Identifying how automation can lose its intended benefit along the development process: A research plan

Simone Rozzi[12], Paola Amaldi[1] and Barry Kirwan[2]
[1]Interaction Design Center
School of Engineering and Information Science
Middlesex University
London, UK

[2]EUROCONTROL Experimental Center
Bretigny-sur-Orge CEDEX
France

simone.rozzi.ext@eurocontrol.int , p.amaldi-trillo@mdx.ac.uk, barry.kirwan@eurocontrol.int

ABSTRACT

Motivation – Automation can fail to deliver the target safety or productivity benefit as intended by those managers and designers advocating its introduction. In a safety critical domain this problem is of significance not only because the unexpected effects of automation might prevent its widespread usage but also because they might turn out to be a contributor to incident and accidents. **Originality/value** – Research on failures of automation to deliver the intended benefit has focused mainly on human automation interaction. This PhD research plan aims at characterizing decisions - taken under productive pressure - for those involved in the automation development process, to identify where and when the initial intention the automation is supposed to deliver can drift from the initial idea. **Expected Finding** – The objective is to develop *Anti-Drift Principles* to identify and compensate proactively for possible sources of drift in the development of new automation. **Research Approach** – The research is based on case study and is currently entering Year 2.

Keywords
Resilience Engineering, Design Process, Safety Critical Domain, Air Traffic Control

INTRODUCTION

The Domain Problem: The Failure of Automation to Deliver the Intended Benefit

In Air Traffic Control, like in many safety critical domains, the introduction of new automated tools is motivated by the anticipated productivity or safety benefits in overall system performance that are expected to arise after tool deployment. But actual performance depends on the performance that the tool actually delivers to the operator. Operational evidence has shown that new automation can be used in ways that differ from the way planned by those designers, managers, and regulators who advocated its introduction. This happens for instance when a last safety net is used for productivity purpose (EUROCONTROL, 2002), or a strategic conflict detection tool is used as a conflict probe (Montoya & Mullan, 2008). In all of these cases the initial intended benefit the tool was supposed to deliver is subverted.

The problem with losing the initial design intention is particularly significant in the case of safety critical automation. Not only widespread use of the tool can be hampered, but the tool itself can also turn out to be a critical contributor to incidents and accidents. The mid air collision over the German airspace between a passenger and a cargo aircraft (BFU, 2004) could be regarded as a failure to coordinate safety critical information in the intended way in the presence of an advanced automated application.

Also, the problem is aggravated by that it is only after the occurrence of incidents or accidents and upon completion of the related investigation report, usually few years later, that regulatory bodies are able to produce the remedial safety recommendations. However while the intent behind the investigation is to avoid that similar accidents happen again in the future, the recommendations tend to focus on short term operational improvements - e.g. new training and/or change in working procedures of the tool, software parameterization - rather than major re-design (Kinnersley & Roelen, 2007). After all, automated systems for safety critical domains are immensely complex and their design and implementation process span across decades; thus only in rare cases major redesign is advocated (Kinnersley et Roelen 2007).

So a sensible question is whether it is possible to intervene early on the development lifecycle. After all, waiting for accidents to happen to improve a new automated tool do not appears to be an effective strategy considering its cost in term of human life.

Related works

The literature on human factors and human computer interaction has long recognized that automated tools do not always function ideally (Alberdi et al. 2005). In a seminal paper on the topic, Parasuraman (Parasuraman, 1997) reports on anecdotal evidence and results from various empirical studies to show how automation might fail. The author identifies three ways in which automation can be ineffectively used by the operator: *disuse*, i.e. when the automated tool is not used, as in the case of warning alarms; *misuse*, when human are rely on the automated tool (even when information is incorrect), rather than on their own judgment; and *abuse,* when functions are automated "without the due regard to consequences for human…performance and operator's authority" (Parasuraman, 1997).

Automation failure to deliver the intended benefit has been studied from a cognitive engineering perspective mainly in socio-technical domains like Aviation (Cummings, 2004), ATC Nuclear Power Plants (Norros, 2004) and Health Care (Vikkelsø, 2005). Such studies have highlighted a number of automation side effects and shown that there is a significant difference between commonly held assumptions on the impact of a new automation on human and system performance and the actual impact that the technology has on the people who have to do the actual work (Sarter, Woods, & Billings, 1997; Vikkelsø, 2005). Examples of such side effects are:

a. Uneven redistribution of workload;
b. Increased complexity of the task;
c. Introduction of new form of errors;
d. Introduction of new coordination dependencies;
e. Indirect Effect on workload of other co-workers.

