When mental models go wrong. Co-occurrences in dynamic, critical systems.

Denis Besnard & David Greathead

School of Computing Science
University of Newcastle upon Tyne
Newcastle upon Tyne NE1 7RU
United Kingdom

Abstract. This short paper is intended to highlight an interesting psychological phenomenon affecting the accuracy of mental models. It occurs when two events happen as expected by an operator but for reasons that are actually not understood. In other words, a mental model of a problem is erroneously taken to be valid as a result of a mere co-occurrence. We discuss this phenomenon with the example of a real commercial air crash. We finally address some implications for systems' design and support tools.

Keywords. Mental models; human error; cognitive psychology; critical systems.

1. INTRODUCTION

During a presentation, one of the authors' colleagues using his laptop for displaying slides got interrupted when his machine's screen went blank. He hit the track pad of his laptop in case the latter had gone to sleep mode but the video signal did not come back. As he was using an adapter for the VGA cable, he suspected the connection had gone loose. After some seconds during which he tightened more firmly the adapter and the VGA cable together, the video came back on. The problem was solved and hopefully, the adapter would now behave normally. Some minutes later, the same scenario happened again and our colleague did the same actions as before. He hit the track pad but this triggered no signal. He then modified the position of the VGA adapter so that the weight of the cable would not pull on it. A couple of seconds later, the video was back on again. The third time the scenario happened, it became obvious, at least to everyone *except* the person using the laptop, that the manipulation on the VGA adapter and the video coming back on was pure coincidence. We discovered that the machine was going to sleep mode and needed about 6 to 8 seconds to exit it. It is likely that any action carried out at the precise moment when the video would come back on would be considered as a solution.

Through this simple -yet real- example, we wish to highlight an interesting cognitive feature. Humans tend to consider that their vision of the world is correct whenever events are happening in concordance with their expectations. When this is the case, problems are erroneously thought to be understood. Co-occurrences cause a lot of disruption in situation awareness to this respect. Two events can happen as expected with their cause not being captured. When this happens, events are erroneously treated as an evidence of valid understanding of a problem.

In the next sections, we will consider a psychological concept (mental models; see section 2) that will shed some light on the mechanisms leading to the aforementioned error. We will then assess the role of this error in critical high-tempo applications through the example of a commercial air crash (section 3). We will finally discuss our paper and assess its relevance for the design of critical systems (section 4).

2.   MENTAL MODELS

Because of limitations in memory and processing capabilities, humans cannot handle the totality of the information displayed in their environment. Instead, they build representations that are meant to support behaviour (Rabardel, 1995). These representations are called mental models (see Byrne, Handley & Johnson-Laird, 1992 for an introduction). When used for describing cognitive activities *in situ*, this concept refers to scarce, goal-driven images of the world that are built to understand the current state of a situation and also to predict the future states of the interaction with this situation.

What characterises best mental models is their incompleteness. Their content is only a partial representation of the environment and their scope is limited (Sanderson, 1990; Sanderson & Murtagh, 1990). They are essentially built from a) the knowledge needed for pursuing a given goal and b) some data extracted from the environment. The resulting image of the world is one where the essential features of a problem are overemphasized whereas the peripheral data can be overlooked (Ochanine, 1978). To this respect, mental models have been called homomorphic representations of the world (Moray, 1987). They are simplified, cognitively acceptable versions of a too complex reality.

Although it is a core activity in process control, human operators do not only have to build representations of their environment. They also have to plan actions, control movements, exchange information with collaborators, etc. This complex combination of tasks has to be executed within a limited amount of processing resources. For this reason, humans tend to save resources whenever it is possible. Forgetting is an instance of such a mechanism. It can also take the form of a heuristic, shortcut-based reasoning (Rasmussen, 1986). As far as mental models are concerned, saving resources causes them to be built on the basis of partial pieces of evidence. However, this has to be seen as the consequence of cognitive limitations where problems are solved according to an intuitive cost-benefit trade-off. Since Simon's (1957) concept of bounded rationality, it is accepted that cheap acceptable solutions are often preferred to costly perfect ones.

