LES CAHIERS DE LA SÉCURITÉ INDUSTRIELLE

2011-09

CONTROL AND ACCOUNTABILITY IN HIGHLY AUTOMATED SYSTEMS

NeTWork'2011

A summary of workshop discussions, prepared by Eric Marsden

FonCSI
FONDATION POUR UNE CULTURE DE SÉCURITÉ INDUSTRIELLE

The *Foundation for an Industrial Safety Culture* (FonCSI) is a french public-interest research foundation created in 2005. It aims to:

- to undertake and fund research activities that contribute to improving safety in hazardous organizations (industrial firms of all sizes, in all industrial sectors);
- to work towards better mutual understanding between high-risk industries and civil society, aiming for a durable compromise and an open debate that covers all the dimensions of risk;
- to foster the acculturation of all stakeholders to the questions, tradeoffs and problems related to risk and safety.

In order to attain these objectives, the FonCSI works to bring together researchers from different scientific disciplines with other stakeholders concerned by industrial safety and the management of technological risk: companies, local government, trade unions, NGOs. We also attempt to build bridges between disciplines and to promote interaction and cross-pollination between engineering, sciences and the humanities.

Foundation for an Industrial Safety Culture
A public-interest research foundation
http://www.icsi-eu.org/

6 allée Émile Monso – BP 34038
31029 Toulouse cedex 4
France

Telephone: +33 534 32 32 00
Fax: +33 534 32 32 01
Email: contact@icsi-eu.org

Editor: Institut pour une Culture de Sécurité Industrielle
Association de loi 1901
http://www.icsi-eu.org/

Introduction

Context

NeTWork (New Technology and Work) is an international, interdisciplinary group of academics, regulators and practitioners with the objective to provide concepts and methods for addressing individual, organizational and societal risks created by technological development, for evaluating the state of the art of technology management, regulation and risk control and debating the way forward. The founder of the core group, responsible for the management of NeTWork, was Prof. Bernhard Wilpert, from Technische Universität, Berlin. The management is now in the hands of Prof. Mathilde Bourrier (University of Geneva), Prof. Gudela Grote (ETH Zürich) and Dr Babette Fahlbruch (TUV Nord).

Over the past 28 years NeTWork has held annual workshops relating to the overall theme of new technologies and work. Workshops have covered a wide range of topics that included human error, accident investigation, training, distributed decision making and management. While the original activities of NeTWork began with a wide coverage of sub-themes, recent preoccupations have focused more specifically on a theme of great scientific and social significance: Safety of high technology systems and the role of human contribution to either failure or resilience of hazardous systems.

The core values of NeTWork are as follows:

- creating opportunities for co-constructive interdisciplinary exchange, integrating social science and engineering/natural science approaches;
- fostering dialogue between domain experts and novices;
- aiming for high impact output based on scientific standards;
- promoting comparative studies across disciplines and industrial sectors;
- considering emerging concepts and revisiting old ones;
- taking account of ethical concerns and fairness;
- providing for intense and genuine social interaction, and creating a setting that inspires trust, respect and constructive criticism.

More information concerning NeTWork can be found on its website:

`http://www.network-network.org/`

The 2011 workshop: Control and accountability in highly automated systems

The topic of the September 2011 workshop, which was held over two days in the Sorèze abbey near Toulouse, France, was *Control and accountability in highly automated systems.* The workshop was organized by Gudela Grote (ETH Zürich) and Johannes Weyer (TU Dortmund), and primarily funded by the *Foundation for an Industrial Safety Culture* (FonCSI).

The workshop brought together academic participants from multiple scientific disciplines and practitioners in several fields. It is the rich discussion between these participants which gave rise to the present document.

Michael Baram	Law	Boston University	USA
Corinne Bieder	Safety and human factors	Airbus	France
Daniel Boos	Telecommunications	Swisscom	Switzerland
Babette Fahlbruch	Social psychology	TU Berlin	Germany
Sylvie Figarol	Psychology	DGAC	France
Robin Fink	Sociology of technology	TU Dortmund	Germany
Tor Olav Groetan	Resilience engineering	SINTEF	Norway
Gudela Grote	Work and organizational psychology	ETH Zürich	Switzerland
Andrew Hale	Safety science	Delft University & Hastam	England
Gisela Hürlimann	History of science	University of Zürich	Switzerland
Toshiyuki Inagaki	Risk engineering	University of Tsukuba	Japan
Chris Johnson	Computing science	University of Glasgow	Scotland
Barry Kirwan	Safety and human factors	Eurocontrol	France
Fabian Lücke	Technology studies	TU Dortmund	Germany
Marc Mölders	Technology studies	TU Dortmund	Germany
Gilles Motet	Risk and dependability	FonCSI	France
Stefan Müller	Law and new technology	TU Berlin	Germany
Thomas Sheridan	Mechanical engineering and applied psychology	MIT	USA
Johan Stahre	Production systems	Chalmers University	Sweden
Neville Stanton	Human factors and ergonomics	University of Southampton	England
Etienne Tarnowski	Test pilot	Airbus	France
Johannes Weyer	Sociology of technology	TU Dortmund	Germany
Ernst Zirngast	Risk analysis	Swiss Re	Switzerland

