THE EMPEROR'S NEW CLOTHES
Or
Whatever Happened To "Human Error"?

Erik Hollnagel
Dept. of Computer and Information Science
University of Linköping, SE-58333
Linköping, Sweden
Email: erik.hollnagel@ida.liu.se

Rene Amalberti
Imassa Cerma
BP 73, 91223 Bretigny sur Orge Cedex
France
Email: ramalberti@imassa.fr

"The road to wisdom?
Well, it's plain and simple to express:
Err
and err
and err again
but less
and less
and less."

(Piet Hein, *Grooks*, 1966)

1. INTRODUCTION

There has for many years been a protracted discussion about the nature and meaning of "human error". Since the term has been part of the daily language for thousands of years (cf. "errare humanum est …), many researchers have taken the term at face value and gone on to propose a considerable number of models, theories, and methods to deal with the "human error". There is considerable face validity to this approach since human action (and inaction) undoubtedly plays a major role in a large number of spectacular incidents and accidents – and in an even larger number of seemingly mundane events. Other researchers have for some time tried to argue that the term "human error" as such is ill defined, that the use of it is bound to be misleading, and that it

therefore better not be used at all. Without delving into tedious historical arguments, one of the early public debates about this issue was the NATO workshop held at Bellagio, Italy, in 1983 (reported in Senders & Moray, 1991), and one of the more eloquent presentations of the case against the use of the term is the book by Woods et al., (1994).

The theoretical arguments against the use of the term "human error" have not changed much over the years, since they from the very start pointed to the fact that the term was both ambiguous and alogical (Hollnagel, 1983). The theoretical arguments have little by little been supplemented by a growing realisation that the process of searching for "human error", and indeed the searching for any kind of root cause, is misguided since it corresponds to an oversimplified conception of how events occur. Most recently, a well-planned attempt to study "human error" empirically has made it clear how difficult it is to define a useful and usable classification system. Indeed, the conclusion of this study could very well be that attempts systematically to identify "human error" through observation in practice are doomed to failure.

In this paper we will try to go through the main arguments in this protracted debate, and try to show that researchers of "human error", like the ill-fated emperor in H. C. Andersen's fairy tale, have nothing on – metaphorically speaking, of course. Just as the emperor's new clothes only existed in the perception (or imagination) of people, so it is with the concept of "human error". Unlike the case of the emperor's new cloths, the reason for the false impression is not to be found in the skilful work of a couple of fraudsters, but rather in the seductiveness that some concepts or ideas seem to have. A more penetrating study of how this could happen is beyond the scope of this paper, and perhaps more a matter for students of the sociology of science. Instead we will concentrate on the arguments themselves, of which there are four.

2. THE ARGUMENT FROM SEMANTICS

In daily life, and in daily language, we use the term "human error" casually on the assumptions: (1) that everyone understands it and (2) that other people's understanding is the same as ours. In the communication between people it is sometimes a problem that a term may have a common denotation but different connotations. In the case of "human error" the situation is interestingly enough that the term has common connotations but different denotations. The fact that everyone does respond to the term because they do understand something by it, because it in a sense is intuitively meaningful, has created the

misconception that it is a well-defined term and that everyone understands it in the same way.

Anyone who has ever attempted to provide a technical definition of the concept of "error" will, however, have realised that it is quite difficult to do so. One reason is the lack of agreement regarding what constitutes the defining qualities of "human error", in many cases due to different premises or different points of departure. An engineer, for instance, may prefer to view the human operator as a system component which can succeed or fail in the same way as a piece of equipment. A psychologists, on the other hand, often assumes that human behaviour is essentially purposive and that it can only be fully understood with reference to subjective goals and intentions. Finally, sociologists have traditionally ascribed the main error modes to features of the prevailing socio-technical system where items such as management style and organisational structure are seen as the main mediating variables.

The fundamental semantic problem is that the term "human error" has at least three different denotations, so that it can mean either the **cause** of something, the **event** itself (the action), or the **outcome** of the action.

- **"Human error" as cause**: "The oil spill was caused by human error". Here the focus is on the action (the "human error") as the alleged cause of the observed outcome (the oil spill).
- **"Human error" as event or action:** "I forgot to check the water level". Here the focus is on the action or process itself, whereas the outcome or the consequence is not considered. In some cases the outcome may not yet have occurred but the person may still feel that an "error" has been made, such as having forgotten to do something. Nevertheless, a forgotten item or action need not always lead directly to a manifest failure.
- **"Human error" as consequence:** "I made the error of putting salt in the coffee". Here the focus is on the outcome, although the linguistic description is of the action. In this example, the fact that the coffee becomes undrinkable is the matter of concern and the action is therefore equated with the consequence. A more serious example is the use of the term "latent human error". This implies, wrongly, that one or more "human errors" are hidden somewhere in the system and that they have yet to manifest themselves. The intended meaning is rather that the system hides is one or more latent consequences of a "human error" that already has occurred.

