Accident Analysis And "Human Error"

(See also: "Why Human Error Is A Meaningless Term")

An accident can be defined as an unexpected event with an undesirable outcome. Both criteria are important, since neither an expected unpleasant event, nor an unexpected pleasant one qualify as an accident. As long as accidents have happened, people have tried to understand them, both to satisfy themselves that the causes were not in conflict with the accepted wisdom of the world, and to help finding ways to prevent that the same - or a similar - accident should occur again.

The concern for accidents and their causes has attracted considerable interest in the last half of the 20th Century, mainly due to a number of spectacular accidents in complex industrial systems. The more famous of these are the explosion at the Flixborough plant (1974); the accident at the nuclear power plant at Three Mile Island (1979); the explosion of the Challenger space shuttle (1986); the meltdown at the nuclear power plant in Chernobyl (1986); the aeroplane collision at Tenerife (1977); and the multitude of problems on the space station MIR (1998). In accidents such as these, failures of human action and judgement have often been seen as part of the causes, and in a growing number of accidents the main cause has been attributed specifically to "human error". In the 1960s the number was around 30%, but grew during the following decades so that the number at present often is put as high as 70-90%. This development is, of course, not the expression of a simple fact or relation. There are rather a number of reasons for this trend, for instance: (1) that technological systems have become more complex, hence more difficult to control; (2) that improved models and method for "human error" analysis have made this cause more likely; (3) that technological systems have become more reliable, hence raising the relative number of other causes; or (4) that it is sometimes cheaper and more convenient to put the blame on a human than to redesign an entire system.

Models And Methods For Accident Analysis

An accident model is a generalised description of how an accident may have happened. Such models are invariably based on the principle of causality, which states that there must be a cause for any observed event. The early accident models tended to see accidents as caused either by failures of the technology or incorrect human actions ("human errors"). This view was gradually extended to recognise both the contribution of latent system states, and the complexity of conditions that could end in an incorrectly performed human action - even leading to the extreme notion of "error forcing" conditions. In contemporary accident models, a distinction is made between actions at the "sharp" end, which often are the initiating events, and actions at the "blunt" end, which create the conditions that either make an action failure near inevitable or turn minor mishaps into major disasters.

Despite these developments, specifically the increasing sophistication in accounting for the organisational determinants of accidents, there is an almost intransigent preference to refer to "human error" as a singular concept. The history of accident analysis clearly demonstrates that the notion of a cause itself is an oversimplification, since a cause represents a judgement in hindsight rather than an unequivocal fact. This acknowledgement notwithstanding, accident models seem to be firmly entrenched both in the idea that a "true" or root cause can be found, and in the idea that "human errors" necessarily must be part of the explanations. This view should be contrasted with the pragmatic, so-called ecological, view which points out, firstly, that action failures are both an unavoidable and necessary element of efficient human performance and, secondly, that the same types of action failures occur across tasks and domains. Without making shortcuts or using heuristics people would be unable to work effectively, and without failing every now and then they would not be able to learn. The challenge for cognitive systems engineering and artefact design is, of course, to make sure that the systems are so robust that minor variations in performance do not lead to fatal consequences, yet so sensitive that users have the freedom necessary to create and optimise control strategies.

The Nature of "Human Error"

Since the 1960s many attempts have been made to provide a technical definition of the concept of error. Yet despite the fact the term error has a relatively simplistic meaning in everyday life, the term is extremely difficult to pin down precisely when considered from a technical point of view. One reason is that the same term, "error", is used to denote either (1) an outcome or consequence, (2) the act or event itself, or (3) the possible cause. Such ambiguity is clearly not conducive for developing accounts of causes and effects. A further reason is that analyses of the nature and origins of error from different professional points of view, often have quite different and at times incompatible premises. Thus, an engineer might prefer to view the human operator as a system component subject to the same kind of successes and failures as equipment. Psychologists, on the other hand, often begin with the assumption that human behaviour is essentially purposive and can only be fully understood with reference to subjective goals. Finally, sociologist have traditionally ascribed the primary error modes to features of the prevailing socio-technical system and in a sociological analysis items such as management style and organisational structure are often hypothesised as the mediating variables that influence error rates.

