the objectives of the control center (maintain reliability, control frequency etc.) and the physical objects and tasks that the system operator uses all day every day.

Figure 5: Mock Up generation engineer dashboard based on function and task analysis
Importantly, once the abstraction hierarchy is developed, it gives clarity to the HMI or EMS display designer to the optimal types of displays, and the information required within the displays to improve operator situational awareness to manage an increasingly complex system. The abstraction hierarchy, once completed can be scaled appropriately. For example, if new physical objects are added to the system or the system operator develops a new functional purpose or role these can be easily added to respective layers and HMI displays developed or adapted to accommodate the new functions.

An example abstraction hierarchy for a system operator is shown in Figure 6 taken from the work proposed by Hilliard, Tran and Jamieson [18]. Compared to the simple function and task analysis developed and illustrated in Figure 4 the abstraction hierarchy is far more complete, complex and detailed; however, the purpose and end goal is the same. Both methods can be used to develop the generator dashboard or video wall display illustrated in Figure 5.

![Figure 6: Abstraction hierarchy for IESO developed by Hilliard, Tran and Jamieson 2017][19]

### 4 Conclusions and Future Work

Transmission system operators are increasingly faced with the task of operating and controlling more complex, challenging and rapidly evolving electricity grids. In tandem with this complexity, the economic and societal consequences of operator mismanagement, causing blackouts has never been
higher. However, the HMI and display designs in the EMS that the operators in control centers have at their disposal to assist them in the face of the increased mental workload associated with control room activities have in general remained static and unchanged, despite an evolution in the systems under control in the past two decades. This paper highlights the need to consider Human Factors Engineering approaches at the core of HMI design for transmission control rooms. Future work would involve a continuation of the work already developed by Hilliard, Tran and Jamieson [18] to encompass an entire control center work domain and the implementation of elements of the abstraction hierarchy into the business processes of system operators control centers and EMS teams. This work can be assisted by analyzing mental workload associated to different tasks and verify how possible peaks could be managed/reduced in the HMI design and or through the control centers work domain analysis by supporting supervisors’ interventions and “keeping in the lop” support functions. Additionally, it is recommended that TSOs and EMS vendors should be provided better guidelines and or reference standards when designing displays and HMI for their TCCs. This would ensure consistency of approach across the industry with the ultimate goal of improving TCC operator situational awareness and improved BES reliability.
References


216

Electroencephalography Studies of Mental Workload – A Neuroergonomics Approach

Pavle Mijović1, Ivan Gligorijević1, Bogdan Mijović1
1 mBrainTrain, Požarevačka 36, 11000 Belgrade, Serbia
{pavle.mijovic, ivan, bogdan}@mbraintrain.com

Abstract. Neuroergonomics is defined as a science discipline that aims at investigating human brain functions in relation to work. With increasing number of wearable sensing solutions for recording brain dynamics at work, this discipline attracted attention of both neuroscience and human factors and ergonomics (HFE) experts. The main advantage of neuroergonomics over the classical HFE methods is that it provides the possibility to investigate the covert cognitive processes that are in functional relationship to the worker’s behavior. One of the most powerful neuroimaging methods that can be applied in neuroergonomics research is electroencephalography (EEG), mainly due to recent trend of miniaturizing the EEG equipment and due to its high temporal resolution. This work briefly introduces neuroergonomics as a science discipline and reviews the recent scientific work where EEG was used as a methodology for investigating mental workload in applied settings.