The consequences of flawed mental models can be disastrous when human beings are interacting with high-tempo critical systems. Human operators (e.g. commercial aircraft pilots) are sometimes faced with unexpected incidents for which they have to find a cause and treat it. This local troubleshooting activity, which is inserted in the more global objective of piloting the aircraft, involves the construction of an explanation in real-time. Because of factors such as limited cognitive resources, confirmation bias and time pressure, pilots are likely to build an erroneous explanation of such incidents. Flaws in mental models are detected when the interaction with the world reveals unexpected events. However, these inaccurate mental models do not always lead to accidents. Very often, they are recovered from. To this respect, error detection and compensation are significant features in human information processing.

The weakness of mental models lies in their poor requirements in terms of validity: If the environmental stream of data is consistent with the operator's expectations, that is enough for the mental model to be reinforced. The understanding of the mechanisms generating the data is not a necessary condition. This is the topic of the paper.

We have to make clear that the scope of the paper is not to know how operators could build exhaustive mental models as their incompleteness reflects a strong need for drastic information selection. Rather, the issue is to understand the conditions in which operators think they have good picture of a situation whereas the underlying causal mechanism has not been captured. We think one of these conditions is a co-occurrence of events. This topic will now be investigated in the context of a critical phase of a commercial flight leading to a crash.

3.   THE KEGWORTH ACCIDENT

On the 8th of January 1989, a British Midland Airways (BMA) Boeing 737-400 aircraft crashed into the embankment of the M1 motorway near Kegworth, resulting in the loss of 47 lives (AAIB, 1989). The crash resulted from the crew's management of a mechanical incident in the left (#1) engine. A fan blade detached from the engine, resulting in vibration (strong enough to be felt by the crew) and the production of smoke and fumes drawn into the aircraft through the air conditioning system. The flight crew mistakenly identified the faulty engine as the right (#2) engine and reduced its power. The cockpit voice recorder showed that there was some hesitation regarding this decision. When the captain asked which engine was faulty, the first officer replied '*It's the le… it's the right one*', at which point the right engine 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 (actually at fault) engine. The flight crew erroneously deduced that the correct decision had been taken, and sought to make an emergency landing at East Midlands Airport. The left engine continued to show increased vibration for some minutes, although this seems to have passed unnoticed by the crew. Soon afterwards, the crew reduced power to the left engine to begin descent, whereupon the vibration in the engine dropped to a point a little above normal. Approximately ten minutes later, power to the left engine was increased to maintain altitude in the final stages of descent. This resulted in greatly increased vibration, the loss of power in the engine and an associated fire warning with that engine. The crew attempted at this point to restart the right engine but this was not achieved in the time before impact, which occurred at some 0.5 nautical mile from runway.

On top of crew's mistakes (see Ladkin, 1996), several factors contributed to the accident:
When later interviewed, both pilots indicated that neither of them remembered seeing any indications of high vibration on the Engine Instrument System (EIS; see Figure 1). The captain stated that he rarely scanned the vibration gauges as he had found them to be unreliable in other aircraft of his experience. It is also worth noting that the aircraft was using a new EIS which used digital displays rather than mechanical pointers. In a survey[1] carried out in June 1989, 64% of BMA pilots indicated that the new EIS was not effective in drawing their attention to rapid changes in engine parameters and 74% preferred the old EIS. The secondary EIS (see Figure 2), on which the vibration indicator was located did not include any audio or additional visual warning to indicate excessive readings[2].

Figure 1: A 737-400 cockpit. The EIS is located in the centre (see white circle). © Pedro Becken.

---

[1] This survey is summarised in the accident report (AAIB, 1989) in section 1.17.3.
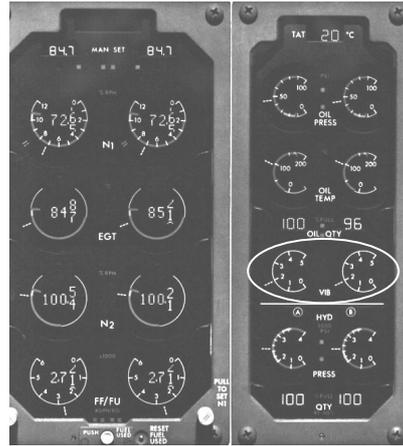[2] This is in accordance with standard design practice.