Overall such unwelcome effects can result in lack of trust, misuses, or disuse; also they are difficult to predict (Parasuraman, 1997), as they arise from the complex interaction that the new tool and the associated procedures will establish with the operational context in which the tool/procedures are embedded.

Other works have looked more specifically at the so called "alarm problem", i.e. the failure to respond to alarm. Woods (1995) notes that evidence from field studies, accident investigation, design reviews, and mathematical models shows that humans can have difficulties to prioritize and respond to abnormal conditions despite the availability of alarm systems (Woods, 1995). Factors that contribute to misuse use of alarm include nuisance alert; ambiguous or underspecified alarm messages etc. Research efforts have focused on finding ways to improve perceptual aspects of alarm, i.e. how to make the physical properties of the alarm – mainly noise, frequency, and color – more salient and distinguishable from the background. However this approach appears as limited in scope as humans react to the significance and meaning that an alarm brings on a specific situation, rather than physical characteristics of signal display only (Norros, 2004; Ponomarenko, 2004; Woods, 1995; Xiao et al. 2004).

Specifically to Air Traffic Control, Bolic & Hansen (2005) have investigated automation usages from an innovation diffusion perspective. They investigated how a conflict detection tool is being adopted and used in US is actually used by air traffic controllers. The authors beyond finding difference of use across centers; across teams of the same center noted that unplanned uses emerged. They concluded that the process of adopting new automation is faster when advantage of use is clearly perceived, and slows when controllers have to change their working practices for accommodating the new tool. The study however brought little insight on how to anticipate unpredicted usages of automation.

Other authors have underlined the later intervention of human factors and safety specialist along the design process of new automation, as cause of misfit between operators and automation. Despite early and continuous involvement of these people from initial design to operational testing is necessary to guarantee the success of the tool (Cardosi, 1998), often these people have access to the design process too late in the process, when many design decisions which have a significant impact on tool usage have been made already, thus leaving little scope for major system improvements (Leveson et al., 2001).). Interestingly an analysis of several accidents occurred in the aviation and nuclear industries, Kinnersely and Roelen (2007) had found that 50% of accidents have their root causes in errors and misconceptions occurring at the design stage the development cycle. This result correlated well with an earlier accident analysis, which identified 59% of incident as having they, root cause in inadequate design (HSE1995). Overall, this evidence suggest to dig deeper into the different roles and organizational processes which sustain the development of new automation so to understand where the initial intention the tool could deliver can be lost along the process.

PROPOSED APPROACH

The present work aims at studying the emergence of failure of automation to deliver the intended benefit as function of holes in the tool development lifecycle. The goal is to identify where the initial intention can be lost along the design

and implementation process. This will make it possible to develop future automated tools by seeking proactively for areas where deviation from initial intention can arise. This approach is informed by a Resilience Engineering perspective and will be outlined in the next section.

RESILIENCE ENGINEERING

Safety depends on creating foresight

According to a Resilience Engineering perspective safety depends on how organizations, groups, and individuals are able to create foresight thus being able to anticipate risk before failure occurs. Conversely, failure is considered as the absence of that ability (Hollnagel, Woods, & Levenson, 2006). Such position is based on the observation that a fundamental trait of high reliability organization is the ability to invest effort to anticipate and plan for unexpected events and future surprises (Weick, 1987; Woods, 2004). Such organizations do not take evidence of past successes a reason for confidence in future successes; rather they constantly challenge their model of risk on the basis that knowledge held by the organization need continuous improvement in face of the continuous hazards inherent in its works process (Hollnagel et al., 2006).

Achieving resilience by helping organizational decision makers to balance pressures

One way to achieve resilience is to help organizational decision makers - such as those people engaged in the implementation/design of new tool - to balance both intense and persisting commercial pressure against decreased safety and increased risk (Hollnagel 2004).

Following the analysis of several accidents across a variety of domains, Rassmussen (1997) has found that accident causes are not rooted just in human error and technical failures and their combination; rather they are also rooted in the drift of the global behavior of the organization under strong pressure towards productivity, in the global context of a competitive environment. This indicates to take into considerations decisions taken by many actors involved in a given work process, in their normal working conditions, but exposed to a persisting competitive pressure. Rasmussen adopts an envelope to delimit the space of acceptable performance of people at different organizational levels which is bounded by administrative, functional and safety related constraints (Figure 1). When pressure is applied along one of the boundaries of the model it will tend to push performance toward the opposed boundaries. For instance if work speed is reduced the system will suffer from reduced productivity thus getting close to the boundary of economic failure. Conversely if productive pressure is increased performance will tend to get close to safety boundaries.