This document

This document is a **summary of the main issues** discussed during the course of the workshop, focusing in particular on questions related to accountability. It aims to provide an overview of the questions most relevant to decision-makers and other interested parties, together with pointers to relevant academic literature. The summary of discussions was produced by Eric Marsden (FonCSI), with valuable feedback from workshop participants.

The papers presented at the workshop will also be published, in the form of a journal special issue or a book, providing more detail and a broader range of views on the issues presented.

We are interested in your feedback! Please send any comments or suggestions for improving this document via email to `cahiers@icsi-eu.org`.

1

I'm sorry Dave, I'm afraid I can't do that

1.1 Context

automation to improve system performance

Effective automation has evolved to best benefit from human operators' strengths (discernment, judgment, adaptability, strategic decision-making) while compensating for their weaknesses (reduced performance under fatigue, decreasing performance on repetitive tasks, difficulty in managing multiple cognitive tasks in parallel). However, automation has not always been effective, and it has always raised questions of whether, and if so how much and when, the automation should take over the responsibility for system performance and safety from humans in the system. In complex systems, improvements in the performance of technology combined with increasing operating pressure have led to increases in the general level of automation in many large safety-critical socio-technical systems (transport, chemicals and refining, energy, *etc.*) (*cf.* Figure 1.1).

Figure 1.1 – *Levels of automation of decision and action selection (adapted from [Sheridan and Verplank 1978])*

Automation is generally implemented with the intention of **improving system performance** (productivity, safety), and most often it does achieve these goals. However, changes in the level of automation lead to several far-reaching changes, whose consequences need to be analyzed. The following issues were discussed during the NeTWork'2011 workshop:

- the far-reaching delegation of authority to more complex automated systems calls for a **redefinition of the role of the human operator**, and for new concepts of shared authority or adaptive automation;

- the shift in control from human operator to automation must be accompanied by a **shift in accountability** from frontline staff to system **designers**, and more generally to the operating organization.
- in some sectors such as air transportation and energy production and distribution where there is a movement from traditional centralized hierarchical control towards more decentralized self-organization, new modes of **governance** may be needed.

These issues are discussed in more detail in the following sections.

1.2 Monitoring automation: an impossible task?

unexpected drawbacks of automation

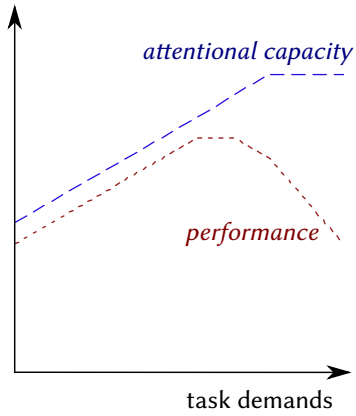
There are a number of known problems associated with human supervisory control[1] of automation [Parasuraman et al. 2000]:

- **Loss of situation awareness**: automation of decision-making functions may reduce the operator's awareness of the system state and of changes to the environment. Humans tend to be less aware of changes in environmental or system states when those changes are under the control of another agent—whether that agent is automation or another human—than when they make the changes themselves. Also, if the automation consistently and repeatedly selects and executes decision choices in a dynamic environment, the human operator may not be able effectively to maintain a good "picture" of the system state, because he is not actively engaged in evaluating the elements of information that led to the decision.
  There is evidence that this effect becomes stronger as the monitored automated subsystem becomes more reliable: the operator's *trust*[2] in the automated system increases, leading to an increase in *reliance* on the system, which in turn leads operators to spend less effort in maintaining an accurate and complete mental representation of the aspects of the situation that are controlled by the automation.
- **Complacency**: if automation is highly but not perfectly reliable in making decisions, the operator may not monitor the automation and its information sources and hence may fail to detect the occasional times when the automation fails. This effect of "over-trust" or "complacency" is greatest when the operator is engaged in multiple parallel tasks.
- **Skill degradation**: if a decision-making function is consistently performed by automation, the human operator becomes progressively less skilled in performing that function, by lack of practice.
- **Loss of attention**: while automation is usually implemented with the intention of **reducing operator workload** and thereby improving performance, current evidence suggests that excessively low mental demands are actually **detrimental to performance**. Cognitive task performance degrades as workload decreases (loss of attention or vigilance, *cf.* figure 1.2) and people are not able to react appropriately with sufficient speed (it takes several seconds to regain full attention, and human decision-making in complex situations which are unexpected and off-normal is relatively slow[3]). This phenomenon has been called *malleable attentional resource pools* [Young and Stanton 2002b]. Indeed, underload is possibly of greater concern than overload, as it is more difficult to detect.

