Paper given at the 5th *plenary CaberNet workshop*, Madeira, 5-7 November 2003.

BUILDING DEPENDABLE SYSTEMS WITH FALLIBLE HUMANS

Denis Besnard
CSR, School of computing science, University of Newcastle
Newcastle upon Tyne NE1 7RU, United Kingdom
denis.besnard@ncl.ac.uk

Abstract: This paper gives an overview of some interdisciplinary issues in the design of computer-based systems. Three examples are borrowed from the DIRC project[1] in order to highlight the importance of taking into account the human factors in modern computing environments. The focus is on the dialogue between humans and automated machines and the necessity of an interdisciplinary collaboration in order to increase the level of reliability of this dialogue.

1. INTRODUCTION

Although the classical model of dependability is focussed on the technical aspects of computer-based systems (Randell, 2000), the combination of technical and human agents is paramount in these systems. At the design level, the integration requires the collaboration of several disciplines, this being the principle that drives the DIRC project.

The necessity for multidisciplinarity becomes obvious when field data are analysed. The latter often show that technical failures alone only account for a small proportion of incidents and accidents. Most often, the failure lies in the dialogue between human agents and the automation. For this reason, it seems worthwhile looking into some forms of socio-technical failures and learn about the potential dimensions of interest to designers and managers. A quick investigation will reveal some diversity in these dimensions and some candidate avenues for progress.

2. DEPENDABLE SYSTEMS WITH HUMAN COMPONENTS: AN OXYMORON?

Until the mid 1980's, the implicit design conception has been that operators have to adapt to the tool they interact with. Later, the ubiquity of computers and the increasing criticality of the processes they controlled made it clear that human-machine interaction was going far beyond just computer science. IT designers then needed to communicate with other disciplines.

However, making automated agents cohabit with human operators is a real challenge. One could decide to cavalierly ignore issues such as trust or acceptance of technology, and still face very hot topics to address. Among these, the question of making humans and machines cooperate reliably remains without a definitive answer. In order to better understand the interdisciplinary challenge raised by modern human-machine interaction, three examples -on which DIRC has produced publications- are discussed. These highlight three different areas of sub-optimality, namely: mental models in critical systems (Besnard, Greathead & Baxter, 2003), usability trade-offs in security (Besnard & Arief, 2003) and workarounds in assembly lines (Voß *et al.*, 2000). This variety is hoped to highlight the width of the spectrum to be considered when reasoning about, and designing socio-technical systems. Comments on these examples will follow (section 3) in which potential improvements and ways forward will be considered.

2.1. Fragile mental models in critical systems

On the 8th of January 1989, a British Midland Airways Boeing 737-400 aircraft crashed into the embankment of the M1 motorway near Kegworth (Leicestershire, UK), resulting in the loss of 47 lives. The crash resulted from the flight crew's management of a mechanical incident in the left engine. A fan blade detached from the engine, resulting in severe vibrations and fumes. The flight crew mistakenly identified the faulty engine as the right engine. The latter was throttled back and eventually shut down. This action coincided with a drop in vibration and the cessation of smoke and fumes from the left (faulty) engine. On the basis on these symptoms, the flight crew deduced that the correct decision had been taken, and sought to make an emergency landing at East Midlands airport. After the crew initially

[1] Visit DIRC at www.dirc.org.uk

reduced power to the left engine at the beginning of descent, the thrust to this engine was increased to maintain altitude during the final stages of descent. This resulted in greatly increased vibration, the loss of power in that engine and a fire warning. The crew attempted at this point to restart the right engine but this was not achieved in the time before impact, which occurred 0.5 nautical miles from the runway (see AAIB, 1989, for the accident report).

Operators attempting to save cognitive resources biases mental models in such a way that partial confirmation is easily accepted. Instead of looking for contradictory evidence, people tend to wait for consistent data. This phenomenon, called confirmation bias (Klayman & Ha, 1989), has already been studied in human-machine interaction (e.g. Yoon & Hammer, 1988). The corollary of confirmation bias is that people overlook contradictory data. In the Kegworth accident, an erroneous decision coincided with a reduction in the level of the symptoms which lasted for some twenty minutes. When it is compatible with the operator's expectations, this type of co-occurrence probably works against rejection of the existing mental model (Moray, 1987). It also makes it harder to integrate any contradictory evidence that may subsequently become available.

Operators are more likely to reject any information that is not consistent with their expectations, rather than update their mental model. The latter has a cost that operators cannot always afford in time-critical situations.