Keywords: Neuroergonomics, Electroencephalography, Wireless EEG, mental workload

1 Introduction

Wearables are becoming ubiquitous in everyday life, since humans are willingly accepting the technological advancement of sensing technology, mainly for improving daily routines and overall well-being. This trend has opened a whole set of new topics in the applied psychology research, as the equipment for physiological sensing is getting smaller, which opens a new possibility for concealed measurements of human brain functions in the applied settings. Following this trend, the field of Human Factors and Ergonomic (HFE) became richer for the new discipline called Neuroergonomics. Neuroergonomics aims at investigating brain functions and their relationship to work [1]. It provides possibility to objectively quantify the
worker’s cognitive state, through investigation of the covert cognitive processes, as opposed to the classical ergonomics methods that mainly rely on the qualitative assessment and behavioral research [2]. Neuroergonomics integrates scientific disciplines of HFE and neuroscience while trying to exploit the benefits of both [3]. The main aim of neuroergonomics studies is to enrich the HFE research by providing precise analytical parameters of brain functioning and behavior in naturalistic settings [1], rather than evaluating human performance through the overt behavior measurements [1, 2], which are mainly based on theoretical constructs [4]. Additionally, dominant approaches in the HFE domain were behaviorism (i.e. stimulus-response psychology) and the cognitive approach in assessing the human performance, while the brain-related mechanisms were largely neglected [1]. Advancement of neuroimaging technologies developed the field of cognitive neuroscience and in the latest years, these technologies were also considered by the HFE specialists. This is important because understanding of the brain processes in the naturalistic environments can lead to improvement of existing industrial processes design and to creation of more efficient and safer working conditions [2], consequently improving the operators’ overall wellbeing. Following this path, the neuroergonomics is expected to especially benefit from the real-time data acquisition and processing, which can enable timely investigation of different workplace parameters that are influencing worker’s cognition, thus providing a valuable input for the workplace optimization [5]. Neuroergonomics has already had a significant success in evaluating brain activity in the various workplace settings, including a large number of industrial sectors, e.g. transportation, aviation, manual assembly work in traditional industrial settings, etc. For instance, neuroergonomics studies have investigated the interaction between the worker and the automated systems, through the studies of dual-task performance [6], operators’ vigilance [7], mental workload assessment [8], assessment of the concurrent physical and cognitive work [9], transport research [10], etc. Additionally, neuroergonomics was successfully applied in neuroadaptive systems that serve for the mutual interaction between an operator and an automated system. The main aim of these systems is that both the operator and the system can initiate the change in the level of automation if needed [11]. Notably, neuroergonomics also covered the field of manual work, where it was shown that simple change in task [12], or providing workers with the frequent micro-breaks [13] can have a positive influence on the operator’s attention level during a monotonous tasks. It was also recently shown that the neuroergonomics methods can be used for the on-line attention monitoring of the operators [14], which could later be implemented in one of the passive brain computer interface (BCI) systems [15].
This work summarizes the neuroimaging techniques that are applicable in neuroergonomics research. Other physiological sensors (e.g. eye-trackers) can also be successfully applied in neuroergonomics studies, but these techniques are not covered by this review. Further, a brief discussion of pros and cons for each neuroimaging technique will be provided, with the main focus on electroencephalography (EEG) studies. A review of the recent neuroergonomics findings on mental workload (MWL) measured by EEG devices will also be provided in the following chapters.