For these reasons, human monitoring of highly reliable automated systems is in a certain sense an **impossible task**.

When automation is changed, human operators adapt to the consequent changes in their activity in various ways (for example, the introduction of ABS systems in automobiles has led

[1] Supervisory control [Sheridan 2006] means that in a person's interaction with automated subsystems, one or more human operators are intermittently giving orders to and receiving information from a computer system that interconnects through actuators and sensors to the controlled process or task environment.
[2] Trust can be seen as a form of substitute for control, a mechanism which allows people to cope with uncontrollable uncertainties or complexity [Luhmann 1979].
[3] Human response times follow a log-normal probability distribution, rather than the better-known normal or "gaussian" distribution commonly encountered for natural phenomena. This means that the response-time distribution is more strongly skewed to the right than the system designer may have anticipated (operator response time may be much larger than anticipated).



Our attentional capacity is not constant, but varies as a function of task demand. When the task requires little supervision, our attentional capacity is low (boredom ...). It increases as the task becomes more demanding, until reaching a plateau of maximal capacity.

Performance decreases as task demands overtake the operator's attentional capacity, and also decreases when the task requires

Figure 1.2 – *Relation between task demands and performance under a malleable attentional resources model, after [Young and Stanton 2002a]*

certain drivers to adopt more dangerous driving styles, a phenomenon known as *risk compensation*). These effects of the human adapting to his new activity must be monitored to ensure that a sufficient level of safety and efficiency of the human-automation team is maintained. Unfortunately, we lack metrics or performance indicators to characterize these aspects of system performance[4].

performance indicators

The crash of Turkish Airlines flight 1951 in 2009 illustrates the difficulties that human operators encounter in monitoring and understanding the performance of the autopilot system, despite their high level of training.

Crash of Turkish Airlines flight 1951

Turkish Airlines Flight 1951 was a passenger flight which crashed during landing to Amsterdam Schiphol airport in February 2009, killing nine passengers and crew, including all three pilots. The aircraft, a Boeing 737-800, crashed into a field approximately 1.5 kilometres north of the runway, breaking into three pieces on impact, without catching fire.

The aircraft's automation was a major contributor to the crash. The aircraft had a faulty radio altimeter, which suddenly reported an invalid negative altitude during the descent. This caused the autothrottle to decrease engine power to idle during approach[5]. The crew noticed this too late to take appropriate action (pointing the nose down and increasing engine power) and recover the aircraft before it stalled and crashed [Dutch Safety Board 2010].

The pilots were very likely not aware that the wrong altitude measurement directed autothrottle action. Indeed, the manuals for use during the flight did not contain any procedures concerning errors in the radio altimeter system. In addition, the pilot's training did not include any detailed system information that would have allowed them to understand the significance of the problem.

The pilots did not detect early signs of their low airspeed (flashing airspeed indication on the primary flight displays, artificial horizon indicating that the nose of the plane was too high). When the stall warning triggered, the first officer (who was inexperienced on this route and under supervision by the commanding officer) manually increased thrust to increase speed. However, the commanding officer (an experienced former Turkish Air Force pilot) also took control of his stick, interrupting the first officer's action on the thrust. The autothrottle—which the pilots were unaware was engaged—pulled back the throttle. By the time the captain noticed that thrust was reduced, it was too late to recover from the stall.

Food for thought in your area of work

- ▷ In your organization, how do you manage the risks of operator complacency and progressive skill degradation related to high-performance automation?
- ▷ Is operator underload, and the consequent decrease in vigilance, managed in the systems that you are involved in?
- ▷ Can you think of situations in the systems you work with where operators may misread certain signals raised by the automation?
- ▷ Does the control-room culture in your organization allow inexperienced operators to challenge the actions of more senior personnel, to improve their understanding of the system, and possibly to warn of an abnormal system state?