Irrespective of the above differences there seem to be at least three intuitive parts to any definition of error:

- First, there must be a clearly specified **performance standard** or **criterion** against which a response can be measured. Engineering reliability analysts have traditionally used objective measures, such as system parameters, as the standard for acceptable performance. In contrast, investigators working from the standpoint of cognitive psychology have tended to rely on subjective criteria such as the momentary intentions, purposes and goal structures of the acting individual. According to this view there are two basic ways that an action can go wrong. In one case the intention to act is adequate but a subsequent act does not go as intended. Here the "error" is conventionally defined as a slip. In another case, actions proceed according to plan but the plan is inadequate. Here the "error" is usually classed as a mistake (Norman, 1981).
- Second, there must be an **event** which results in a measurable **performance shortfall** such that the expected level of performance is not met by the acting agent. Many researchers of human performance ascribe to a pessimistic view according to which "human errors" provide strong evidence of design-defects of the human information processing system. Considerable research has therefore aimed at identifying the design-defects of the human mind. Once identified the assumption is that guidelines can be developed to determine for which situations the human operator can and cannot be trusted with the care of complex plant and equipment. A more optimistic viewpoint is found in a line of research that emphasises that most "human errors" have their origins in processes that perform a useful and adaptive function. This approach takes a much more beneficial view of the variability of human cognition and performance, and relate "error mechanisms" to processes that underpin intelligent functioning and especially the human ability to deal with complex data that are characterised by a high degree of ambiguous and uncertainty.
- Third, there must be a degree of **volition** such that the actor has the opportunity to act in a way that will not be considered erroneous. Thus, if something is not avoidable by some action of the person, it is neither reasonable nor acceptable to speak of an error. This, presumably, also includes the so-called error forcing conditions, even though the full implications of that approach have not yet been realised. Factors that occur outside the control of the individual, for example, "Acts-of-God" or acts of nature, are therefore better defined as accidents.

Accident Analysis

Accident analysis denotes the set of methods and principles that are used to find, in a systematic manner, the cause – or causes - of an accident. The engineering practice of accident analysis is based on the assumption that there is a true or real cause of an accident – sometimes referred to as the root cause. This should be contrasted with the more recent view of, for instance, cognitive systems engineering, according to which the cause is a social construct (Woods et al., 1994). Formally speaking, a cause is the identification, after the fact, of a limited set of aspects of the situation that are seen as the necessary and sufficient conditions for the effect(s) to have occurred. In order for a cause to be an acceptable explanation, it must be possible to associate it unequivocally with a system structure or function (people, components, procedures, etc.), and to do something to reduce or eliminate the cause within accepted limits of cost and time.

Accident analysis methods are usually based on a hierarchy of causes, although most hierarchies represent convenience rather than a taxonomy. The classification causes has undergone a significant development during the last 30 years or so (Figure 1). In the early days of accident analysis, i.e., the 1950s and 1960s, causes were roughly seen as belonging to three groups called: technical failures, "human error", and anything else – called "other". The development in the categories, paralleled by a development in accident analysis methods, has in particular taken place within the groups of "human error" and "other". As far as the latter is concerned, the major new categories have described various causes of organisational failures, including safety culture, quality assurance, and pathogenic organisations. As far as the former is concerned, the development has been to distinguish between causes in different types of work (management, maintenance, design, and operation), and further to increase significantly the categories for "human errors", going from omissions and commissions to detailed information processing failures or "cognitive errors".