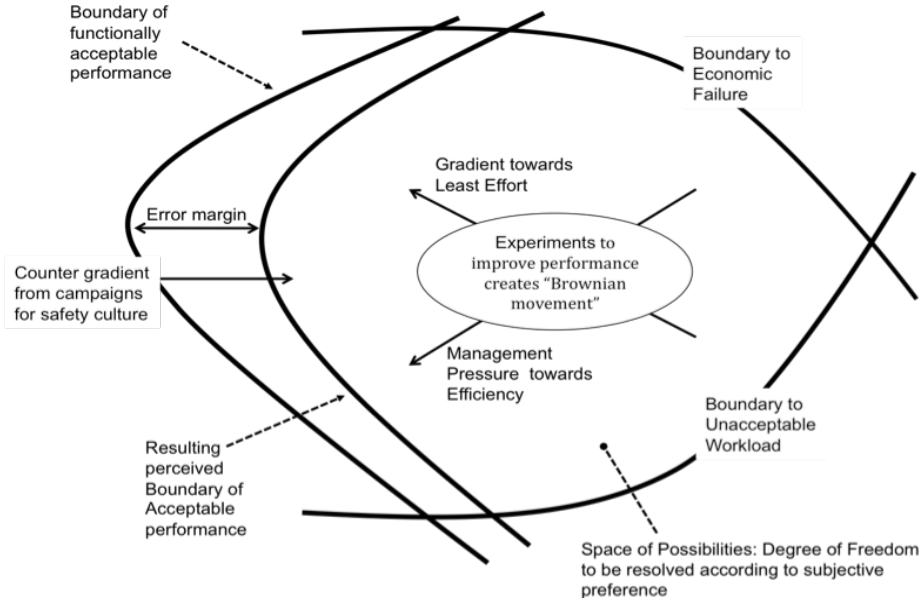

Figure 1. Rasmussen's space of acceptable performance (1997).

In conclusion, applying this line of reasoning along the automation design process implies to hypothesize that initial design ideas undergoes some modification since its early conception. Therefore two hypotheses are explored in this PhD research:

> (a) "Organizational Drifts" from the original safety intentions, while reflecting the need to compromise between competing stakeholder interests (Rasmussen, 1997) might later cause the "front end" to search for adaptations – unplanned usages - to less than ideal conditions for the tool effectiveness.

(b) Unplanned, "situated" usages of the tool might affect the original safety benefit the tool was intended to deliver;

If both hypotheses receive support, then new specific Resilience Engineering principles can be identified along with methodological implications for the design cycle of new/improved tools and innovations for the ATM system.

METHODOLOGY

At present the work has completed an extensive review of the literature - available in Rozzi et al. (2008) - which contributed to define the research approach and the data requirements reported here; thus completing PhD Phase 1. From this point onward the work is entering Year 2 and the following two phases of work remain to complete.

Anti Drift Principles Development, in the context of MSAW.

This phase consists in investigating the development process of an automated alarm called Minimum Safe Altitude Warning System (MSAW) - which is detailed in a later section - and will cover the whole duration of PhD Year 2. In this context the objective is (i) to trace the path from the initial conception phase where needs, ideas, views are expressed by the various stakeholders involved in the definition phase, (ii) to the phase where the alarm is implemented and unplanned usages emerge. This phase has been broken down in three blocks of work:

2.1 – Analysis of MSAW design/implementation process. This block focuses on tracing the actual development process of MSAW, to identify point of drift from initial safety intent, and motivation behind the drift;

2.2 – Analysis of MSAW usages. This block will look at controller perspective as embedded in the actual usages of the MSAW, i.e. how the MSAW is used and why it is used in a particular way;

2.3 – Integration. This block will map the results of Blocks 1 and 2 in search for classes of drifts that are critical during the development process as they will later impact on the adaptation of the tool in the operational environment. This phase is expected to result in the formulation of anti drift principles.

Main data sources for Year 2 include interviews with stakeholders involved in MSAW definition and set up, air traffic controllers engaged of the tool, and accident analysis. Table 1 details (i) objective, (ii) duration, (iii) data collection tasks, and (iv) expected outcomes for each of the three blocks of Year 2.

Anti-Drift Principles validation

The objective of this phase is to validate the formulated anti-drift principles. It is anticipated that validation of the principles will happen by apply them to an on going implementation of MSAW or in the context of another application, yet to be defined, to make sure safety intent is not lost during the design/implementation phase. The application chosen for the validation will affect the validity of PhD results.