1.3 New modes of governance in partly decentralized systems

The safety of large socio-technical systems is increasingly dependent on the coordination between multiple human and non-human actors who belong to different professions and organizations, and may be in different geographical locations. These subsystems and operators may have differing objectives, operating constraints and underlying beliefs concerning safety. Air transportation is an example of a sector which is slowly evolving from a centralized view of safety with hierarchical control (pilots obeying instructions from ATC[6]) to more decentralized modes of operation which will require more cooperation between parties[7].

decentralized operation

This trend towards increased decentralization is also illustrated by the mid-air collision in Überlingen in 2002.

Überlingen mid-air collision

In July 2002 a Bashkirian Airlines Tupolev passenger jet and a DHL Boeing 757 cargo jet collided in mid-air over the town of Überlingen in southern Germany. All 71 people on board the two aircraft were killed.

The German accident investigation found that the accident was caused by shortcomings in Swiss air traffic control (which was supervising the flights at the time of the accident) and by ambiguities in the use of TCAS[8], the on-board collision avoidance system.

At the time of the accident (near midnight), a single air traffic controller was handling the airspace (his colleague was resting in another room, a practice tolerated by management though against regulations), and was covering two workstations. The controller was occupied with another aircraft arriving late at Friedrichshafen, and maintenance on the phone system obliged him to use a less convenient mode of communication[9]. He did not spot the two planes' collision course until a minute before impact.

The controller ordered the Tupolev jet to descend to avoid collision. Seconds after the Russian crew initiated the descent, however, their TCAS instructed them to climb, while at about the same time the TCAS on the DHL cargo instructed the pilots of that aircraft to descend. Had both aircraft followed those automated instructions, it is likely that the collision would not have occurred. The cargo plane initially followed the TCAS instructions and initiated a descent, but could not immediately inform the controller due to the fact that he was dealing with another flight. The Russian pilot on the Tupolev disregarded the TCAS instruction to climb and instead began to descend, as instructed by the controller; both planes were now descending. Unaware of the TCAS-issued alerts, the air traffic controller repeated his instruction to the Russian jet to descend, giving the Tupolev crew incorrect information as to the position of the DHL plane. The aircraft collided at an altitude of around 10 km.

[4]Certain authors argue that this is because concepts such as situational awareness and complacency—though intuitive to many people—are under-specified, cover too broad a range of underlying conditions, are unmeasurable and thus unfalsifiable as hypothesized contributing factors to accidents [Dekker and Hollnagel 2004].
[5]This "retard" mode is designed automatically to decrease thrust shortly before the aircraft touches down; when engaged it displays RETARD on the primary flight display.
[6]ATC: Air Traffic Control
[7]Work on new air traffic management paradigms is being undertaken in large R&D projects called SESAR in Europe and NextGen in the United States of America.
[8]TCAS: Traffic Collision Avoidance System

conflicting orders

The accident raised questions as to how pilots should react when they receive **conflicting orders** from the TCAS and from ATC[10]. The TCAS is designed assuming that both crews will promptly follow the system's instructions. The manual clearly states that TCAS should always take precedence over any ATC commands: If an instruction to manœuvre is received simultaneously from an RA (resolution advisory, the instruction issued by the TCAS) and from ATC, then the RA should be followed. However, this training had not been given to the Russian pilots, whose safety model was based on the assumption that ATC has a global view of all objects in the sky, and is therefore best placed to instruct pilots.

Pilots are not required to notify ATC prior to responding to an RA. This manœuvre does not require any ATC clearance, since TCAS takes into account the position of all other aircraft with transponders in the surrounding area. This illustrates a change in the governance from a centralized model (ATC has a global view of all objects) to a partly decentralized model allowing local optimizations [Weyer 2006]. Who is responsible in the case of an incident?

Technical changes to TCAS. Prior to this accident occurring, a change proposal for the TCAS II system had been issued. This proposal would have detected that one aircraft was not following the initial TCAS advisory, and created a "reversal" of the original warning, asking the DHL plane to climb and the Tupolev crew to descend (this would likely have avoided the Überlingen collision; it is now implemented). Additionally, an automatic downlink for the TCAS—which would have alerted the air traffic controller—had not been deployed worldwide at the time of the accident.