2.2. Usability trade-offs in IT security

Although new approaches towards authentication have been proposed (Brostoff & Sasse, 2000, 2003; Jermyn, 1999), complex passwords still remain a widespread security mechanism. This state of affairs probably originates from a desire to control accesses more and more tightly. But even complex passwords are not totally secure (Nielsen, 2000). When one actually looks at what happens in the workplace, human cognitive limitations become obvious: users cannot remember their passwords and need external memories (e.g. sticky notes on monitors). The use of passwords raises several usability problems (Adams, Sasse & Lunt, 1997). Ultimately, security faces a nice paradox where by increasing the complexity and number of passwords, the level of protection can actually decrease (Weirich & Sasse, 2002).

Users who write down passwords act in a way which is the result of an equation where risks, costs and benefits are core factors. Given their perception of risk, users trade-off the cost of memorising passwords against the benefits of seeking ease-of-work. This intuitive usability trade-off also applies to file sharing, patching software or updating anti-virus programs. Risk-unaware users seek immediate benefits at the cost of expensive failures. Their behaviour can be seen as driven by a rule of least effort (Rasmussen, 1986) where any security measure hanging in the way of accomplishing the main task is unlikely to be followed.

2.3. Workarounds in assembly lines

A company assembling diesel engines relies on a computer-based tool in order to track orders, deadlines, special customisations for particular customers, etc. The software drives the entire supervision of the process from the management of the stocks, all the way down to delivery dates. The software is designed in such a way that all the parts needed for an engine have to be in-stock before the assembly can begin. This appears to be an unworkable constraint for the operators who can still work on areas of an engine for which parts are available. However, because the software system does not allow this sort of ad-hoc adaptations to contingencies, operators create items in the stock for the parts that are missing and begin the assembly. Later, when the missing parts get delivered, they are mounted on the engine and the stock is set back to zero.

Generally speaking, procedures do not rule human behaviour (Fujita, 2000). Humans do not obey rules if the latter are perceived to uselessly obstruct the acomplishment of a task. Here, the cost of waiting is avoided by a deviation from the procedure. When this happens, it is usually a symptom that the tools and procedures in use do not match the operators' needs or intentions. This mismatch, which often involves managerial decisions (Reason, 1995), has already caused very serious accidents in other areas of the industry (see the Tokaimura incident report by Furuta *et al.*, 2000).

3. THE FUTURE OF DEPENDABLE SYSTEMS

Humans as system components will probably evolve less than their technical counterparts. Humans' cognitive architecture will not change in some years' time, their processing capabilities will not double each year from now on and they will continue to face the same cognitive limitations as their grandfathers. So the way out is to support humans as they are: fallible agents. From a design policy standpoint, the solution no longer lies in radical technological inventions but in further improvements of the dialogue between human operators and the automation. The three examples presented in section 2 are now going to be reviewed and possible ways forward will be suggested.

3.1. Pro-active assistance to operators in critical systems

One way forward to intelligent decision support systems (Hollnagel, 1987) is to design support tools that incorporate models of the system and its users. This would allow machines to predict the future states they are going to enter, flag dangerous future states (see Hazard Monitor in Bass *et al.*, 1997), detect potential cognitive conflicts (see Rushby *et al.*, 1999), anticipate operator's decisions, provide more appropriate context-sensitive alarms and support for critical decisions. Expected benefits include the provision of some assistance in emergency situations before matters become too critical. It implies that systems at large have to be designed in such a way that even unexpected events can be identified as such and appropriately handled by support tools.

3.2. Improving the human side of security

Security must be user-centred (Zurko & Simon, 1996). Generally speaking, the design of security products and policies should rely more on the rules of human-computer interaction, as suggested in Patrick *et al.*, 2003; Sasse, 2003). Passwords must be, at least, easy to remember and reduced in number as much as possible. As far as end-users are concerned, the ideal number of passwords is zero, so any measure getting closer to an *effortless* security is a step forward to better security in general. People should not have to remember about IT security or even think about it. Security should be transparent for whom it is not a primary objective.

3.3. Workarounds as symptoms of system's flaws

Workarounds are unsupported configurations. They are violations revealing a lack of flexibility and a need for configurability. One way of avoiding violations and dangerous workarounds is to design around human practices. If this is not a design decision, it will be enforced in an ad-hoc manner by users, out of any control. Integration of practices into design is not a novel idea but it does not seem to be as applied as is actually needed. Design, from the early stages, must reflect the way the work is done. This also applies to subsequent designs and evolution of systems. It is an important issue. If we want to assume that a given system is used by the rules, these rules have to be workable.