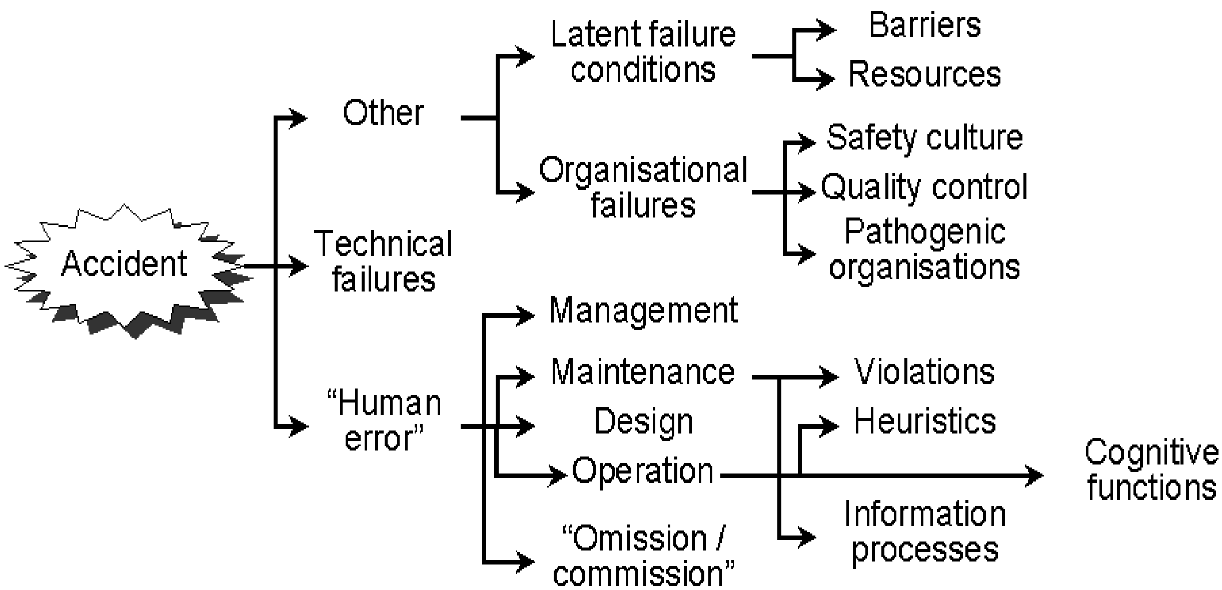

Figure 1: Development in the classification of causes.

Traditional Human Factors and Ergonomic Approaches. Models of operator error developed within a human factors and ergonomic tradition generally provide a useful framework the development of practical tools to classify errors in terms of their external manifestations. The objective of such analyses is usually to predict the probability of **simple** error forms, which may occur in human-machine interactions, such as the distinction between "errors of omission" and "errors of commission". Although more complex models of human error have been specified at the level of observable behaviour, the basic limitation of the approach is that it provides little information regarding the possible psychological causes. Thus, the analytic capability of traditional ergonomic models is typically quite low and resultant models contain few general principles that can be used to construct a description of failures in terms of underlying psychological functions.

Information Processing Models. In contrast to traditional human factors approaches, information processing models have a highly developed basis and analytic capability, but cannot always easily be converted to useful and practical tools for performance analysis or prediction. The analytic capability of information processing models derives mainly from the large number of general statements relating to error tendencies or mechanisms that can be incorporated into such models. Since the models often are of a rather loose nature, many of these explanations are unfortunately *ad hoc* and lack a clear theoretical foundation. (For example, it is commonly assumed that short-term memory is capable of processing a strictly limited number of chunks of information and that demands in excess of this amount typically cause the short-term memory system to fail.) Furthermore, when information processing models permit the formulation of general predictions that certain types of error will occur, the extent to which such predictions transfer to a real-world setting is unclear. Analyses of accident and near-miss reports, for example, frequently describe situations where information processing failures occurred although the situations were well within the presumed performance capabilities of the human operator. Conversely, there are many well documented instances where operators have succeeded to control a situation where information processing models would have predicted failure.

Information processing models also face the problem that the capabilities of the human are far from being stable or constant. In analogy with physical artefacts, each of the components of the human information processing system are assumed to have identifiable capabilities, hence also boundaries (high and low). Yet despite heroic attempts to define and measure the basic capabilities, success has been strictly limited. This fact could be taken as an indication that the underlying approach is wrong, i.e., that the human is not really an information processing system - although people in many ways can be described as such with considerable benefits.

Cognitive Systems Engineering. Models of erroneous actions generated from the standpoint of cognitive systems engineering tend to do quite well in terms of analytic capability, predictive capability, technical basis, relation to existing taxonomies, practicality, and cost/effectiveness. Currently, the main shortcoming is that they fall short of the ideal with regard to their ability to predict likely error forms. They are, however, not bettered in this respect by either ergonomic models or information processing models. Moreover, cognitive systems models are particularly strong in terms of their technical content because they are based on viable and well-articulated models of human action. Cognitive systems engineering focuses more on how human performance is influence by the internal and external context than on how it can be explained by hypothetical models of cognitive functions and structures. The models therefore provide a better basis for the development of practical tools for error analysis and subsequent reduction.