Automation creep. Analysis of pilot responses to TCAS resolution advisories shows that in some cases pilots fail to follow advisories (either climbing when instructed to descend, or descending when instructed to climb, or not responding at all), which could lead to accidents. On new Airbus aircraft, TCAS resolution advisories are now input to the autopilot, reducing the need for pilot action. This "automation creep", while positive for safety in the medium term, generates a long-term risk of lowering pilots' skill level:

- as their active implication in piloting the aircraft decreases, their level of attention and situational awareness tend to decrease, as explained in §1.2;
- as the importance of pilots' role in civil aviation decreases, the level of prestige and attractiveness of the occupation tends to decrease, possibly leading (over time) to a decrease in the skills and qualifications of the people choosing this job.

deskilling

Governance patterns. As in all hazardous industries, airplane manufacturers and airlines put in place operational experience feedback systems to learn from incidents and accidents and improve safety. Accident reports issued by accident investigation boards sometimes lead to recommendations of changes to aircraft design, equipment, maintenance or operating procedures. From a political point of view, it can be easier for an accident investigation board to suggest changes than for the manufacturer or an airline, who may fear that issuing a change recommendation amounts to admitting liability for the accident.

Certain industries are characterized by weak regulatory oversight (road transportation, computer security, …), which—among other consequences—may reduce opportunities for industry-wide learning.

Food for thought in your area of work

- In the systems that you are involved with, are there any areas where modes of control and modes of governance are evolving (perhaps silently), due to new technology or organizational changes? Does this evolution introduce new risks?
- In your sector, is there an organization with sufficient independence and competence to raise concerns related to safety, which other organizations might attempt to handle in a less transparent manner?

[9]The faulty communication system also prevented controllers from Karlsruhe from warning him of the impending problem.

[10]It also highlighted issues related to operator's personal responsibility for errors, since the air traffic controller on duty at the time of the accident was later stabbed to death by the father of three of the victims of the crash.

1.4 Moving accountability upstream

Accountability requires **understanding** and **possibility of acting** [Boy and Grote 2011]. When we blame technology or automation for contributing to an accident, we are not blaming the inanimate electronic circuits, but rather a "blunt-end" worker instead of the "sharp end" operator:

- the system designer, for insufficient foresight (ability to anticipate the sequence of events that led to the accident) or for a design error;
- the manufacturer, for defects in the production process;
- the maintainer, for inadequate maintenance;
- the managers, for deciding that the system should be operated (potentially, despite warnings that the level of safety was low).

In response to this increasing pressure on accountability, designers and managers should adapt their work processes in various ways:

- increased **transparency** on available design options, their advantages and disadvantages for different stakeholders;
- increased attention to **documenting design choices** and the criteria used by decision-makers (state of scientific knowledge at the time the decision was made, external constraints such as cost pressures, *etc.*);
- more attention paid to **learning from experience** in design successes and design failures (though this is difficult in particular concerning products whose life-cycle may span several decades…).

normal accident theory

Some argue that accidents in very complex and tightly coupled socio-technical systems (where failures of subsystems can interact in unexpected ways, where it is difficult for a single person to comprehend the entire system and understand all the factors that contribute to its safety) are inevitable, or "normal" (*Normal accident theory*, after [Perrow 1984]).

act of god

The legal system is generally poorly equipped when faced with "normal" accidents, since it is focused on determining the responsibility of an individual or a legal body. However, some legal systems include a distinction between an *injustice*, where someone must pay, and *misfortune*, where no-one is responsible (an "act of god" or *force majeur* situation)[11].

Social controls on the hazards caused by new technology

Definition

A number of **social controls** aim to ensure that technological change does not lead to unreasonable risks for health, safety and the environment [Baram 2007]:

- **consumer choice in the marketplace**: if consumers are well informed on the risks related to a product or service, they can make a choice which is compatible with their preferences;
- **self-regulation** by companies producing products or operating facilities, with voluntary programmes such as *Responsible Care* in the chemical industry [ICCA 2006];
- **government regulation** of authorized levels of pollution, of technological risk generated by a hazardous facility, of risk to consumers of a product;
- **tort liability** law (in certain countries, most famously the USA), which guarantees the right of a person harmed by another party's negligent behaviour or unsafe product to receive compensation for damages.

Several arguments can be put forward[12] against the applicability of tort liability in highly autonomous systems:

- since highly automated systems are destined to exceed governance by man, they defy a legal framework such as tort law which is ultimately based on individual responsibility;

[11] The example cited is the constant jurisprudence of the German Federal Court of Justice (Bundesgerichtshof) in civil matters when applying article 823 of the German civil code, concerning extracontractual responsibility.
[12] Presented during the workshop by Prof. Stefan Müller, School of Economics and Management, Technische Universiteit Berlin.