Figure 2: A 737-400 EIS. The secondary EIS is on the right-hand half of the picture. The vibration indicators are circled in white.

As another contributing factor, the crew workload increased out of control. Some time after the #2 (working) engine had been erroneously shut down, the captain tried without success to stay in phase with the evolution of the incident. He was heard on the CVR saying: '*Now what indications did we actually get (it) just rapid vibrations in the aeroplane – smoke…*'. At this point, the crew were interrupted with a radio communication from the air traffic control. Later, the Flight Service Manager entered the flight deck and reported that the passengers were very panicky. This further distracted the flight crew and the captain had to broadcast a message to the passengers. Both the captain and first officer were also required to make further radio communications and perform other duties in preparation for the landing. All of these actions impacted on the degree of control of the emergency.

Finally, it is worth noting that while both the captain and the first officer were experienced (13,000+ hours and 3,200+ hours flying time respectively), together they had only 76 hours of flying experience in the 737-400 series.

4. DISCUSSION

In this section, we are going to briefly address some general issues about mental models in dynamic systems (for a definition of these systems, see Cellier, Eyrolle & Mariné, 1997 or Brehmer, 1996). Stemming from the accident data exposed in section 3, we will then suggest some reflections on dependable systems' design. We will finally conclude the discussion by a quick look at what limits have to be made clear when studying human error.

Although we are studying a specific flaw in mental models on the basis of an aircraft accident, we wish to highlight that our approach is different from mode confusion (Rushby *et al.*, 1999; Crow *et al.,* 2000; Rushby, 2001; Leveson *et al.*, 1997). We are not focussed on automation surprises. Instead, we are interested in the nature of some of the cues involved in flawed mental models in man-machine interaction.

4.1. Mental models in a dynamic world

Saving cognitive resources biases mental models in such a way that partial confirmation is easily accepted. Instead of looking for contradictory pieces of evidence, mental models tend to "wait" for consistent data. This phenomenon, called *confirmation bias* (see Klayman & Ha, 1989 for a definition)*,* has already been studied in man-machine interaction (see for instance Yoon & Hammer, 1988). Its main side-effect consists in overlooking contradictory data. This is one explanation for the reinforcement of flawed mental models. In the case of the accident reported in this paper, an erroneous decision was made that coincided with a drop in symptoms. This probably confirmed the crew's belief that they had shut down the faulty engine. Moreover, the

drop in symptoms persisted for some twenty minutes, thereby reinforcing the crew's conviction to have solved the problem. This type of co-occurrence probably reinforces mental models, causing contradictory evidence hard to integrate even if is available.

Human operators can erroneously maintain as valid, representations that have already departed from a reasonable picture of the reality. In high-tempo situations, one reason is that operators try to avoid the cost of revising their mental model as long as it allows to stay more or less[3] in control. We can investigate this mechanism a little further. Because mental models are constantly matched against the feedback from the process they control, they are fed with a constant stream of data. However, there exist situations were the situational feedback is discrepant from the operator's expectations. When this discrepancy provokes such a loss of control that required tasks cannot be run anymore, some costly revision of the mental model as well as diagnostic actions are needed (Rasmussen, 1993). This is a non trivial task in dynamic situations such as piloting an aircraft: Whereas some situation awareness is already lost, the crew is required to run, coordinate and share two processes at the same time. One is a rule-based control of the flight parameters (the plane must continue to fly). The other process is information gathering and integration. The potential work overload caused by this dual activity may explain why out-of-date mental models are kept "alive" even after the detection of some mismatches. Providing they can keep the system within safe boundaries, operators in critical situations can feel more comfortable with loosing some situation awareness rather than spending time gathering data at the cost of a total loss of control (Amalberti, 1996).