MSAW SYSTEM DESCRIPTION

MSAW is an automated safety net that was developed and introduced in the 70s in US as a protection against a type of incident known as Controlled Flight Into Terrain accidents (CFIT). CFIT are one of the most serious aviation killers since the emergence of civil aviation and occur whenever a flight suffering no engine failure or other equipment impairment is flew into terrain, sea or other obstacles, without the crew being aware of the impending collision (Philips, 1999). CFIT occurs especially during landing in proximity of the airport, however they might occur also during en route phase of flight. Thus MSAW system can be found in tower, approach and en route control centers.

MSAW software has to alert air traffic controllers whenever aircraft descent (or are predicted to descent) below a minimum predetermined safe altitude. The system receives altitude information from tracked aircraft and compares it against terrain database. If the aircraft altitude is below or predicted to be below MSAW system generate a visual alert on the controller radar display and/or a aural alert in the control room. Upon generation of MSAW alert, the controller has to identify the aircraft that triggered the alert and inform the aircraft crew of the hazard (Greenwell, Strunk, & Knight, 2004). To note that responsibility to maintain separation from terrain remains with the flight crew. Also the MSAW does not suggest any resolution maneuver.

The MSAW has been chosen as a suitable application for two reasons. Firstly despite being an effective safety barrier in principle, the ineffectiveness of MSAW to result in the controller issuing an advisory alert to the crew has been cited as a contributory cause in a number of accidents since its early deployment (Rosenker, 2006). During these accidents the tool either generated the alert at appropriate time but was unnoticed by the controller, or failed to alert the controller at appropriate time, thus losing its intended benefit. Also, the analysis of these accidents indicates that root causes for poor tool performance can be traced to the design of the tool and its adaptation to local operational conditions which are unique for each control center (NTSB, 2006). The second reason is more pragmatic: Part of this research will study MSAW development process in US, and this can offer some valid lessons to drawn that can inform and optimize the deployment of MSAW in Europe.

Table 1. Workplan for PhD Year 2

Block	**2.1** **Analysis of MSAW design process**	**2.2** **Analysis of MSAW usages.**	**2.3** **Integration**
(i).Objective	This block will look at the evolution of MSAW concept. Main objectives are to: (a). Trace the development of the MSAW application from the initial need, through concept, requirements definition phase and implementation; (b). Identify along this process points where the concept idea has drifted from initial managers' and engineers' idea; (c). Understand (i) the motivation behind the drift, e.g. aggressive schedule, competing interests, and (ii) their acceptance criteria, which ultimately depends on the underlying decision model, e.g. historical success.	This block will study unexpected adaptation associated to the usage of MSAW in operational settings. More specifically it will study how the MSAW is used by controllers, why it is used in a particular way, and whether such (under/over)use/s differs from the usage/s as intended by MSAW designers.	The objective of this block is to map the (i) drifts observed along the design process, their associated motivations and acceptance criterion, with the (ii) associated MSAW related unplanned usages found in the operational environment.
(ii).Duration	Jannuary – March 2009	April – June 2009	July – December 2009
(iii).Data Collection Task	(T1). Interviews with managers and engineers involved in MSAW development; (T2). Attendance of project meetings and Review of relevant documentation, which depending on availability might include minutes and presentations from meetings, requirements documents, validation documents, relevant e-mails.	(T3). Analysis of Incidents and Accidents reports related to MSAW; (T4). Field Observations and Interviews with controllers/supervisors in Centre/s where MSAW is in use.	(T5). Analysis and integration of finding from previous activities and deliverable preparation
(iv).Expected Ouctomes	- Description of MSAW development process, from initial need to conception and implementation; - Taxonomy of Drifts, and Motivation behind the Drift, and Acceptance criteria.	-Taxonomy of unplanned adaptations associated to MSAW.	- Model of the design and implementation process as influenced by drift and adaptation; - Resilience engineering "Anti-drift" principles.

CONCLUSION

In conclusion, the emergence of unplanned usages associated to new automation appears still as a persisting issue in safety critical domains. Existing literature has highlighted side effects of automation and – in the case of alarm – research has focused on perceptual and informative aspects of tool design. The present work, while leveraging on a Resilience Engineering framework, explores the problem by investigating unintended usages as function of decisions happening during the design/implementation process. No studies in Air Traffic Control and Aviation have been found based on such perspective, despite evidence across a range of industries indicates that accidents often have their roots cause in decisions taken along the design process.