- ▷ the inherent complexity of highly-automated systems and the fast pace of technological change poses an insurmountable obstacle to the timely translation of concepts of responsibility in legal terms;
- ▷ in systems where the causation of damage and allocation of fault are likely to become increasingly difficult to decide, more cost-effective ways of managing risks and handling compensation for damages can be found[13].

Workshop participants largely agreed to the following statement:

In large, complex, tightly-coupled socio-technical systems where safety-critical functions are assured by automation supervised by human operators, certain accidents will have *systemic causes*, where it is difficult—and to a large extent meaningless—to allocate responsibility to an individual or a group of people. The use of **insurance schemes** for compensation of victims should be encouraged over the use of tort liability in these classes of systems.

insurance

Emphasis should be placed on identifying the **lessons that can be learned** from the systemic accident, which must be a trigger which leads system stakeholders to reexamine all facets of the complex, interacting system and the ways in which it can fail. Seeking to allocate individual or organizational responsibility is likely to hinder this search for understanding.

Expedited handling of financial compensation to victims of the Concorde crash

Air France Flight 4590 was a Concorde flight operated by Air France from Paris to New York City. In July 2000, having sustained damage from a foreign object dropped onto the runway by another passenger jet, it crashed shortly after takeoff, leading to the death of all 109 people on board and four people on the ground.

The accident investigation report was published at the end of 2004. A criminal investigation was launched in March 2005, and a ruling delivered in December 2010, more than 10 years after the accident. However, the airline and its insurers had arranged for full financial compensation to victims and their relatives to be paid within a year of the accident, without the need to go to court. This expedited compensation process has become an established practice among major airlines.

Training of system operators can fulfill multiple objectives:

- ▷ improve their skills and task performance;
- ▷ help compensate for system deficiencies which would be too expensive to resolve directly (fitting operators to the "Procrustean bed"[14]);
- ▷ attempt to **transfer liability for errors** to operators.

Training is necessary to allow operators to understand how the system that they will supervise works, to become familiar with its limitations and the possible failure modes of the technological components, and the recommended recovery strategies. Naturally, this training should not be used as a justification for limiting the effort put into the design of the system's technical components and the interaction between operators and technology.

Other issues discussed during the workshop, in brief:

- ▷ Some systems are **supra-national** (air traffic management, cyberspace, …), which may lead to conflicts between the different bodies of applicable legislation and questions as to which jurisdiction applies in case of a dispute.
- ▷ Should product liability issues depend on whether the product is marketed for use by *consumers* or by *professionals* (who can be expected to benefit from more training and specific qualifications)? There are slight differences in tort law, since the concept of "for intended use" is in some American states extended to "for use which can reasonably be anticipated".

"intended use"

[13] For instance, in many countries, compensation of victims of traffic accidents is handled via a compulsory third-party insurance scheme. This does not exclude criminal trials where drivers are judged for voluntarily reckless conduct.

[14] In Greek mythology, Procrustes (*"the stretcher"*) was a bandit who lived in Attica. He had an iron bed on which travelers he intercepted were made to spend the night. His humor was to stretch the ones who were too short until they died, or, if they were too tall, to cut off enough of their legs to make them fit. None could resist him, and the surrounding countryside became a desert.

proportional liability

- How is **shared responsibility** handled in tort liability law? Certain jurisdictions allow for "comparative responsibility" or "proportional liability", in which the fault is divided among the responsible parties, and compensation divided accordingly. This doctrine was used in the litigation concerning the Deepwater Horizon oil spill in April 2010, where responsibility was shared between the operator of the oil exploration project (BP), the owner of the platform (Transocean) and the subcontractor responsible for cementing the well (Halliburton).

Food for thought in your area of work

- Do system designers and managers accept that they may one day be held accountable for their contribution to an accident? Have they changed their way of working as a consequence?
- What is the balance between focus on understanding the underlying causes of accidents, and identifying people to blame?
- What is the real balance between the three possible objectives for training described above (improve operators' performance; compensate for system deficiencies; transfer liability for errors to operators) in the systems that you are involved in?

