Human Errors as an Invaluable Source for Experienced Decision Making

Matthias RAUTERBERG & Roger AEPPLI

Work and Organization Psychology Unit, Swiss Federal Institute of Technology (ETH)
Nelkenstrasse 11, CH-8092 Zürich, Switzerland, Tel: +41-1-632 7082

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. The most well known arguments against the artificial intelligence approach are presented and discussed in relation to expertise, intuition and implicit knowledge. The importance of faults and errors are discussed to describe expertise, and how knowledge about unsuccessful behaviour influences the actual decision making process of experts.

1. Introduction

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. In general, incident surveys in a variety of industries attribute high percentages of critical events to the category 'human error' (cf. [10]). The label 'human error' is very controversial (e.g., [10]). Labeling actions and assessments as errors identifies a symptom, not a cause ([9] p. 120)! The symptom should call forth a more in-depth investigation of how a system comprising people, organizations, and technologies both functions and malfunctions [9] [14] [20]. The conventional 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 [14] 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". Human errors are the most important cause of accidents or near-miss accidents. Nagel [13] presents the results of his analysis: Approximately 70 percent of accidents in aviation operations are classified as human errors. Accidents are categorised as caused by either unsafe acts of persons (e.g., operator error) or by unsafe conditions (cf. [20]). 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.

In the context of cognitive ergonomics most of the known modelling approaches are 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" ([23] p. 258). 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 task solving behaviour and the complexity of the assumed corresponding mental model.

Learning how to solve a specific task with a given system--in this conventional way of interpretation--means that the behavioural complexity is *positively* correlated with the cognitive complexity of the underlying mental model. But this assumption seems to be--in this generality--wrong. 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 classical artificial intelligence paradigm. 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 [12] 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'? To come up with an answer, we have to look on the discussion and review coming from the Dreyfuses.

To Dreyfus [5], the world of the subjective is more important than that of the objective; reality is defined from within--in terms of the individual and his power to perceive and act, to know truths that are unutterable. Dreyfus concludes that some of the things' people do are intrinsically human and cannot be mechanised. To Dreyfus they are *intuition*, *insight*, and *comprehension*--the ability to immediately grasp complex situations, re

M. Rauterberg & R. Aeppli / Human errors as an invaluable source for decision making

solving ambiguities, weeding the relevant from the irrelevant. According to Dreyfus, the conviction that we can formalise reality, explaining everything with rules, began--as far back as the days of ancient Greece--and have become so dominant in the twentieth century that few people question it. This is one explanation for the doctrine of 'mental logic'. Mary Henle [8] declares: "I have never found errors which could unambiguously be attributed to faulty reasoning." She suggests that mistakes arise because people misunderstand or forget premises, and because they import additional and unwarranted factual assumptions into their reasoning (see [9]).

The Dreyfuses [4] argument that only novices use facts and rules. But as we become expert, we forget the rules and act intuitively, automatically adjusting our behaviour to the perceived constraints. Most scientists assume that these kinds of abilities are based on the unconscious and simultaneous processing of signals coming from the eyes, the ears, and the hands. But the Dreyfuses [4] believe that intuition defies rational powers of description, that it can't be computerised. Like judgement and wisdom it is one of the *atomic element* of our world (i.e. *irreducible*). We share the critique of the Dreyfuses, but we do not follow their conclusions. To overcome the deadlock and 'mystical' situation following the Dreyfuses we need a new understanding of knowledge that gives an expert the ability to act intuitively.

2. Empirical Studies of 'Erroneous' Behaviour

Our basic assumption is that human behaviour cannot be erroneous. Of course, human decisions and the behavioural consequences of these decisions can be classified afterwards 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. In this sense we share the position of the Dreyfuses.

2.1. The 'law of requisite variety'

Humans need variety to behave and to adapt. A total static environment is insufferable. Ashby ([2] 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 the observable variety of 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*. Husseiny and Sabri [11] counted 644 "critical incidents" in a representative study analysing complex systems (this is equivalent to an error rate of 16%); they noted that in "non nuclear complex systems" the rate of slips lies only between 3% and 7%. Most complex systems are explicitly designed to constrain the operator's behaviour to a minimum of variety. Ulich [21] 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'. "Error and correct performance are two sides of the same coin" ([19], p. 82). We interpret--in a first step--all deviations of the correct solution path as exploratory behaviour caused by the need for variety, and not as faults or errors (cf. [33])!

