Human Errors: Disadvantages and Advantages

Matthias Rauterberg[1] and Daniel Felix[2]

[1]Work and Organization Psychology Unit
[2]Institute of Hygiene & Applied Physiology
Swiss Federal Institute of Technology (ETH)
Nelkenstrasse 11, CH-8092 Zürich, Switzerland
URL: http://www.ifap.bepr.ethz.ch/~rauterberg/

ABSTRACT

The traditional paradigm for learning and training of operators in complex systems is discussed and criticised to react on the 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 a new metaphor for cognitive structures to describe expertise, and how knowledge about unsuccessful behavior influences the actual decision making process of experts.

Keywords: human error, meta learning, mental model, experience, expertise

1. INTRODUCTION

Why is this type of statements "I learned more from my defeats than from my victories" (Napoleon, ca. 1819) sometimes (or always) true? To answer this question we need a new understanding of human errors, inefficient behavior, and expertise. In this paper we will discuss the importance of learning from unsuccessful behavior. 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. Hollnagel, 1993). The label 'human error' is very controversial (e.g., Hollnagel, 1991). Labeling actions and assessments as errors identifies a symptom, not a cause (Hollnagel, 1991, 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 (Rasmussen, 1986; Reason, 1990; Hollnagel, 1993).

The conventional understanding of human errors is characterized by a *negative* valuation of erroneous behavior, something that must be avoided. The Western Culture is constrained by this *taboo*: Not to talk about faults, errors and other dangerous behavior! 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, because we do not promote a negative public image of ourself. The Webster's New Encyclopedic Dictionary defines five different meanings of an error (1996, p. 340): "(1) a: deviation from a code of behavior, b: an act involving an unintentional deviation from truth or accuracy, c: an act that through ignorance, deficiency, or accident fails to achieve what should be done, …; (2): the quality or state of erring; (3): a false belief or a set of false beliefs; (4): something produces by mistake; (5): the difference between an observed or calculated value and the true value". There are two important components in Webster's definition: (I) a norm or standard (e.g., code, truth, accuracy, true value, etc.) to refer to, and (II) an unintentional act of an actor (e.g., human, production system, etc.). One central question is: Are we--as humans--able to make *intentionally* an error? Ofcourse, we can intentionally decide to behave in a 'wrong' way, to make foolish decisions, etc.; but is such kind of situations a valide perspective on real life situations in which 'real' human errors occur (e.g., a pilot error in a flying plane)?

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" (p. 149). He stresses the point that faults and errors "can only be defined with reference to human intentions or expectations" (p. 149). One important aspect in this definition is the 'reference to human *intentions* or *expectations*'. Let us discuss the importance of expectation and prediction in a more detailed way in the context of action theory.

2. ACTION THEORY AND HUMAN ERRORS

Leontyev's three-level schema (Leontyev, 1978, see Figure 1) describes the sphere of analysis of human activities and directs the attention to the transformations going on between three levels: motive–activity, goal–action, and instrumental conditions–operations (Leontyev, 1978; Engeström, 1991; Ulich, 1994; see Figure 1). These three levels are organised in a hierarchical structure where the top level of activities includes several actions that are performed by appropriate operations. In a 'pure' objective way only the operational level can be observed and analysed. The goal setting and motivational level must be derived or investigated by indirect methods (e.g., questionnaire, interview, thinking aloud, etc.) based on the introspective observations of the investigated subjects. (Note the similarity to the model of Rasmussen (1986): knowledge based, rule based, and skill based level.)
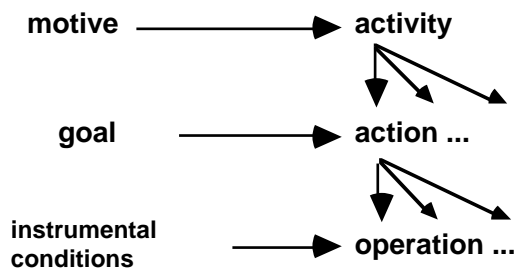

Figure 1. The three levels schema of the activity theory of (Leontyev, 1978).
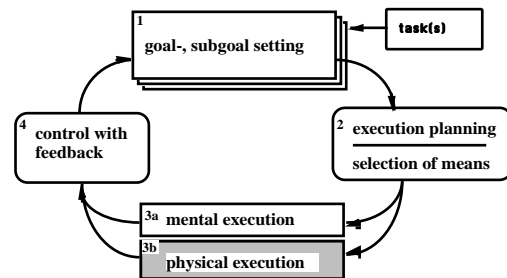

Figure 2. The complete action-cycle: the central concept of the action regulation theory (Hacker, 1994).