1.5 Governance of automation innovation

Conflict between fostering innovation and legal liability. Tort liability law is often criticized for its negative impact on innovation: firms are reluctant to introduce new and unproven technology—such as more sophisticated automation—for fear of future litigation. Indeed, some product liability laws employ the principle of *absolute liability* ("liability without fault" or "liability without negligence"), meaning that the manufacturer is liable for damages if injury is caused by a defect in a product, regardless of intent to produce harm. This concern is mitigated by two arguments that protect the firm which introduces new technology:

ultimate responsibility of the operator

- In the case of automation for supervisory control, the notion of "ultimate responsibility" of the human operator tends to protect the designer and manufacturer from liability while the technology is "young". Over time however, this argument becomes invalid; there is a tipping point where a court will judge that the "reasonable expectation" of the user is that the automated system works correctly.

development risks

- The **development risks defence** (present in most jurisdictions[15]) shields the producer from liability if the producer can demonstrate that the objective state of scientific and technical knowledge at the time the product was supplied was insufficient to enable the risk or fault to be discovered. The burden of proof lies with the producer/manufacturer. Note that this defence does not apply to risks which are known but which the current state of technology does not allow adequately to be managed.

Slow introduction of Advanced driver assistance systems

These systems are designed to improve road safety and reduce environmental impacts of automobile transport by providing help to the driver. Example technologies are adaptive cruise control (automatically slowing down the vehicle when approaching the car ahead), blind spot monitoring, lane departure warning systems, collision avoidance systems (which automatically apply the vehicle's brakes when a likely collision with an obstacle ahead is detected), and vehicular communication systems (networks in which vehicles and roadside units are the communicating nodes, providing each other with information, such as safety warnings and traffic information). Some of these systems actively take control away from the driver, by applying the brakes, tightening the seat-belt, closing windows prior to a collision, etc.

Some of these technologies have been available for at least a decade. Automobile manufacturers have been somewhat slow to bring them to the market, in part because of fears related to legal liability. For the same reasons, they are sometimes marketed as advisory support rather than as a safety function.

[15]Exceptions include certain EU states and specific areas such as pharmaceutical development.

Role of certification authorities. The introduction of new automation technologies in safety-critical systems is also subject to approval by certification authorities or regulatory bodies. The increasing complexity of systems managed by automation leads designers to implement complex non-deterministic control mechanisms, based on technologies such as rule-based automatons or fuzzy logic. However, certification authorities tend—for understandable reasons—to be conservative, demanding guarantees of predictable and deterministic operation of the automation.

conservatism impeding innovation in safety

Identifying success factors for automation projects. Many large and ambitious projects aiming to improve performance by introducing new automation have failed (due to inability to attain the level of performance demanded within the allocated budget and time frame, or due to lack of acceptance by politically powerful human operators[16]). Most learning concentrates on analyzing failures, but it would be interesting to understand what went *right* with certain successful automation projects, such as TCAS[17] and the short-term conflict-alert system in air traffic management.

what went right?

1.6 Automation philosophy

Different manufacturers can have different doctrines or philosophies with respect to automation. For example, Airbus is known to be more favorable to automation in the cockpit than is Boeing: its planes have a greater tendency to restrict the actions of the pilot when the automation decides that the consequences would be dangerous (flying into stall, for example). This difference may be attributed in part to different client populations: pilots in the USA are typically ex-US Air Force pilots, with excellent training in demanding situations, whereas European pilots more frequently learn at "school". The more experienced pilots tend to be more resistant to the introduction of more automation[18].

Transparency and "understandability" of automation: When designing a consumer device including high levels of automation (such as a smartphone), manufacturers attempt to "hide" internal system details from the user, aiming for "user friendliness". In contrast, the design of automation for safety-critical systems should aim to allow the professional operator to understand what is happening at lower levels within the system, in order to allow him to understand the reasons underlying decisions made by the automation and avoid "automation surprise[19].

However, care should be taken to avoid the opposite situation of **over-alarming**. Automation in control rooms uses alarms to inform the operators of abnormal conditions (excessively high temperature or pressure, failure of a system component, *etc.*), which are typically signaled to the operator using flashing lights and audible alerts. In crisis situations, a large number of alarms can be triggered simultaneously, which can overwhelm the operator. **Alarm management systems** should be implemented to filter alarms dynamically based on the current system state so that only the currently significant alarms are annunciated and to direct the operator's attention to the most important problem that he or she needs to act upon. More generally, operator interfaces should be designed to be [Woods and Branlat 2010]:

- ▷ **transition oriented**, to capture and display events and sequences;
- ▷ **future oriented**, to reveal what should or can happen next;
- ▷ **pattern based**, to allow quick recognition of unexpected or abnormal conditions;
- ▷ **abstract**, to capture the state and trends in higher order system properties.

Automation and safety/security compromises. Is the driver of an automobile equipped with driver assistance automation authorized to disable the safety function? It does seem difficult

[16] Air traffic management is an example of an area where operators wield considerable political power, and have refused to use certain new automated systems.