2 Neuroimaging Techniques applicable in Neuroergonomics studies

Extensive review of neuroimaging techniques applicable to neuroergonomics research has been recently published by Mehta and Parasuraman [1]. The neuroimaging techniques can be divided in two distinct groups according to their recording mechanisms: the ones that use techniques for indirect metabolic indicators of neural activity, and the ones that utilize the direct measurements of the brain activity [1]. The first group consists of techniques such as functional magnetic resonance imaging (fMRI), positron emission topography (PET) and functional near infrared spectroscopy (fNIRS). On the other hand, Electroencephalography (EEG) and therefrom derived event related potentials (ERPs) belong to the neuroimaging techniques that directly measure brain activity [1, 16], together with the Magnetoencephalography (MEG). The main distinction between neuroergonomics and neuroscience is that the former aims at investigating the brain functioning in relation to work and therefore when evaluating which neuroimaging method should be used for neuroergonomics study following three important criteria should be considered [1]: (1) Temporal resolution, (2) Spatial resolution, and (3) The degree of mobility. The temporal and spatial resolutions present the ability of the recording device to discriminate between two data points in time and space, respectively, while the degree of mobility relates to the dimensions of the recording equipment and its usability for usage in naturalistic environments [1]. Although fMRI and PET provide high spatial resolution, their low temporal resolution and big dimensions of recording equipment limit their usability for recording brain activity in natural environment [16]. Thus, from the first group fNIRS remains the single convenient technique for the neuroergonomics research in naturalistic setting due to it is lightweight, being consequently wearable [7]. However, since fNIRS also suffers from low temporal resolution, which is on the order of several seconds per data samples, its use in dynamic naturalistic settings is still somewhat limited [16]. In order
directly investigate the sub-second brain processes the neuroimaging technique has to have a very good temporal resolution. Two widely employed methods are EEG and MEG. However, MEG is still contained solely to laboratory conditions due to the size of the recording equipment [1], thus leaving EEG as the most favorable method for investigating brain dynamics in unrestricted environments [16].

2.2 EEG and its metrics applicable in Neuroergonomics studies

EEG is a non-invasive recording technique, which measures the electrical brain activity that originates in neocortex, using the recording electrodes that are placed on the subject’s scalp [16]. The millisecond temporal resolution of EEG systems makes them suitable for the real-time investigation of brain dynamics in complex environments. Although it suffers from low spatial resolution, the information that is obtained from its frequency bands and ERP components allows researchers to understand how diverse situations influence the processing of the brain. Apart from using ERPs for assessing the patients’ various clinical conditions, over the past 40 years ERPs were recorded from healthy individuals for assessing various covert cognitive mechanisms. Back in 1990, Parasuraman proposed the introduction of ERPs in ergonomics research and discussed about possible benefits of its application in various HFE problem areas. However, until recently the traditional EEG recording suffered from a long wiring between electrode cap and amplifier unit, which engender the artificial artifacts that degrade the quality of the recorded signal [17]. Additionally, EEG recordings usually required shielded dimly lit and sound attenuated room, which was one of the main preconditions for its recording, thus limiting its use in naturalistic environments [16]. These problems were recently overcome by the development of the wearable EEG systems, empowering their use in naturalistic and applied environments [17]. This was mainly attributed to the miniaturization of the research-grade EEG amplifiers (such as, e.g. SMARTING, mBrainTrain), which can now be mounted on the participants head, thus reducing the cables length and eventually diminishing the artificial noise which are caused by the long wiring. The electromagnetic interference, mostly generated by power lines (50Hz or 60Hz noise) is solved by smart electronic designs. Since the covert cognitive context is usually encrypted in complex brain dynamics it is hard to isolate the specific cognitive processes in naturalistic settings (Bulling and Zander, 2014). In order to do so, it is important to first carefully design the study, since it is important to precisely evoke some of the cognitive features, such as ERP. Another way of accessing
the meaningful brain related features is the investigation of the frequency bands of the EEG signals. In this way, it is possible to continuously observe the changes in the covert cognitive processes. The ERPs represent a voltage fluctuation in the EEG signals that are related to the occurrence of the specific stimuli (which can be visual, auditory or tactile). One of the most interesting ERP components for the neuroergonomics studies is the P300 component that usually starts in the time window between 200 and 400 ms after the stimuli presentation. It has been already confirmed that the P300 component is not related to the physical attributes of the stimuli, meaning that it is rather related to the depth of the cognitive processing of the specific stimuli [18]. When assessing the cognitive state from EEG frequency domains one could use the basic index, which consists of solely calculating power ratios for each of the frequency bands, or the ratio index that is derived from the ratio between power of frequency bands [19]. The brain rhythms are generally divided in frequency bands: δ (delta: < 4 Hz), θ (theta: 4-7.5 Hz), α (alpha: 8-12), β (beta: 13-35 Hz) and γ (gamma: >35 Hz). However, depending on the literature, the spans of the frequency bands can vary. The low frequency waves (such as delta, alpha and theta) are usually high in amplitude and are notable in the state of rest, relaxation, sleepiness, low alertness etc. On the other hand, the high frequency and low amplitude waves are reflecting the alert state, state of wakefulness, state of task engagement, etc. One of the interesting feature that has been used in neuroergonomics studies is the engagement index (EI), that represents the ratio between the high and low frequency waves (EI = β/(α+θ)). The increase of the EI represents that the person is more cognitively engaged in the task, while the lower values of the EI represent the lower cognitive engagement on the task at hands [8].