Action regulation theory offers a coherent body of principles for human-centered task and work analysis and design (Hacker 1994). For Hacker (1986, p. 61), the work order, its interpretation or its acceptance as a work task is "the central category of psychological consideration of activity..., as decisive specifications for the regulation and organisation of the activities occur with the 'objective logic' of its contents". This quotation makes clear that for all, who follow the activity or action theory, the *task* becomes very important in the analysis of behavior. Great importance is attached to the concept of the *complete task.*

Characteristics of complete tasks, which must be taken into consideration when analysing and/or designing human activities (e.g., job design), are, according to the concept of action regulation theory presented here (see Figure 2, and Ulich, 1994, p. 168): (1) Independent setting of goals which are embedded in the superimposed goals; (2) Independent action preparation in the sense of taking on planning functions; and, selection of the means including the necessary interaction for goal attainment; (3) Mental or physical performance functions with feedback on performance pertaining to possible corrections of actions; (4) Control with feedback on results and the possibility of checking the results of one's own actions against the set goals.

Incomplete activities--or partialized actions--"lack, to a large extent, possibilities for independent goal-setting and decision-making, for developing individual work styles or sufficiently precise feedback" (Hacker, 1987, p. 35). Complete activities (or tasks) offer the possibility of setting goals and sub-goals, as well as of offering decision-making possibilities during the various phases of task completion, and therefore provide latitude of activity or action. Complete activities are therefore becoming fundamental for realising the concept of action regulation. Goals are organised in a hierarchical tree structure (Hacker, 1986). The complete action cycle (see Figure 2) has a fractal or recursive structure: Each component of the complete action cycle can be analysed with an embedded and subordinated action cycle. Learning in form of chunking occurs over all three levels in Figure 1.

Learning is a permanent and irreversible process that is regulated by optimising the range of expectations based on situated decision making processes. An self regulated actor (e.g., human) maximises his or her anticipation processing by optimising the situated decison making: the gap between the range of the actor's expectations and the range of the system's behavior and feedback, resp. (see Figure 3). If the difference between the range of actor's expectations and the range of real system's outcome exceeds a particular threshold, a breakdown occurs. These breakdowns are unexpected inconsistencies between the actor's predictions and the real situation. This kind of a breakdown is caused either by internal factors (e.g., the incorrect or incomplete actor's mental model of all relevant circumstances in the actual situation), or by external factors (e.g., non ergonomical envrionment conditions). 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 (cf. Berkson and Wettersten, 1982), and how are these insights and the derived knowledge embedded in the individual cognitive structure.

One consequence of the traditional approach in the context of artificial intelligence (AI) is the fact that all inferences to build the expectations and predictions 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). As the saying is, *experience* is a wonderful thing; it enables you to

26

recognize an error when you make it again! The most important aspect for humans is to get a real chance to learn from errors (see Figure 4). Learning in this sense is the only way to enable humans to survive as race.
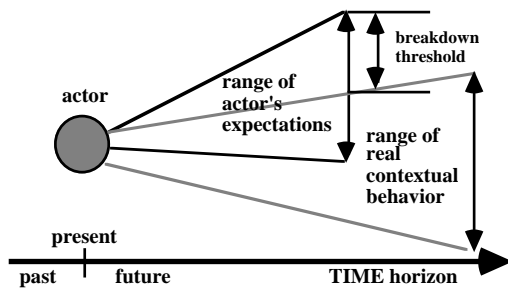

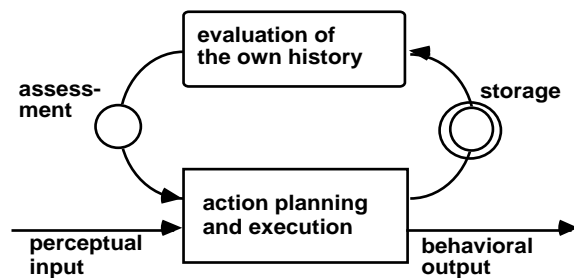**Figure 3.** The actor's expectations on his time horizon.

Figure 4. The different components of human learning.