Critical situations can be caused by the combination of an emergency followed by some loss of control. When this happens, there is little room for recovery. The Kegworth accident probably falls into this category. By the time they attempted to read data on the engines, the crew got caught by other tasks. The emergency of the situation and the emerging workload delayed the revision of the mental model which eventually was not resumed.

4.2.   Implications for the design of dependable systems'

The Kegworth crash highlights that automation is a real dialogue between humans and machines. When this dialogue fails because information flow does not help situation awareness, incidents are likely to be processed in a sub-optimal manner. The following discussion will not focus explicitly on co-occurrences as we see them as a very local mechanism as compared to the complexity of cognitive activities involved in the control of dynamic processes. Instead, we think a wider discussion is needed in order to assess more precisely the stakes of a more reliable interaction between man and automated systems. We speculate human-machine interaction could be improved in three complementary ways.

- Firstly, humans must be made more aware of their own functioning, by such means as training and education. Some psychological mechanisms can then become more obvious to the operators themselves and positively influence the perception that they have of their own performance. To the best of our knowledge, this human-factors vision has now been integrated in air pilots' training for about 15 years. However, benefits cannot be immediate. In a near future, it may be the case that more and more pilots, who will have been educated in human factors from the early stages, will contribute to a even higher degree to critical systems' dependability.

- Secondly, the automation must be aware of the human operators by having embedded, during the design process, some knowledge of the human reasoning as well as some screening functions (as already suggested by Rasmussen, 1991). This would allow machines to anticipate human's decisions, provide context-sensitive alarms and support

---

[3] The wording may seem a bit loose but actually, humans typically accept not to understand everything of a problem as long as they reach a precision of control that matches their intentions.

for critical decisions. Such an approach was already investigated by Boy as early as 1987. Expected benefits include the provision of some assistance for exceptional emergency situations before humans face critical problems. Operators need more help on the situations for which they have not been trained, than on nominal settings. It implies that systems at large have to be designed in such a way that unexpected events are handled in some way by support tools. Wageman (1998) argues that interfaces can typically flood operators with extra data at a time of the process (e.g. emergencies) where few resources are still available. From our point of view, we think it is precisely because human properties and intentions are not captured by automated systems that over-information occurs. This issue has been addressed by Filgueras (1999) and more extensively developed by Hollnagel (1987) who proposed the concept of *intelligent decision support systems*.

- Lastly, a way forward may be to design support tools having models of the system they are a part of. This would permit the automation to predict the future states it is going to enter given the inputs coming from the environment and the operator. Without this kind of assistance, humans will have to continue looking for data during critical phases of process control instead of having pro-active support.

The constant increase in commercial aviation traffic has not been followed by an increase in accident rate. A contribution to this rather positive state of facts is the steady technical improvement of modern aircrafts. Nevertheless, as reported by Amalberti (1996), a flat accident rate persists since the 1970s. This is why we think more efforts have to be invested in the reliability of the dialogue between operators and automation. We think, following Amalberti, that the limit to modern aviation safety now lies in the extent to which we can improve cooperation in the dialogue between automated systems and human agents. This assertion undoubtedly extends beyond aircraft piloting and hits any critical system where humans have to take decisions.

4.3. Limits

We wish to emphasise that mental models can fail for several other causes than co-occurrence. Instances of such causes include complexity, lack of knowledge, workload. We want to give co-occurrences the attention they deserve. They can lead to catastrophes but only account for a small portion of the causes of failure of mental models.

Although we have focussed on the weaknesses of mental models, we also have to emphasize that human errors are not cognitive dysfunctions. Often, errors must be seen are marginal events caused by the same mechanisms that generate correct actions most of the time (Johnson *et al.*, 1992). As a consequence, errors have to be considered in this paper as the side-effects of a cost/benefit driven reasoning process aimed at getting an optimal performance for the lowest mental cost (Amalberti, 1991, 1996).