3 Neuroergonomics studies of MWL

EEG is a powerful tool for obtaining the objective measurement on the influence of mental workload on the available mental resources during the regular work routine. For instance, Raabe et al. [20] investigated how different workload conditions are influencing the amplitude of the P300 ERP component, during the driving task while the secondary task was listening to the auditory oddball task. They found that the task conditions significantly influenced the P300 amplitude, as in the high MWL condition, the P300 amplitude was significantly lower, as compared to the low MWL condition [20], proving that during the high MWL the attentional resources were significantly lower than in the low MWL task. Another study investigated the influence of the two distinct visual operation alarm systems on cognitive
resources in the simulated air-traffic control (ATC) environment [21], also through the evaluation of the P300 component amplitude that was related to the secondary auditory task. The authors reported here that the more salient alarm system showed better performance, but also induced higher P300 amplitude in the operators, meaning that the enhanced visual design allowed more cognitive resources to be dedicated to the brain processing of the secondary, auditory stimuli [21]. In another study, the workload was estimated using both the frequency features and the ERP measurements in an n-back psychological task, where the workload was varying by varying the n (number of preceding) letters to the one that was presented in the screen [22]. They found that the EEG power that the best discriminates between the MWL conditions was the modulation of the alpha band (between 0-back and 2-back task). In the ERP measurement, the authors reported that again the P300 component was significantly lower in the high MWL condition (2-back) as compared to the low MWL condition (0-back). They also proposed a model for the MWL estimation based either solely on ERPs, or frequency features, but also the fusion model that combined both measures and reported that the classification accuracy of the fusion model showed the highest results, up to 90% [22]. In the study of monotonous manual assembly work it has also been found that the MWL modulation can influence the attentional resources of the operators through the investigation of the P300 component amplitude, E1 and the behavioral measurement of the operator’s movements that are not directly related to the task [8]. In fact, it was found that slight modification of the manual assembly work routine (that is “automatic” by nature), in the sense that workers should be instructed with the way of initiating the operation (slightly increasing the MWL) can be beneficial for the worker’s cognitive state.

4 Concluding Remarks

This work aims in disseminating neuroergonomics as a powerful research method in the MWL domain. Only a limited number of the scientific studies related to the objective measurement of the influence of the MWL on the operators cognitive state was presented and the main focus was to explain how physiological measurements (specifically EEG) can be utilized in MWL research. Although Parasaruman and Wilson [23] modestly stated that neuroergonomics is not revolutionary, but rather another step in HFE research, the growing body of neuroergonomics research refuted their statement. In fact, ever advancing technology supported neuroergonomics research and only fifteen years from its emergence, it tends to becoming one of
the main directions in HFE research. This trend is currently driving the establishment of neuroergonomics as the new scientific discipline and in the years to come it is expected that an increasing number of HFE specialist will also adopt the neuroergonomics methods as one of the standard tools for the MWL research.
References