2.2. About the relationship between behavioural and cognitive complexity

In one of our experiments [16] we compared the task solving behaviour of novices (subjects without experiences of electronic data processing EDP) with the behaviour of experts (subjects with a lot of EDP experience). We could show, that the complexity of the observable task solving process (the 'behavioural complexity') of novices is significantly larger than the complexity of the observable behaviour of experts. All six novices and all six experts solved exactly the same four different tasks completely (for more details see [15] and [16]). This important result seems to be--at the first glance--trivial. But, what does it really mean? If we assume--and this assumption seems to be very plausible--that the cognitive complexity of the real experts is significantly higher than the cognitive complexity of the novices then we must conclude that the correlation between behavioural and cognitive complexity is *negative*! And, *this* result is not trivial!

Note that in our experiment the minimal task complexity of all four tasks was only reached by one expert (task one: behavioural complexity = 6, the 'one best way'). We do not argument that this minimum cannot be

M. Rauterberg & R. Aeppli / Human errors as an invaluable source for decision making

reached, but to constrain human behaviour only to this minimum leads directly to the paradox of 'errors' of high skilled and over-trained experts. "The 'slip of action' considered here seem to have occurred almost exclusively during the largely automatic execution of highly practised and 'routinized' activities" ([19], p. 83).

Based on the important result, 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 corresponding mental model. If the cognitive structure is too simple and therefor incomplete, then the concrete observable task solving process must be filled up with many heuristics or trial and error behaviour [16]. This negative correlation can be interpreted in three different ways (see Figure 1).
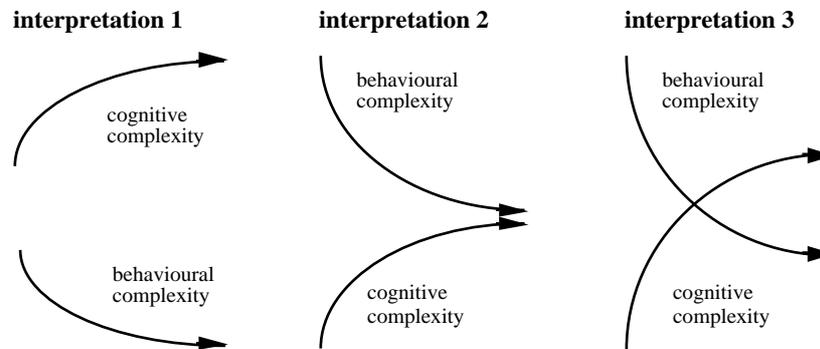

Figure 1. Three possibilities for a negative correlation between behavioural and cognitive complexity.

The *first interpretation* (see Figure 1) means that the complexity of the particular mental task model is at the very early beginning of a learning process significantly higher than the complexity of the observable behaviour. This interpretation is obviously wrong (see also [18]). It is not plausible to assume that humans have substantial knowledge about an unknown task, beforehand!

The *second interpretation* (see Figure 1) means that the absolute value of the behavioural complexity approaches approximately to the absolute complexity value of the corresponding mental model. If the mental model is completed (the tangential point), then we can observe a positive correlation from that point on. This interpretation is congruent with the artificial intelligence assumption of *a positive correlation*--after the tangential point. One problem with this interpretation is that the most mental models are incomplete; they will never reach this tangential point (the "buggy" knowledge, see [3]). But, without this tangential point we can not assume a *positive correlation*.

The *third interpretation* (see Figure 1) means that--at the beginning--the behavioural complexity is significantly higher than the cognitive complexity. After a massive learning process with trial and error behaviour the novice excellerates to an advanced and later on to an expert. During this learning process a hughe amount of knowledge about unsuccessful trials is acquired. The complexity of this specific knowledge is responsible for the complexity overhead in the experts' mental models (see [17]). This is the most important reason why the crossing point does not mean that the complexity of the mental model is equivalent to the complexity of the complete and error free mental model.

The behavioural complexity shrinks to a minimum (constrained by the concrete task) and the cognitive complexity goes tangentially to a maximum that is significantly higher than the minimum of the behavioural complexity: The complexity of the--more or less--complete mental model *plus* the complexity of the knowledge about unsuccessful trials!

2.3. What does we learn from errors?

Arnold and Roe assume ([1] 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 [6]. Frese et al [6] 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 ([7] 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 [24] 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." Van der

M. Rauterberg & R. Aeppli / Human errors as an invaluable source for decision making

Schaaf ([22] p. 85) concludes, that "every time an operator, manager, procedure, or piece of equipment 'behaves' in an unexpected way and thereby prevents a likely breakdown of the production system … or restores the required levels of safety and reliability, these *positive deviations* could be detected, reported and analysed in order to improve the qualitative insight into system functioning on the whole." This conclusion is not only valid for the global 'accident driven' design process "on the whole", this statement is also valid on the individual level of operating a complex system. Breaking through the system's boundaries is the only way to get a really deep understanding of the whole system. "In unusual or novel situations, however, it may be essential to have a thorough understanding of the functional structure of the automated systems and to be able to use this knowledge in operationally effective ways"([25] p. 57).