Socio-Technical Approaches. Since the late 1980s, a further approach to human error modelling has been on the rise, which places the focus on the organisational and operational context in which the work takes place (Reason, 1997). The main background for this approach is the recognition that a common factor in a large number of accidents is the organisational condition. As the view on human error gradually became more refined it was acknowledged that incidents evolved through a conjunction of several failures and factors, and that many of these were conditioned by the context. This evolved into the previously mentioned notion of people at the "sharp end" of a system as contrasted with people and events at the "blunt end". The basic idea is that human performance is a compromise between the demands from the monitored process and the resources and constraints that are part of the working environment. The demands and resources meet at the sharp end, and the result shows itself in the actions that people make. The demands come from the process, but by implication also from the organisational environment in which the process exists. The resources and the constraints are more explicitly given by the organisational context, e.g. in terms of procedures, rules, limits, tools, etc.

The socio-technical approach is in many ways the inverse of the classical information processing approach, which concentrates on the internal human information processing mechanisms. The socio-technical approaches therefore run the same danger of focusing on one aspect of the situation, the context, thereby neglecting the other. The socio-technical approaches are nevertheless valuable as reminders of the need to consider both sides - cognition and context. At present the socio-technical approaches have not been developed to the stage where they can provide a potential explanation for human erroneous actions. As a matter of fact, they tend to reduce the relative contribution of individuals and of the operator's cognitive functions. It is not inconceivable that in the future more mature socio-technical approaches will be developed, which can enrich the explanations of human erroneous actions. At present, however, they have not reached this stage.

Conclusions

One lesson of almost 40 years of research into "human error" is that it is not a specific category. Rather, everything is a "human error" if only one goes back far enough. The logical consequence of that is that the analysis may also go a step beyond the "human error", and thereby avoid simple-minded explanations. Although the relativity of the analysis is an unresolved issue, one inescapable conclusion is the realisation that human actions are not a special type of causes. Human erroneous actions are rather a symptom of the conditions of work, and of the demands from the organisation and the environment (the end users). Technology is often used in clumsy ways that are not well adapted to the needs of the people at the "sharp end". This results in unnatural and unusual demands on the practitioner that tend to congregate at the higher tempo or higher criticality periods of activity. The manifestation of this has been an apparent epidemic of failures labelled as "human error". Yet it is only by seeing erroneous actions as an indication of missing adaptation, that the findings from accident analyses can be used constructively to understand and prevent accidents.

Literature

Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, *88*, 1-15.

Reason, J. T. (1997). *Managing the risks of organizational accidents*. Aldershot, UK: Ashgate.

Woods, D. D., Johannesen, L. J., Cook, R. I. & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers and hindsight*. Columbus, Ohio: CSERIAC.

Why "Human Error" Is A Meaningless Concept

A number of people, including myself, have for some years argued against an oversimplified use of the concept of "human error", and specifically against developing specific theories of "human error". The arguments were first put forward in the heyday of "human information processing" (which, by the way, is another term that should be abandoned), hence had relatively little effect. The popularity in recent years of "contextual" or "situated" approaches, and specifically of the school of naturalistic decision making, has to some extent preempted the objections to the use of the term "human error" - although not completely.

To illustrate the longevity of this debate, I have found a position paper written in preparation of a conference on "human error". This position paper was never published, although it was referred to in the proceedings that (with some delay) followed the conference. Apart from correcting spelling errors, the contents of the position paper is unchanged - bad grammar and all. Some newer references are mentioned after the position paper.

Position Paper for

NATO CONFERENCE ON HUMAN ERROR

August 1983, Bellagio, Italy

Erik Hollnagel
OECD Halden Reactor Project, Norway

INTRODUCTION

The organisers have provided us with a both stimulating and irritating list of questions relating to the topic of the conference: "Human Error". My first intention was to try to answer the questions one by one, or at least group by group. However, after some consideration it appeared to me that the questions contained an important bias, and that it was necessary to discuss this before trying to answer the questions.

The bias that I find is the assumption that there exists something called "Human Error" about which meaningful questions can be posed - and answered! "Human Error" thereby gets the status of a concrete phenomenon as, for instance, decision making. It is, however, obvious that "Human Error" does not refer to something observable, in the same sense as decision making does.

Decision making is an example of a function on the psychological level. It can be used to denote the activity of making a decision as well as the covert function behind action. In the first sense it is observable in an everyday meaning of the term. In the latter it may have to be inferred from observed behaviour, but the inferences need not be very elaborate. And in both cases it is regarded as a function. As a rule I will assume that functions, such as decision making, can be detected or seen in a straightforward way by an observer - although they may not be directly observable from a more stringent philosophical point of view.