4. CONCLUSION

This paper has presented three cases where DIRC investigated the dialogue between humans and some form of automation. The author is of the opinion that the technical aspects of the failures described herein only account, if at all, for a portion of the picture. Within the scope of this paper, human factors ranging from cognitive ergonomics to sociology are needed in order to capture the complexity of designing collaborative, secure and workable automation. Without taking into account such basic recommendations as in section 3, the final dependability of socio-technical systems will not go beyond its current weakest link: the reliability of human-machine interaction.

5. REFERENCES

Randell, B (2000). Facing up to faults. *The Computer Journal*, 43, 95-106.

Reason, J. (1995). A systems approach to organized error. *Ergonomics*, 38, 1708-1721.

AAIB (1989). Report on the accident to Boeing 737-400-G-OBME near Kegworth, Leicestershire on 8 January 1989. Report available online at http://www.dft.gov.uk/stellent/groups/dft_control/documents/contentservertemplate/dft_index.hcst?n=5236&l=4

Besnard, D. & Greathead, D. (2003). When mental models go wrong. Co-occurrences in dynamic, critical systems. To appear in *International Journal of Human-Computer Studies*.

Besnard, D. & Arief, B. (2003). Computer security impaired by legal users. To appear in *Journal of Computers & Security*.

Voß, A., Procter, R., Slack, R., Hartswood, M., Williams, R. & Rouncefield, M. (2000). Production management as ordinary action: an investigation of situated, resourceful action in production planning and control. Proceedings of the 20th UK *Planning and Scheduling (SIG) Workshop*, Edinburgh, UK (pp. 230-243).

Klayman, J. & Ha, Y.-W. (1989). Hypothesis testing in rule discovery: Strategy, structure and content. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 5, 596-604.

Yoon, W. C. & Hammer, J. M. (1988). Deep-reasoning fault diagnosis: an aid and a model. *IEEE Transactions on Systems, Man and Cybernetics*, 18, 659-676.

Moray, N. (1987). Intelligent aids, mental models, and the theory of machines. *International Journal of Man-Machine Studies*, 27, 619-629.

Brostoff, S. and M. A. Sasse. (2000). Are passfaces more usable than passwords? *People and Computers-XIV, Usability or Else!* Proceedings of *HCI2000*, Sunderland, UK, (pp. 405-424).

Brostoff, S. & Sasse M. A. (2003). Ten strikes and you're out: Increasing the number of login attempts can improve password usability. Proceedings of *CHI 2003 Workshop on HCI and Security Systems*, Fort Lauderdale, Florida.

Jermyn, I. (1999). The design and analysis of graphical passwords. *Proceedings of the 8th USENIX Security Symposium*, Washington DC.

Nielsen, J. (2000). Security and human factors. Article available online at http://www.useit.com/alertbox/20001126.html.

Adams, A., Sasse, M. A. & Lunt, P. (1997). Making passwords secure and usable. Proceedings of *HCI'97 People & Computers XII* (pp. 1-19).

Weirich, D. & Sasse M. A. (2002). Pretty Good Persuasion: A First Step Towards Effective Password Security in the Real World. Proceedings of *New Security Paradigms Workshop*, Cloudcroft, NM (pp. 137-144).

Rasmussen, J. (1986). *Information processing and human-machine interaction.* North Holland: Elsevier Science.

Fujita, Y. (2000). Actualities need to be captured. *Cognition, Technology & Work*, 2, 212-214.

Furuta, K., Sasou, K., Kubota, R., Ujita, H., Shuto, Y. & Yagi, E. (2000). Analysis report. *Cognition, Technology & Work*, 2, 182-203.

Hollnagel, E. (1987). Information and reasoning in intelligent decision support systems. *International Journal of Man-Machine Studies*, 27, 665-678.

Rushby, J., Crow, J. & Palmer, E. (1999). An automated method to detect potential mode confusions. Proceedings of the *18th AIAA/IEEE Digital Avionics Systems Conference*, St Louis, MO, USA.

Zurko, M. E. & Simon R. T. (1996). User-Centred Security. Proceedings of *Workshop on New Security Paradigms*, Lake Arrowhead, CA (pp. 27-33).

Patrick, A. S., Long, A. C. & Flinn, S. (2003). HCI and security systems. Proceedings of *CHI 2003 Workshop on HCI and Security Systems*, Fort Lauderdale, Florida.

Sasse, A., (2003). Computer security: Anatomy of a usability disaster, and a plan for recovery. Proceedings of *CHI 2003 Workshop on HCI and Security Systems*, Fort Lauderdale, Florida.

6. ACKNOWLEDGEMENTS

This paper was written at the University of Newcastle upon Tyne within the DIRC project (http://www.dirc.org.uk) on dependability of computer-based systems. The author wishes to thank Cliff Jones for useful comments and the sponsor EPSRC for funding this work.