In A. Özok & G. Salvendy (Eds.), Advances in Applied Ergonomics (pp. 827-830). West Lafayette: USA Publishing.

WHY AND WHAT CAN WE LEARN FROM HUMAN ERRORS?

Matthias RAUTERBERG

Work and Organizational Psychology Unit (IfAP), Swiss Federal Institute of Technology (ETH)
Nelkenstrasse 11, CH-8092 Zürich, SWITZERLAND
Tel.: +41-1-6327082, Fax: +41-1-6321186, Email: rauterberg@ifap.bepr.ethz.ch

Keywords:
Error; learning; cognitive structure; expertise

Abstract:
In this paper the traditional paradigm for learning and training of operators in complex systems is discussed and criticised. There is a strong influence (the doctrine of 'mental logic') coming from research carried out in artificial intelligence (AI). The most well known arguments against the AI-approach are presented and discussed in relation to expertise, intuition and implicit knowledge. The importance of faults and errors are discussed in the context of cognitive structures to describe expertise, and how knowledge about unsuccessful behaviour influences the actual decisions.

INTRODUCTION

Why is this statement sometimes (or always) true? To answer this question we need a new understanding of human errors, inefficient behaviour, and expertise. In this paper we will discuss the importance of learning from unsuccessful behaviour. What percentage of unanticipated events (e.g., accidents) is caused by human error? This is a question that vexed researchers for years in the context of human interaction with complex systems.

The classical understanding of human errors is characterized by a *negative* valuation of erroneous behaviour, something that must be avoided. The Western Culture is constrained by this *taboo*: Not to talk about faults, errors and other dangerous behaviour! This taboo keeps us to present our self as successful as possible. We are--normally--not allowed to discuss in public how and what we could learn from our faults and errors.

Rasmussen (1986) defines human errors as follows: "if a system performs less satisfactorily than it normally does–because of a human act–the cause will very likely be identified as a human error". Accidents are categorised as caused by either unsafe acts of persons (e.g., operator error) or by unsafe conditions (cf. Reason, 1990). One consequence of using this dichotomy is often to blame the individual who was injured or who was in charge of the machine that was involved in the accident. In fact, it is probably meaningless even to ask what proportions of accidents were due to human error. The more important question is what can one learn from his or her errors, and how are these insights and the derived knowledge embedded in the individual cognitive structure.

How would it be if the majority of the knowledge of the long-term memory of humans consists only of *unsuccessful trials?* Nearly all modelling approaches seem to be a representatives of a theory driven approach for *error-free skilled behaviour* (Booth, 1991, pp. 80ff). Why do we believe that an empirical driven approach--looking to the concrete task solving behaviour of people--is better than a theory driven approach? The answer refers to the following assumption.

Most of the known modelling approaches is based on the implicit assumption that the mental model maps completely to the relevant part of the conceptual model, e.g. the user virtual machine. Unexpected effects and errors point to inconsistency between the mental model and the conceptual model. This one-to-one mapping between the mental model and the conceptual model of the interactive system implies a *positive* correlation between the complexity of the observable behaviour and the complexity of the assumed mental model. But this assumption seems to be – in this generality – wrong.

Based on the empirical result in (Rauterberg, 1993), that the complexity of the observable behaviour of novices is larger than the complexity of experts' behaviour, we must conclude that the behavioural complexity is *negatively* correlated with the complexity of the mental model. If the cognitive structure is too simple, then the concrete task solving process must be filled up with many heuristics or trial and error behaviour.

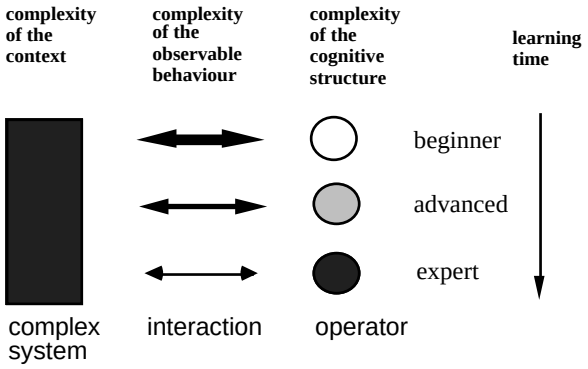

Figure 1. The relationship between the complexity of the human interaction and of the cognitive structure.

Learning how to solve a specific task with a given system means that the behavioural complexity decreases and the cognitive complexity increases (cf. Figure 1). Now, one of the central question is: What kind of knowledge is stored in the cognitive structure? Before we are able to give a preliminary answer to this question, we have to discuss the consequences of the traditional AI-paradigm.