"Human Error" is, however, not a function, but a cause (or, to be precise: an assumed cause). We can use the term in a functional sense, as when we say that someone is making a mistake or an error. But in neither case is the "Human Error" an activity, nor the result of an intention. It is simply a contradiction of any reasonable definition to say that a person can make an error intentionally. Accordingly, it would be meaningless to call it a function.

It may be argued that "Human Error" characterises the outcome of an action rather than the cause. Our classifying an outcome as a "Human Error" is a misuse of the terminology. What is meant is rather that the outcome was caused by a "Human Error". Neither can the "Human Error" be the activity that leads to the outcome. We cannot classify an activity as being a "Human Error", since that would assume that making the error was intentional. As that is not the case, it will be more correct to classify the activity as a failure to accomplish the intended outcome.

Being a cause, "Human Error" must be inferred from observations rather than observed directly. Other examples of such non-observables are " goal", " memory", etc. Consequently we must specify the observations from which the inferences are made. These observations will normally be about a particular performance or segment of a performance. We may observe the performance of an operator, classify it as being incorrect, and determine the cause to be a "Human Error". But in no case can we observe the "Human Error" directly.

Since "Human Error" is inferred, it is not necessarily unique. Another way of saying this is by noting that "Human Error" is just one explanation out of several possible for an observed performance (or more precisely, a part of an actual performance description, cf. Hollnagel et al., 1981). The analysis is normally carried just far enough to find a plausible explanation. If an explanation, which refers to the technological parts of the system, cannot be found the category "Human Error" is normally used (cf. Rasmussen, 1981). It is only when the analysis is carried beyond this point that we may realise that an explanation in terms of "Human Error" is insufficient.

I am not saying this to begin a philosophical discussion. The point I want to make is that we should start with an analysis of the empirical data we have, and from that derive what "Human Error" is. I will try to do so in the following, using a functional analysis based on systems theory.

Since my major source of experience is operators in control of a complex process (a nuclear power plant), I will assume that the system we deal with is a Man-Machine System (MMS) that functions as a process control system. By an MMS I mean a system that is composed of one or more operators and one or more machines (usually computers) that are designed to support the control of the process. A particular example of this approach is the Cognitive Systems Engineering (cf. Hollnagel & Woods, 1983). In the following, I will address the six groups of questions, although in a different order than presented by the organisers.

THEORY

When the performance of an MMS is being observed (and evaluated) a mismatch may be detected between the actual and the intended system states, or between the achieved results and the goal. The detection of this mismatch presumes that a description of the intended system state (or goal) is available. The mismatch is assumed not to be random, hence to have an identifiable cause. Finding the cause amounts to accounting for the observed variance in the system's performance. If faults in the technological parts of the system cannot be found, the solution is generality to assign the variance (or residual variance) to the human component, hence to use "Human Error" as an explanation.

The detection of this mismatch is thus the observational basis for inferring the existence of a "Human Error". It should be noted that if there is no observed mismatch, there will be no reason to look for a cause. Variations in performance do not necessarily lead to undesired outcomes, hence mismatches. They may, for instance, be detected and corrected by the system at an early stage or the environment can be sufficiently friendly and forgiving. There will consequently be cases of performance variability that remain unnoticed. From the point of view of a theory of "Human Error" they are, however, just as important as the cases where a mismatch is observed, and should therefore be accounted for by it.

The crucial point thus is a mismatch between intended and actual outcomes of action. If the functional analysis is carried one step further, it will show that the cause of the mismatch can be located either in the selection of the goal for the action (the formation of the intention) or in the execution of the actions designed to achieve that goal. One may even distinguish between a larger number of categories by using one of the models of human decision making, or a theory of human performance. But this actually reduces the need for a specific theory of "Human Error", since the observed discrepancies instead can be explained by referring to, for instance, a performance theory. That may furthermore have the virtue of focusing on the situation and context in which the MMS must function, and the interaction between its inherent characteristics and the environmental constraints.

Consequently, I do not think that there can be a specific theory of "Human Error", nor that there is any need for it, This is not because each error, as a "something" requiring an explanation, is unique, but precisely because it is not, i.e., because it is one out of several possible causes. Instead we should develop a theory of human action, including a theory of decision making, which may be used as a basis for explaining any observed mismatch A theory of action must include an account of performance variability, and by that also the cases of where "Human Error" is invoked as a cause.