3. Conclusion

The following thought experiment can help us to demonstrate the importance of knowledge about unsuccessful trials: In a given--maybe problematic--situation the operator has only five different possibilities to overcome the troubles (say A, B, C, D, E), but only one possibility is the correct one (e.g. C), then the chance to find C-- for a total unexperienced person--is 20%. After the experience that A, B and E are not the correct one then the chance to find C in the third trial increases up to 50%; the operator has only to choose between C and D! This is the reason why we think that knowledge about unsuccessful behaviour and errors is an invaluable source for experienced decision making! An expert is not only a person who knows how to solve a given problem, he knows also, what--and sometimes why--specific possibilities are not helpful or--in particular cases--dangerous.

References

[1] 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.

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

[3] Chi, M., Glaser, R. & Farr, M. (1988) The nature of expertise. Hillsdale: Erlbaum.

[4] Dreyfus, H. & Dreyfus, S.(1979) Mind over machine–the power of human intuition and expertise in the era of the computer. New York: The Free Press.

[5] Dreyfus, H. (1979) What computers can't do–the limits of artificial intelligence. New York: The Free Press.

[6] 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.

[7] 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.

[8] Henle, M. (1978) Foreword to R. Revlin & R. E. Mayer (Eds.) Human reasoning. Washington: Winston.

[9] Hollnagel, E. (1991) The phenotype of erroneous actions: implications for HCI design. In: G. Weir & J. Alty (eds.) Human Computer Interaction and Complex Systems. London: Academic Press, pp. 73-121.

[10] Hollnagel, E. (1993) Human reliability analysis: context and control. London: Academic Press.

[11] Husseiny, A. & Sabri, Z. (1980) Analysis of human factor in operation of nuclear plants. *Atomenergie, Kerntechnik* vol. 2.

[12] Johnson-Laird, P. (1983) Mental models. Cambridge (UK): Cambridge University Press.

[13] Nagel, D. (1988) Human error in aviation operations. In: E. Wiener and D. Nagel (Eds.) Human Factors in Aviation. San Diego: Academic Press, pp. 263-303.

[14] Rasmussen, J. (1986) Information processing and human-machine interaction. (System Science and Engineering Vol 12, A. Sage, Ed.) New York: North-Holland.

[15] Rauterberg, M. (1992) A method of a quantitative measurement of cognitive complexity. In: G. van der Veer, M. Tauber, S. Bagnara & M. Antalovits (Eds.) Human-Computer Interaction: Tasks and Organisation. Roma, pp. 295-307.

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

[17] Rauterberg, M. (1995) From novice to expert decision behaviour: a qualitative modelling approach with Petri nets. In: Y. Anzai, K. Ogawa & H. Mori (Eds.) Symbiosis of Human and Artifact: Human and Social Aspects of Human-Computer Interaction. (Advances in Human Factors/Ergonomics, Vol. 20B, pp. 449-454), Amsterdam: Elsevier.

[18] Rauterberg, M. & Aeppli, R. (1995) Learning in man-machine systems: the measurement of behavioural and cognitive complexity. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics 'Intelligent Systems for the 21st Century'. (Vol. 5, IEEE Catalog No. 95CH3576-7), pp. 4685-4690.

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

[20] Reason, J. (1990) Human error. New York: Cambridge University Press.

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

[22] Van der Schaaf, T. (1992) Near miss reporting in the chemical process industry. Proefschrift, TU Eindhoven.

[23] Van der Veer, G., Guest, S., Haselager, P., Innocent, P., McDaid, E., Oesterreicher, L., Tauber, M., Vos, U. & Waern, Y. (1990) Desiging for the mental model: an interdisziplinary approach to the definition of a user interface for electronic mail systems. In: D. Ackermann and M. Tauber (Eds.) Mental Models and Human-Computer Interaction 1. Amsterdam: North-Holland, pp. 253-288.

[24] 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.

[25] Woods, D., Johannesen, L., Cook, R. & Sarter, N. (1994) Behind human error: cognitive systems, computers, and hindsight. (State-of-the-Art Report CSERIAC SOAR 94-01) Wright-Patterson Air Force Base, Ohio (USA).