ACKNOWLEDGMENTS

This study is part of an ongoing doctoral research programe sponsored by EUROCONTROL Experimental Centre, Bretigny-sur-Orge CEDEX, France. The authors are grateful to Ben Bakker and Hans Hagerman for the insightful discussions on development process of safety critical automation in ATC. The views expressed herein do not necessarily reflect the official views or policy of the Agency.

REFERENCES

Alberdi, E., Ayton, P., Povyakalo, A. A., & Strigini, L. (2005). *Automation bias and system design: a case study in a medical application*. Paper presented at the IEE People & System Symposium.

BFU. (2004). *Investigation Report* (Accident Report No. AX001-1-2/02): German Federal Bureau of Aircraft Accident Investigation.

Bolic, T., & Hansen, M. (2005). *USER REQUEST EVALUATION TOOL (URET) ADOPTION AND ADAPTATION, THREE CENTER CASE STUDY*. Paper presented at the 6th Usa-Europe ATM Seminar.

Cardosi, K. (1998). Human factors Lessons Learned in the Design and Implementation of Air Traffic Control Systems. *The Controller, First quater*, 11-15.

Cummings, M. L. (2004). *Automation Bias in Intelligent Time Critical Decision Support Systems*. Paper presented at the AIAA 1st Intelligent System Technical Conference.

EUROCONTROL, C. (2002). RITA 2: Replay Interface for TCAS Alerts.

Executive, H. a. S. (1995). *Out of Control, Why control systems go wrong and how to prevent failure*. London, UK: HMSO.

Greenwell, W., Strunk, E., & Knight, J. (2004). *Failure Analysis and the Safety-Case Lifecycle*. Paper presented at the IFP Working Conference on Human Error, Safety, and System Development.

Hollnagel, E., Woods, D., & Levenson, N. (2006). *Resilience Engineering: Concepts and Precepts*: Ashgate.

Kinnersley, S., & Roelen, A. (2007). The Contribution of Design to Accidents. *Safety Science 45*(1-2), 31-60.

Leveson, N., de Villepin, M., Daouk, M., Bellingham, J., Srinivasan, J., Neogi, N., et al. (2001). *A Safety and Human-Centered Approach to Developing New Air Traffic Management Tools*. Paper presented at the ATM2001.

Montoya, F., & Mullan, C. (2008). *Validation of a European Gate to Gate Operational Concept for 2005-2010* (Technical Report No. D4.4.4. Report prepared for Gate to Gate, European Founded project, Programme rference number GTG-44-EURO-SIM-D444-V0100).

Norros, L. (2004). *Acting Under Uncertainty. The Core Task Analysis in ecological study of work*. Espoo: VTT Technical Research Centre of Finland.

NTSB. (2006). *Safety Recommendation A-06-44 through -47*. Washington, D.C.: National Transportation Safety Board.

Parasuraman, R. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39*(2), 230-253.

Philips, R. O. (1999). *Descriptions of flight paths for selected controlled flight into terrain (CFIT) Aircraft Accidents, 1985-1997*: US Departement of Transportation Research and Special Programs Administration Volpe National Transportation Systems Centre.

Ponomarenko, V. A. (2004). The significance of theoretical concepts in activity theory for applied research in aviation. *Theoretical Issues in Ergonomic Science, 5*(4), 297-312.

Rasmussen, J. (1997). Risk management in a dynamic society: A modeling problem. *Safety Science, 27*, 183-213.

Safety Reccomendation A-06-44 through -47, (2006).

Rozzi, S., Amaldi, P., & Fields, B. (2008). *Task Analysis and Contextual Models of Controllers Activity for Interactive System Design. A literature review.* Paper presented at the 7th EUROCONTROL Innovative Research Workshop & Exhibition, Bretigny-sur-Orge, France.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation Surprises. In G. Salvendry (Ed.), *Handbook of Human Factors & Ergonomics*: Wiley.

Vikkelsø, S. (2005). Subtle Redistribution of Work, Attention and Risks: Electronic Partient Records and Organizational Consequences. *Scandinavian Journal of Information Systems, 17*(1), 3 - 30.

Weick, K. E. (1987). Organizational culture as source of high reliability. *California Management Review, 29*, 112-127.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics, 38*(11), 2371-2393.

Woods, D. D. (2004). Creating Foresight: Lessons for Enhancing Resilience from Columbia. In B. Starbuck & M. Farjoum (Eds.), *Learning from the Columbia Accident*: Blackwell.

Xiao, Y., Seagull, J., F., Nieves-Khouw, F., Barckzak, N., & Perkins, S. (2004). Organizational-Historical Analysis of the Failure to Respond to Alarm Problems. *IEEE Transactions on Systems, Man, and Cybernetics, 34*(6), 772-778.