The famous classification of Rasmussen (1986) in skill-based, rule-based, and knowledge-based behavior is one consequence of taking the AI-approach seriously. Human behaviour can be therefor classified (1) as error-free skilled behaviour, (2) as inefficient behaviour, and (3) as erroneous behaviour (cf. Wickens, 1992).

Berkson and Wettersten (1982) can show why the AI-approach is not able to cover human errors. The classical AI-approach has a high affinity with the theory of Selz (1913). Berkson and Wettersten compared the theory of Selz with the epistemological consequences of Popper's conception (1974): Selz describes a problem as an *incomplete structure* that must be completed in an evolutionary way, while Popper describes a *conflict* between the old and--more or less--complete theory and new contradictory facts that must be overcome by reconstruction. Following Selz' ideas the AI-approach developed the impasse-driven learning theory.

One consequence of the traditional AI-approach is the fact that all inferences of a heuristic problem or task solving process must have a 'mental logic'. The most glaring problem is that people make mistakes. They draw invalid conclusions, which should not occur if deduction is guided by a 'mental logic' (cf. Reason, 1979; Wehner, 1984).

The doctrine of 'mental logic' can certainly be formulated in a way that meets the methodological criterion of *effectiveness*. The trouble with mental logic is thus empirical. Johnson-Laird (1983) describes six main problems: (1) People make fallacious inferences. (2) There is no definitive answer to the question: Which logic, or logic's, are to be found in the mind? (3) How is logic formulated in the mind? (4) How does a system of logic arise in the mind? (5) What evidence there is about the psychology of reasoning suggests that deductions are not immune to the content of the premises. (6) People follow extra-logical heuristics when they make spontaneous inferences. Why does cognitive psychology constrain the modern research to the doctrine of 'mental logic'?

THE LAW OF REQUISITE VARIETY

Our basic assumption is that human behaviour cannot be erroneous. Of course, human decisions and the behavioural consequences of these decisions can be classified as erroneous and faulty, but from a pure introspective standpoint--from the internal *psycho*-logic of the subject--each decision is the best solution fulfilling all actual constrains and restrictions: lack of information

and/or motivation, lack of knowledge and/or qualification, over or under estimation of the task and/or context complexity etc.

Humans need variety to behave and to adapt. A total static environment is insufferable. Ashby (Ashby, 1958, p. 90) summarises his analysis of regulation and adaptation of biological systems as follows: "The concept of regulation is applicable when there is a set D of disturbances, to which the organism has a set R of responses, of which on any occasion it produces some one, r_j say. The physico-chemical or other nature of the whole system then determines the outcome. This will have some value for the organism, either Good or Bad say. If the organism is well adapted, or has the know-how, its response r_j as a variable, will be such a function of the disturbance d_i that the outcome will always lie in the subset marked Good. The law of requisite variety then says that such regulation cannot be achieved unless the regulator R, as a channel of communication, has more than a certain capacity. Thus, if D threatens to introduce a variety of 10 bits into the outcomes, and if survival demands that the outcomes be restricted to 2 bits, then at each action R must provide variety of at least 8 bits." This condition is one important reason to call human behaviour as *incompressible*.

If we try to translate this 'law of requisite variety' to normal human behaviour then we can describe it as follows: All human behaviour is characterized by a specific extent of variety. This variety of human behaviour can not be reduced to only a 'one best way'. If the system--in which the human has to behave--constrains this normal variety then we can observe 'errors'. In this sense an error is the necessary violation of system's restrictions caused by inappropriate system constrains.

If a system constrains human behaviour to only one possible 'correct solution path' then we can observe a maximum of violations, say errors. Most complex systems are explicitly designed to constrain the operator's behaviour to a minimum of variety. Ulich (1994) arguments against this 'one best way' doctrine of system design because users differ inter- and intra-individually. A system must have a minimum of flexibility to give all users the opportunity to behave in an error-free way. To investigate the relationship between behavioural and cognitive complexity we observe individual behaviour in its 'natural sense'. All deviations of the correct solution path are interpreted as exploratory behaviour caused by the need for variety.

EMPIRICAL STUDIES OF 'ERRONEOUS' BEHAVIOUR

Arnold and Roe assume (1987, p. 205), "that errors may have great functionality for the user, especially during learning. When the user is able to find out what has caused the error and how to correct it, errors may be highly informative. This implies that one should not try to prevent all errors." This hypothesis was tested later in an empirical investigation by Frese et al (1991).

Frese et al (1991) describe the following four reasons for the positive role of errors in training: (1) "the mental model of a system is enhanced when a person makes an error … (2) mental models are better when they also encompass potential pitfalls and error prone problem areas … (3) when error-free training is aspired, the trainer will restrict the kind of strategies used by the trainees, because unrestricted strategies increase the chance to error … (4) errors not only appear in training but also in the actual work situation." They compared two groups: one group with an error training (N=15), and a second group with an error-avoidant training (N=8). In a speed test the error-training subjects produced significant fewer errors than the error-avoidant group.