Following the tradition of Gestalt-psychology (cf. Duncker, 1935) Ulich (1994, p. 319) differentiates between the 'foolish' and the 'good' error. Köhler (1917) could show that the 'good' error is an important precondition for insight learning and optimal problem solving processes. "Knowledge and error flow from the same mental sources, only success can tell the one from the other" (Mach, 1905, p.84). Success in problem solving can be paraphrased as: 'If you tried to do something and failed, you are vastly better off than if you had tried to do nothing and succeeded.' Only activity--successful *and* unsuccessful behavior--coupled with self optimizing learning can modify the mental knowledge in an appropriate direction. Let's prove this hypothesis with empirical studies.

3. EMPIRICAL STUDIES OF 'ERRONEOUS' BEHAVIOR

Husseiny and Sabri (1980) 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 behavior to a minimum of variety. All deviations of the correct solution path are interpreted as errors! 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*.

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 behavior, (3) the significance of errors and faults can only be understood as part of the whole problem solving process, and (4) successful and unsuccessful behavior coexist."

Van der Schaaf (1992, 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 (cf. Wehner, 1984). "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"(Woods et al., 1994, p. 57).

4. CONCLUSION

In this paper the traditional paradigm for learning of humans in complex environments is discussed and criticised. There is a strong influence (the doctrine of 'mental logic') coming from research carried out in the AI context. 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 learning and activity. Activity coupled with self optimizing learning has a strong influences on the development of cognitive structures to generate expertise. Knowledge about unsuccessful behavior influences the actual decision making process of experts by *exclusion*, knowledge about successful behavior by *inclusion* of the correct aspects and steps.

"Action Theory seems to be an integrative long-term approach that is still developing especially with the development of hierarchically subordinate sub approaches. Action Theory is still more a heuristic broad-range framework than a final theory. … The integrative power of Action Theory will bridge some interrelated gaps: the gap between cognition (and knowledge) and action, the gap between cognition and motivation (see goal orientation), and even the gap between basic and more applied approaches …" (Hacker, 1994, p. 113). Our research interest is bridging both gaps. The sum of all cognitions are determined by the mental knowledge: the mental model.

Our basic assumption is that human behavior cannot be erroneous. Of course, human decisions and the behavioral 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 expectation into the near future fulfilling all actual constrains and restrictions. A breakdown--an error--is caused by internal and external factors: lack of information and/or motivation, lack of knowledge and/or qualification, over or under estimation of the task and/or context complexity, non ergnonomical design of tools, inappropriate feedback etc.

REFERENCES

Arnold, B. and Roe, R. (1987). User errors in human-computer interaction. In: M. Frese, E. Ulich and W. Dzida (Eds.) Human Computer Interaction in the Work Place. Elsevier, pp. 203-220.

Berkson, W. and Wettersten, J. (1982). Lernen aus dem Irrtum [Learning from error]. Hoffmann & Campe.

Duncker, K. (1935). Zur Psychologie des produktiven Denkens [The psychology of productive thinking]. Springer.

Engeström, Y. (1991). Activity Theory and Individual and Social Transformation. Activity Theory 7/8: 6-17.

Frese, M., Brodbeck, F., Heinbokel, T., Mooser, C., Schleiffenbaum, E. and 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]. In: E. Rümmele (ed.) Sicherheit im Sport--eine Herausforderung für die Sportwissenschaft. Strauss, pp. 91-100.

Hacker, W. (1986). Arbeitspsychologie [Work psychology]. Huber.

Hacker, W. (1987). Software-Gestaltung als Arbeitsgestaltung [Software design as work design]. In: K.-P. Fähnrich (ed.) Software-Ergonomie. State of the Art 5 (pp. 29-42). Oldenbourg.

Hacker, W. (1994). Action regulation theory and occupational psychology. Review of German empirical research since 1987. The German Journal of Psychology 18(2): 91-120.

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

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

Husseiny, A. and Sabri, Z. (1980). Analysis of human factor in operation of nuclear plants. *Atomenergie, Kerntechnik* Vol. 2.

Köhler, W. (1917). Intelligenzprüfungen an Menschenaffen [Intelligence tests with antropoids]. (New edition 1973) Springer.

Leontyev, A. N. (1978). Activity, consciousness, and personality. Prentice Hall.

Mach, E. (1905). Knowledge and error. (New edition 1976) Dordrecht.

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

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

Reason, J. (1990). Human error. Cambridge University Press.

Ulich, E. (1994, 3rd edition). Arbeitspsychologie [Work psychology]. Schäffer-Poeschel.

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

Webster's (1996). New Encyclopedic Dictionary. Könemann.

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), University of Bremen.

Woods, D., Johannesen, L., Cook, R. and 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).