5. CONCLUSION

In this paper, we have emphasised the negative impact of co-occurrences on the accuracy of mental models. We have attempted to analyse the process by which some co-occurrent events can be erroneously treated as a confirmation of understanding. This mechanism partly explains the crash of a commercial airplane in United Kingdom in 1989 and has to be taken into account when building critical applications in socio-technical systems.

6. ACKNOWLEDGEMENTS

7. REFERENCES

Air Accidents Investigation Branch (1989). Report on the accident to Boeing 737-400 - G-OBME near Kegworth, Leicestershire on 8 January 1989. Available online at http://www.aaib.dft.gov.uk/formal/gobme/gobmerep.htm

Amalberti, R. (1991). Modèles de raisonnement en ergonomie cognitive. in *Science et Défense 91. Sécurité des systèmes. Neurosciences et ergonomie cognitive.*

Amalberti, R. (1996). *La conduite de systèmes à risques.* Paris : P.U.F.

Boy, G. (1987). Operator assistant systems. *International Journal of Man-Machine Studies*, 27, 541-554.

Brehmer, B. (1996). Man as a stabilizer of systems. From static snapshots of judgement processes to dynamic decision making. *Thinking and Reasoning*, 2, 225-238.

Byrne, R.M, Handley, J.-H. & Johnson-Laird, P. N. (1992). Advances in the psychology of reasoning : meta-deduction. in M. T. Keane & K. J. Gilhooly (Eds) *Advances in the Psychology of thinking, vol. 1.* Harvester Wheastsheaf: New York.

Cellier, J.-M., Eyrolle, H. & Mariné, C. (1997). Expertise in dynamic systems. *Ergonomics*, 40, 28-50.

Crow, J., Javaux, D. & Rushby, J. (2000). Models of mechanised methods that integrate human factors into automation design. *International Conference on Human-Computer Interaction in Aeronautics: HCI-Aero 2000*, Toulouse, France.

Filgueras, L. V. L. (1999). Human performance reliability in the design-for-usability life cycle for safety human computer interfaces. in M. Felici, K. Kannoun & A. Pasquini (Eds). *SAFECOMP'99*, Springer-Verlag: Heidelberg (pp. 79-88).

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

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.

Ladkin, P. (1996). *Extracts from UK AAIB Report 4/90 on the 8 January 1989 accident of a British Midland B737-400 at Kegworth, Leicestershire, England.* Available online at http://www.rvs.uni-bielefeld.de/publications/Incidents/DOCS/ComAndRep/Kegworth/kegworth-ladkin.html

Leveson, N., Pinnel, L. D., Sandys, S. D., Koga, S. & Reese, J. D. (1997). Analysing software specifications for mode confusion potential. in C. W. Johnson (Ed) *Proceedings of a workshop on human error and system development*, Glasgow, Scotland (pp. 132-146).

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

Ochanine, D. (1978). Le rôle des images opératives dans la régulation des activités de travail. *Psychologie et Education*, 2, 63-72.

Rabardel, P. (1995). *Les hommes et les technologies.* Paris : Armand Colin.

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

Rasmussen, J. (1991). Technologie de l'information et analyse de l'activité cognitive. in R. Amalberti, M. de Montmollin & J. Theureau *Modèles en analyse du travail.* Liège : Mardaga (pp. 49-73).

Rushby, J. (2001). Modelling the human in human factors. Invited paper, *Safecomp 2001*, Budapest, Hungary (pp. 86-91).

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.

Sanderson, P. M. & Murtagh, J. M. (1990). Predicting fault diagnosis performance : why are some bugs hard to find ? *IEEE Transactions on Systems, Man and Cybernetics*, 20, 274-283.

Sanderson, P. M. (1990). Knowledge acquisition and fault diagnosis : experiments with PLAULT. *IEEE Transactions on Systems, Man and Cybernetics*, 20, 255-242.

Simon, H. A. (1957). *Models of man*. New York : Wiley.

Wageman, L. (1998). Analyse des représentations initiales liées aux interactions homme-automatique (IHA) en situation de contrôle simulée. *Le Travail Humain*, 61, 129-151.

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.