Gürtler (1988, p. 95) got the same results in the context of sports: "there, where more accidents were counted in the training phase, appeared less – above all of less grave consequences – accidents during the match. Few accidents during the training correlate with accidents of grave consequences during the match."

Wehner (1984) meta-analysed several important articles about human errors and came to the following conclusions: "(1) wrong actions are neither diffused nor irregular, (2) wrong actions appear in the context of successful problem solving behaviour, (3) the significance of errors and faults can only be understood as part of the whole problem solving process, and (4) successful and unsuccessful behaviour coexist."

CONCLUSIONS

First, let us shortly summarise the traditional approach for learning based on training. To avoid unnecessary knowledge about unsafe acts beyond stable system's reaction operators are only trained on key emergency procedures. The beneficial effects of *extensive* training of these key emergency procedures are that they become the dominant and easily retrieved habits from long-term memory when stress imposes that bias. Sometimes emergency procedures are inconsistent with normal operations. To minimise the uncertainty coming from these inconsistencies Wickens demands the following design: "Clearly, where possible, systems should be designed so that procedures followed under emergencies are as consistent as possible with those followed under normal operations" (Wickens, 1992, p. 422).

We try to argument against this position. But, what is wrong with this traditional position? Nothing, of course not! Except the assumption that "knowledge about 'unsafe acts beyond stable system reactions' is *unnecessary* or *dangerous"*. If our experimental results (the *negative* correlation between behavioural and cognitive complexity, see Rauterberg, 1993) are correct (and there is no evidence that they are not correct), then we must conclude that the cognitive structure of experts contains knowledge about unsuccessful trials. What does this result mean for the cognitive structure of mental models about complex systems? Our conclusion is that humans need for effective and correct behaviour in critical situations a huge amount of knowledge 'about unsafe acts beyond stable system reactions'.

REFERENCES

ARNOLD, B. & ROE, R. (1987) User errors in Human-Computer Interaction. In: M. Frese, E. Ulich & W. Dzida (Eds.) Human Computer Interaction in the Work Place. Amsterdam: Elsevier, pp. 203-220.

ASHBY, W. R. (1958) Requisite variety and its implications for the control of complex systems. *Cybernetica* 1(2):83-99.

BERKSON, W. & WETTERSTEN, J. (1982) Lernen aus dem Irrtum [Learning from error]. Hamburg: Hoffmann & Campe.

BOOTH, P. A. (1991) Errors and theory in human–computer interaction. *Acta Psychologica* 78: 69-96.

FRESE, M., BRODBECK, F., HEINBOKEL, T., MOOSER, C., SCHLEIFFENBAUM, E. & THIEMANN, P. (1991) Errors in training computer skills: on the positive function of errors. *Human-Computer Interaction* 6:77-93.

GÜRTLER, H. (1988) Unfallschwerpunktanalyse des Sportspiels [Analysis of accidents in sports games]. In: E. Rümmele (Ed.) Sicherheit im Sport – eine Herausforderung für die Sportwissenschaft. Köln: Strauss, pp. 91-100.

JOHNSON-LAIRD, P. (1983) Mental models. Cambridge (UK): Cambridge University Press.

POPPER, K. (1974) Objektive Erkenntnis: ein evolutionärer Entwurf. Hamburg.

RASMUSSEN, J. (1986) Information Processing and Human-Machine Interaction. (System Science and Engineering Vol 12, A. Sage, Ed.) New York: North-Holland.

RAUTERBERG, M. (1993) AMME: an automatic mental model evaluation to analyze user behaviour traced in a finite, discrete state space. *Ergonomics* 36: 1369-1380.

REASON, J. (1979) Actions not as planned: the price of automization. In: G. Underwood & R. Stevens (Eds.) Aspects of consciousness. London.

REASON, J. (1990) Human Error. New York: Cambridge University Press.

SELZ, O. (1913) Über die Gesetze des geordenten Denkverlaufes–Erster Teil [The laws of thinking–part 1]. Stuttgart: Spemann.

ULICH, E. (1994, 3rd edition) Arbeitspsychologie [Work Psychology]. Stuttgart: Poeschel.

WEHNER, T. (1984) Im Schatten des Fehlers–einige methodisch bedeutsame Arbeiten zur Fehlerforschung [In the shadow of errors–methodological considerations]. (Bremer Beiträge zur Psychologie, Nr. 34, Reihe A-11/84), Bremen: Universität Bremen.

WICKENS, C. (1992) Engineering Psychology and Human Performance (2nd edition). New York: HarperCollins.