Observed mismatches in performance are always caused, in the sense that they can be analysed until the necessary and sufficient conditions for their occurrence have been established. In some cases they may be classified as random, but that just means that the natural performance variability is sufficient to account for the mismatch, hence that no definite "other" cause has been identified.

Since errors are not intentional, and since we do not need a particular theory of errors, it is meaningless to talk about mechanisms that produce errors. Instead, we must be concerned with the mechanisms that are behind normal action. If we are going to use the term psychological mechanisms at all, we should refer to "faults" in the functioning of psychological mechanisms rather than "error producing mechanisms". We must not forget that in a theory of action, the very same mechanisms must also account for the correct performance which is the rule rather than the exception. Inventing separate mechanisms for every singe kind of "Human Error" may be great fun, but is not very sensible from a scientific point of view.

Even though we do not have a "Theory of Error", it makes sense to distinguish between endogenous and exogenous causes for the performance mismatch. There are certainly cases where the mismatch can be attributed to external causes, such as a bad interface design, lack of operational support, misleading messages, etc. Similarly, there are cases where the causes are of an internal rather than external nature. I do, however, believe that in most cases the cause is best described as a mixture. Stress, for instance, is often caused by (situationally) unreasonable demands to the operator. And deficiencies in the design of the interface may often be compensated by the adaptability of the operator (cf. Taylor & Garvey, 1959). Replacing a "Theory of Error" with a theory of human action increases rather than reduces the importance of both internal and external causes, and emphasises the need to carry the analysis as far as possible.

To conclude, a theory of error must be a theory of the interaction between human performance variability and the situational constraints.

TAXONOMY

The taxonomy of the terms will obviously follow from the theory. Alternatively it may be considered a part of it. Since the theory is about human action rather than "Human Error", the taxonomy should be concerned with the situations where mismatches can be observed, rather than with the inferred "Human Errors".

There are several obvious dimensions for such a taxonomy. One already mentioned is whether the mismatch can be attributed to external or internal factors. In terms of the parts of an MMS, the question is whether the causes should be sought in the machine alone, in the operator alone, or in the interaction between the two. If the cause is assumed to lie with the operator, we have already seen how the analysis can be further refined using a decision making model.

Another possible dimension is whether the mismatch is detected by the operator, by the machine, or by an external agent e.g., a Technical Support Centre or a supervisor . In the first case one can further ask whether the operator tried to correct the mismatch, and how that influenced his activities.

Other dimensions can easily be found, and several complete taxonomies are available. One good example is the CSNI taxonomy (cf. Rasmussen et at., 1981), which is an attempt to characterise the situation where a mismatch occurs, rather than the "Human Errors". In this taxonomy "Human Error" is simply one of the many possible causes for a reported incident. Other taxonomies can rather easily be suggested once a proper theoretical background has been established. The choice of a taxonomy must depend on the purpose of the description, e.g., whether one wants to reduce the frequency of reported incidents, or improve the understanding of human decision making.

DEFINITION OF KEY TERMS

Before the key terms are defined, it is important to make sure that they are properly selected. One can, of course, make a potpourri of terms that are normally used to characterise situations where humans make mistakes or errors, and then define them, e.g., by using a recognised dictionary. But if the definitions are to serve a purpose, it is essential that they have a common basis, for instance a theory. By the same rationale it also is essential that the terms have a common basis.

To repeat what has been said above, I believe we should attempt to come forward with a theory for "Human Action" rather than "Human Error", and that this should be used for selecting and defining the key terms. Such a theory is not yet available, but I will nevertheless attempt to give a definition of some of the terms the organisers have listed, using intentional action as a basis.

- **Error**: Undefined. This term should be substituted by " action" or " activity".
- **Mistake**: Incorrect selection of goal state; incorrect goal decision.
- **Fault**: Incorrect selection of action to reach a goal, or incorrect execution of that action.
- **Slip**: Unintentional substitution of a correct performance segment (action) with

an incorrect one.

- **Accident**: External disturbance of intended performance.
- **Cause**: Accepted explanation for some performance characteristic, normally a performance mismatch.
- **Reason**: Subjective explanation of goal state or intention.
- **Origin**: Undefined. I am not sure why this is included in the list.
- **Responsibility**: Attribution of cause for the mismatch to a specific part of the MMS.

