Position Paper for
NATO CONFERENCE ON HUMAN ERROR

August 1983, Bellagio, Italy

Erik Hollnagel
OECD Halden Reactor Project, Norway

1. INTRODUCTION

The organizers 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 behavior, 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' characterizes the outcome of an action rather than the cause. But 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 realize 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 organizers.

2. 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 generally 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 single 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 endogeneous and exogeneous 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 emphasizes 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.

3. 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 taconomy. 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 Center 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 completed taxonomies are available. One good example is the CSNI taxonomy (cf. Rasmussen et al., 1981), which is an attempt to characterize 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.

4. 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 characterize situations where humans make mistakes or errors, and then define them, e.g by using a recognized 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 organizers 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.

5. 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 lead 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' depends more on the interaction with the environment than on any stable inherent characteristic of the operator. 'Simple' 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, situational 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 summarize, 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.

6. THERAPY

From a practical point of view the most important questions 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 beause 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.

7. 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.

8. REFERENCES

Hollnagel, E., Pedersen, O. M. & Rasmussen, J. 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 "procustean" approach to the optimization of man-machine systems. Ergonomics, 2, 187-194.