Detecting Human Machine Interaction
Fingerprints in Continuous Event Data

Audrey Reinert

Abstract: Human factors experiments conducted on human operators using a system in a laboratory setting can include direct measures of human performance including physiological measurements and survey results. This precise human performance data can be used to model the environmental and systematic conditions which precipitate a human factors event. However, these results have limited ecological validity since such results difficult to replicate when the system is used outside of laboratory conditions. This is because the methods used to obtain human performance data are disruptive to a task environment. It difficult to obtain a high volume of human performance data because of the cost or the disruptive nature of these measures. Further, this data has yet to be used in an adaptive manner to fit the task to the operator. The same system, when used in a live context, generates a continuous stream of ecologically valid state data. Thousands of individuals can interact with the live version of the system simultaneously to produce a high volumes of state data. What data from live systems gains in ecological validity and volume it sacrifices in the ability to directly measure and record human performance. What measures of human performance which are included in live data are often indirect measures of human performance. The human factors community does not know how to convert indirect measures of human performance data collected from live systems into direct measures of human performance data. The human factors community does not know how continuous system data can be used “adaptively” to predict user delay in responding to a system event. These gaps will be addressed using the following approach. A system will be constructed which produces and records continuous state data. This system will be distributed to participants to generate a large testing dataset. The following analyses will be performed on the data. First, the continuous state data will be analyzed to identify periods of High, Medium and Low Workload. Second, the continuous state data will be analyzed predict a participant’s delay in responding to a malfunction. The proposed research is divisible into seven stages. The first stage entails the development of a Web-based version of MATB-II which will be used to collect the data used for subsequent analysis. The second stage is a pilot study intended to identify the delineations in the data between high, medium, and low workload.
conditions. Once clear delineations in the data have been defined, the MATB-II task will be distributed via Prolific Academic to collect a high volume of continuous system state data. Once a sufficiently high number of participants has completed the experiment, the data will be cleaned to remove incomplete, bad and null responses. Once the data has been cleaned, a series of classification algorithms will be constructed which use derived measures in the data to categorize participants by into the categories of high, medium and low workload. These classification algorithms will be trained on a subset of the data and tested on the remaining subset. Stage six uses a similar process to determine if indirect (derived measures), or direct measures, can correctly classify the delay in responding to a malfunction. A sensitivity analysis will be performed in the seventh stage to determine if the data’s sampling rate, measures of participant expertise and or measures of participant’s experience affect the ability to classify workload or delay. This project tests the following hypotheses. The null hypothesis is that the number of correct classifications of participant workload is equal to chance \((1/3)\). The first hypothesis is that the number of correct classifications of participant workload is greater than chance. The second hypothesis states that the number of correct classifications of participant groupings is equal to chance \((1/N)\) where \(N\) is the number of clusters which best describes the data. The third hypothesis that the number of correct classifications of participant group is greater than chance \((1/N)\). This research makes the following unique contributions to the human factors literature. The first contribution is knowledge of whether direct or derived features in the data can be used to characterize workload. The second contribution is knowledge of whether direct or derived features in the data can be used to predict delay in responding to a malfunction and which features are the most predictive of delay. The third and final contribution is the identification of some of the characteristics of data required for proper characterization, classification and prediction of human machine interaction events in continuous state data.
The new workload of new technologies

Catherine Burns

1 Systems Design Engineering, University of Waterloo
Waterloo, Canada
catherine.burns@uwaterloo.ca

Abstract. Cognitive work analysis (CWA) is a framework that has been used in many settings to describe various aspects of work. This paper outlines how CWA can be used to understand work and workload. The work domain, control task, and strategies analysis can be useful to understand the nature of work, work allocation and workload. Finally, the prediction of work patterns is discussed. Predicting the influence of new technologies on human work is a critical capability for the human factors practitioners of the future.

Keywords: cognitive work analysis, workload, task analysis, function allocation

1 Introduction

We are living in an exciting era of technological change. Just as the development of the internet revolutionized work, future developments in automation and artificial intelligence may create a second revolution. These technologies offer great promise to improve human work but also present some challenges. In the following sections I have introduced four challenges of introducing automation and new technologies into our control rooms. I am considering “control rooms” quite broadly, as these new technologies are also creating a broader spectrum of control rooms. We now have control centres at hospitals, at financial trading companies, and at the operations of smart cities. Indeed, our automated vehicles and automated home technologies are placing the average person in their own custom built control worlds.