PREDICTION

Assuming that we try to establish a theory of human action rather than "Human Error", the predictions must be about actions. They must specifically be about the variability of human action that leads to mismatches. We can, of course, make a count of the instances where an operator makes a mistake, i.e., where the cause of the mismatch is attributed to a "Human Error". But that does not mean that it is sensible to attempt to assess the reliability of the operator, even if we refrain from considering the operator in mechanistic terms. Making such a count furthermore assumes that a meaningful measurement has been defined.

It is obvious for anyone who has worked with the reliability aspect of the human operator, that the occurrence and frequency of human errors depend more on the interaction with the environment than on any stable inherent characteristic of the operator. Similarly, quantitative measures, such as error rates, will therefore be inadequate and even misleading. Instead we need detailed and consistent descriptions of the conditions where mismatches occur. These qualitative descriptions may eventually be used as a basis for more straightforward measurements.

With regard to the specific questions relating to prediction, it will at our present state of knowledge only be the frequency of mismatches and typical causes that can be predicted. We know from experimental psychology, particularly the studies of attention and performance, that there are important regularities, as diurnal variations, situation dependencies, etc. Even though most of these data come from simplified laboratory situations, there is no reason to assume that they cannot be applied to realistic work situations. This has been confirmed, for instance, by studies of shift-work. It is also highly plausible that there are significant individual differences in "Error Proneness".

To summarise, making predictions requires an adequate definition of what the predictions are about. Unless frequencies and probabilities are sufficient, one must have a theory, or at least a set of good hypotheses, in order to make the predictions. It is furthermore logical that predictions cannot be about causes, unless we assume a strictly deterministic world. Consequently, the predictions must be about outcomes, i.e., observed mismatches, and possibly the actions leading to them. In the sense that "Human Errors" are causes, we can therefore not make predictions of human errors.

THERAPY

From a practical point of view the most important question is how mismatches can be prevented. One clue to this is found in the cases where mismatches do not occur, either because they are detected and corrected by the operator, or because the system is sufficiently forgiving. It would be reasonable to look further into these possibilities for preventing mismatches, hence reducing "Human Error"

There are probably very many ways in which an MMS can be designed to facilitate the detection and correction of errors. A good working theory of human action will be invaluable in this respect, since it will make it possible to indicate more precisely when and how interventions to change the course of action can be made. It is probably better to design for general detection and correction rather than for specific prevention. The experience from all types of process control clearly shows that Murphy's law cannot be beaten.

However, even if the best of systems has been designed, there will remain a basic variability in human performance that will lead to mismatches when the circumstances are right (or wrong, rather). If the operator was turned into an automaton (or even replaced by one), we might produce an error-free system, provided the degree of complexity was sufficiently low. But that is precisely the crux of the matter. The mismatches may occur not just because of mistakes made by the operator during operation, but also because of mistakes made by the designer during earlier phases. These mistakes would not be contained unless a theory of human action was applied to literally every aspect of the system.

SPECULATION

The questions raised in this group are very mixed Most of them seem to refer to fundamental problems of human beings, such as the evolution of learning and knowledge. I will save them for the, hopefully, warm nights at Bellagio. Some of them may have been answered indirectly by the considerations given in the preceding. From a cybernetic point of view there is definitely a virtue in error, seen as mismatches. It is only by becoming aware of, or being informed about, our failures to achieve the goals, including making clear what the goals are, that we can improve our performance. That certainly also includes the position I have exposed in this paper.

REFERENCES

Hollnagel, E., Pedersen, O. M. & Rasmussen, J. (1981). *Notes on human performance analysis (Risø-M-2285)*. Risø National Laboratory: Roskilde, Denmark.

Hollnagel, E. & Woods, D. D. (1983). Cognitive systems engineering. New wine in new bottles. *International Journal of Man-Machine Studies*, *18*, 583-600.

Rasmussen, J. (1981). Human Errors. *A taxonomy for describing human malfunctioning in industrial installations (Risø-M-2304)*. Risø National Laboratory: Roskilde, Denmark.

Rasmussen, J., Pedersen, O. M., Mancini, G., Carnino, A., Griffon, M. & Gagnolet, P. (1981). *Classification system for reporting events involving human malfunctions (Risø-M-2240), SINDOC(81) 14*. Risø National Laboratory: Roskilde, Denmark.

Taylor, F V. & Garvey, W. D. (1959). The limitations of a procrustean approach to the optimisation of man-machine systems. *Ergonomics*, *2*, 187-194.

Back