[17] The introduction of TCAS did pose problems (among them the Überlingen accident), but it was successful in introducing safety automation in a sector with significant organizational complexity.

[18] However, this doesn't explain their reticence to fly by wire, which existed since the F-16 fighter plane, more than 15 years before its introduction in civil aircraft.

[19] An *automation surprise* [Palmer 1995] occurs when the automation behaves in a manner that is different from what the operator is expecting. This can, for instance, be due to mode error or to a task knowledge gap [Johnson 1992] between operator and system.

to argue against the driver being able to switch off certain functions (particularly when they are unproven in use) if they are clearly failing. However, this introduces new risks related to external intervention (whether malicious or accidental).

Food for thought in your area of work

- ▷ Does the automation interface always allow operators to "dig down" into lower system layers, to understand why the technology is behaving as it is?
- ▷ Does the automation interface help operators by suggesting what should or could happen next?
- ▷ When dealing with very complex technology, is the automation interface expressed in terms of abstract notions, to allow operators quickly to understand the system's state?

Bibliography

Baram, M. (2007). Liability and its influence on designing for product and process safety. *Safety Science*, 45(1-2):11–30. DOI: 10.1016/j.ssci.2006.08.022. 8

Boy, G. A. and Grote, G. (2011). Chapter *The authority issue in organizational automation* in *The handbook of Human-Machine Interaction: A Human-centered design approach* (Boy, G. A., Ed.). Ashgate. 8

Dekker, S. and Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work*, 6(2):79–86. DOI: doi:10.1007/s10111-003-0136-9. 6

Dutch Safety Board (2010). Crashed during approach, Boeing 737-800, near Amsterdam Schiphol airport. Technical report, Dutch Safety Board. Available at the URL: http://www.onderzoeksraad.nl/docs/rapporten/Rapport_TA_ENG_web.pdf. 5

ICCA (2006). Responsible care global charter. Technical report, International Council of Chemical Associations. Available at the URL: http://www.icca-chem.org/ICCADocs/09_RCGC_EN_Feb2006.pdf. 8

Johnson, P. (1992). *Human computer interaction: psychology, task analysis, and software engineering*. McGraw-Hill, London. ISBN 9780077072353, 217 pages. 11

Luhmann, N. (1979). *Trust: a mechanism for the reduction of social complexity*. John Wiley & Sons, New York. 103 pages. 4

Palmer, E. (1995). "Oops, it didn't arm." A case study of two automation surprises. In Jensen, R. S. and Rakovan, L. A., Ed., *Proceedings of the eighth International Symposium on Aviation Psychology*, 227–232 pages. Available at the URL: http://human-factors.arc.nasa.gov/IHpersonnel/ev/OSU95_Oops/PalmerOops.html. 11

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 30(3):286–297. 4

Perrow, C. (1984). *Normal accidents: Living with High-Risk Technologies*. Basic Books, New York. ISBN 978-0465051427. 8

Sheridan, T. B. (2006). Chapter *Supervisory Control* in *Handbook of Human Factors and Ergonomics*, pages 1025–1052. John Wiley & Sons. DOI: 10.1002/0470048204.ch38. 4

Sheridan, T. B. and Verplank, W. L. (1978). Human and computer control of undersea teleoperators. Technical report, MIT Man-Machine Systems Laboratory, Cambridge, MA, USA. 3

Weyer, J. (2006). Modes of governance of hybrid systems. The mid-air collision at Ueberlingen and the impact of smart technology. *Science, Technology & Innovation Studies*, 2(2). Available at the URL: http://www.sti-studies.de/ojs/index.php/sti/article/view/95/76. 7

Woods, D. and Branlat, M. (2010). Hollnagel's test: being 'in control' of highly interdependent multi-layered networked systems. *Cognition, Technology & Work*. DOI: 10.1007/s10111-010-0144-5. 11

Young, M. S. and Stanton, N. A. (2002a). Attention and automation: new perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, 3(2):178–194. DOI: 10.1080/14639220210123789. 5

Young, M. S. and Stanton, N. A. (2002b). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44(3):365–375. DOI: 10.1518/0018720024497709. 4

Reproducing this document

Fondation pour une Culture de Sécurité Industrielle
a public interest research foundation
http://www.icsi-eu.org/

6 allée Émile Monso – BP 34038
31029 Toulouse cedex 4
France

Telephone: +33 534 32 32 00
Fax: +33 534 32 32 01
Email: contact@icsi-eu.org

ISSN 2100-3874

6 allée Emile Monso
Zac du Palays - BP 34038
31029 Toulouse Cedex 4
www.icsi-eu.org