The first challenge is the function allocation challenge that our field has worked with for decades. Now, however, functional allocation is between human and technology, with technology covering more functional space and decision making than before. New technologies can provide anything from information analysis, to recommendations to full automation. However, in all
interactions there is still a human “in the loop” though that human can be pushed further and further out of that loop. The supervisory control problems of decades ago are now at a larger scale. Automation is more pervasive, and may be interacting with less trained users such as the drivers of automated vehicles.

The second challenge is the movement of automated technologies to new users. These new users may have less awareness of the need to remain in the loop with these technologies, as evidenced by some of the inappropriate behaviors we have noted with Tesla drivers. Our field study of Tesla drivers indicated that these users were quite comfortable with the autopilot automation and willing to test and challenge the automation by using it inappropriately. Essentially the risk acceptance of these users was very high, whether that was due to their own sense of self-efficacy, the nature of early technology adopters, or unreasonable trust in the product is yet to be determined. Our research has indicated that the people behave differently when working with automated systems. This effect can be attributed to automation complacency or overtrust, but these behaviors need to be understood more clearly. Automation complacency/overtrust are too broad for the range of behavior we have been seeing. We have seen automation experimentation with Tesla drivers, financial risk-taking with automated traders, and lack of results checking with naval operators. Each is a distinctly different behavior that may be a reaction to the particular automation, the individual differences of the users, or the work context. These factors need to be understood more clearly. The third challenge is the necessity of a priori prediction of the effects of the new automation. Meeting this challenge requires a combination of strong human factors work methods and careful well-constructed research that begins to establish generalizable principles of the influence of automation. As automation is applied, its functional behavior must be characterized in detail so that the work performance of the human-automation team can be understood correctly. The fourth challenge is the need to take a human-automation-system perspective. Accepting a situation where the automation is great but the human needs training to “fit in”, or “to stay in the loop” is not acceptable. A correctly designed human-automation system should meet an increased performance level and the automation should augment human intelligence, not push the human into a role where humans are not effective. Correctly designed human-automation includes the user in the loop, in an engaging and healthy work experience from the beginning of the design process. Requiring a user to tap a steering wheel, or perform a series of routine checks to stay in the loop is an indicator that the automation has not been designed with the human in mind. Rather than be discouraged
by these challenges, I believe that these challenges present a new era of exciting exploration of how humans work and how design can become more human-centered. Automation should be seen as a way to advance human objectives and create richer and more rewarding lives for everyone.
Tell me how you kit, I’ll tell you how you think
the use of behavior analyses to optimize parts supplies
in automotive industry

Lisa Jeanson1,2, J.M. Christian Bastien1, Alexandre Morais2, Javier Barcenilla1,

1 Université de Lorraine, PErSEUs, EA 7312, Metz, F-57045, France
   {christian.bastien, javier.barcenilla}@univ-lorraine.fr
2 PSA 2-10 Bd de l’Europe Poissy 78092 Yvelines Cedex09
   {lisa.jeanson, alexandre.morais}@mpsa.com

Keywords: Mental Workload; Cognitive processes; Attention; Performance; Prevention of error.

1. Purpose and methodology

In 2004, the PSA Group, a French automotive manufacturer, developed the PSA Excellent System. This organizational system is based on the Lean Manufacturing principles and aims to optimize vehicle production. One of the pillars of this system is the follow up of a "work standard" designed by the methods engineers [1]. In theory, work standards allow for the balancing of shifts, i.e. the organization of tasks that can be performed by operators within a given time period. The target here is to maintain operators' performance and health. In spite of this approach, errors and complaints on the assembly lines have emerged. In order to understand these phenomena, we carried out a detailed analysis of the operators' activity. For the data collection, we combined several methodologies: hierarchical analysis of the prescribed and actual operators’ tasks, filmed observations and behavioral coding (with The Observer XT[2,3]) [4,5]. Here we present the analysis of the ‘kitting’ activity for vehicles doors assembly as an example of this particular approach. In fact, the production managers of the doors assembly lines in the factory of Sochaux (France) requested our help to improve the layout and workstation design in the kitting area. Kitting is a process often used in mixed model assembly, in which different objects demanding different component are assembled, in
alternative to continuous supply (also known as line stocking). It allows manufacturer to reduce their costs in stock management and parts supply. In this process, the operator (called the ‘picker’) takes components in stores and sorts them in a ‘kit’ driven by an Automated Guided Vehicle (AGV). A kit is defined as ‘... a specific collection of components and/or subassemblies that together (i.e. in the same container) support one or more assembly operations for a given product or shop order’ [6]. That is to say, one kit contains just the components that operators need to assemble one object. Each picker follows one AGV, one AGV can drive several kits (in our case, 4 kits) and each part is put in a dedicated location on the kits. A light system, called ‘pick to light’ assists the operator in choosing the right components in the stores and storing this component in the right kit (kits are differentiated by colors). Even if this process seems simple and mechanical since it does not involve any complex task (to see the light and the color, to take the component, and to put the component in the corresponding kit), many errors occur. In fact, from Mars 9th to April 19th 2018, we reported 91 errors due to kitting in the area we studied (assembly of vehicle doors in Sochaux, France). Those errors create disruptions in assembly lines since operators often have to stop their tasks and leave their workstations to look for the right parts in the kitting area. This questions one of the major benefits put forward for the use of kitting over line stocking: time efficiency [7,8]. We observed and recorded pickers, before coding their behavior to identify mental workload constraints in kitting tasks.

2. Findings

Analysis of the data revealed discrepancies between work standards and actual tasks. The ‘anticipated regulations’, or strategies used collectively by operators to cope with production constraints could explain these differences. For example, when the number of components to kit or the speed of the AGV increases, pickers tent to stop it. Similarly, when the number of components to kit or the speed of the AGV decrease, pickers tent to run in front of the AGV. Then, we noted that pickers often changed the prescribed order of the component to kit, sorting them by colors (thus by kits) and types of components. This could mean that pickers need a logical order of picking to lower the cognitive load imposed by kitting. At last, we observed that pickers tent to pick some components a few seconds before the lights went on. This suggests that instead of using the pick to light to select components, pickers automatically pick them. Whichever strategy pickers choose, its aims at keeping ahead of the AGV. These strategies are costly for the pickers because
it involves operations, steps, reasoning, that are not taken into account in the
design of workstations. This added time pressure forces them to walk and kit
faster, leading them to errors. During our 2 hours observations, we reported 7
situations in which the pickers forgot one or more components and had to
stop the AGV to pick them up. Furthermore, each strategy is also a symptom
of a dichotomy between Lean Manufacturing rules shaping workstation design
and real production constraints. In other words, pickers do not kit how we
think they kit. For example, the variation of task intensity (number of
components to kit and AGV speed) has a direct impact on their activity.
Considering time pressure on workstations, pickers need their operations to be
automated actions to gain time and lower cognitive load. However, to do so,
they need a stable and repetitive environment. That is why each time their
work rhythm tent to change, they have to readapt to that new rhythm. Thus,
they are disturbed and find strategies to cope with that disruption. That’s
also why pickers tent to ritualize the order of the components they kit by
sorting them by colors and types.

3. Conclusion

Observations on workstations allowed us to understand how operators think
on assembly lines and to identify anticipated regulations. We see these
strategies as a symptom of a dichotomy between the Lean Manufacturing
rules shaping workstation design and real production constraints. In this way,
we attempt to help engineers creating environments where operator’s mental
workload is optimized, thus where operators can automatically accomplish
their tasks without putting their health and performance at risk. The ultimate
goal of our study is therefore to improve the PSA Excellent System's
predictive performance models.
References