The differences between the three denotations are illustrated in Figure 1. In each case there is a similar connotation, that "human error" is bad and something that should be avoided. The problem remains even if the distinction is simplified to be between "error as event" and "error as consequence". The multiplicity of usage has been criticised by Woods et al. (1994), who eloquently argued that "human error" always is a judgement in hindsight.

> "... 'human error' is not a well defined category of human performance. Attributing error to the actions of some person, team, or organization is fundamentally a social and psychological process and not an objective, technical one."
> (Woods et al., 1994, p. xvii)

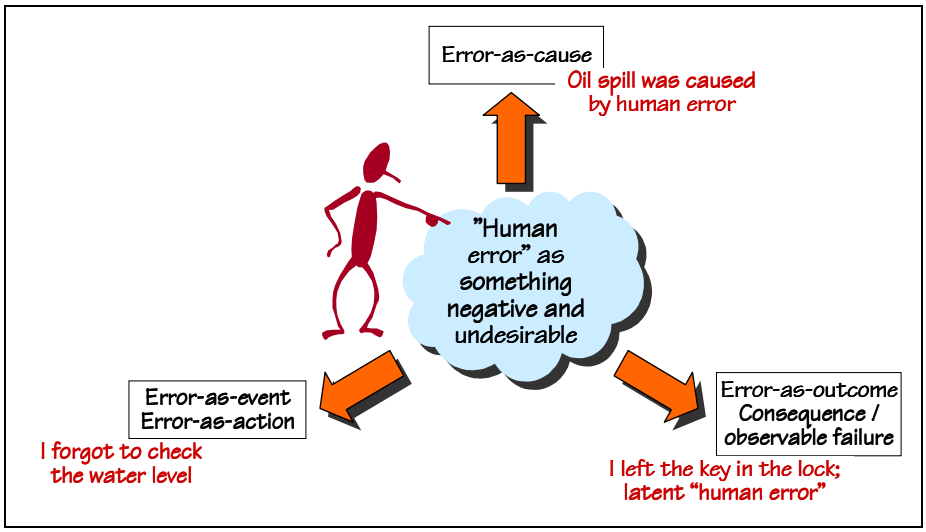

Figure 1: Three denotations of the term "error".

The multiple denotations are unfortunately not the only problem with "human error". Another is that it alludes to the notion of right and wrong or correct and incorrect, that is, a binary distinction. Yet even if we limit the use of the term "human error" to denote "error as event", the notion of an action gone wrong is a serious oversimplification. In practice, people may often realise, consciously or subconsciously, that something has gone awry before the consequences have had time to manifest themselves and therefore make attempts to compensate for or adjust the development of events (Sellen, 1994). Following the proposal of Amalberti (1996), this leads to the following classification, cf. Figure 2.

- Actions for which the actual outcome matches the intended outcome, i.e., actions that seem to achieve their goal. These actions are usually regarded as correctly performed actions, hence give little cause for concern, even though it is possible that the outcome came about in other ways.

- Actions that are perceived as having been carried out incorrectly in some way, but where the discrepancy is detected and corrected. This can either happen as the action is being carried out, where typing mistakes are a typical example, or immediately after as long as the system makes a recovery possible. If the system is sufficiently forgiving, the actual and intended outcomes may still match and the action may therefore for all intents and purposes be considered as correct.

- Actions that are recognised as being carried out incorrectly, and where recovery is not possible. A recovery can be impossible for several reasons, for instance that the system has entered an irreversible state, that there is insufficient time or resources, etc. In these cases the actual and intended outcomes do not match, and the action is therefore characterised as an error.

- Actions that are recognised as being carried out incorrectly, but where the discrepancy is ignored. This usually happens because the person considers the expected consequences of the action failure as unimportant in an absolute or relative sense. This assessment may either be correct or incorrect, depending among other things on the users' knowledge of the system in question. If it turns out that the consequences were not negligible, the action is in retrospect classified as an error.

- Actions that are carried out incorrectly, but which are not detected at the time, and therefore not recovered.

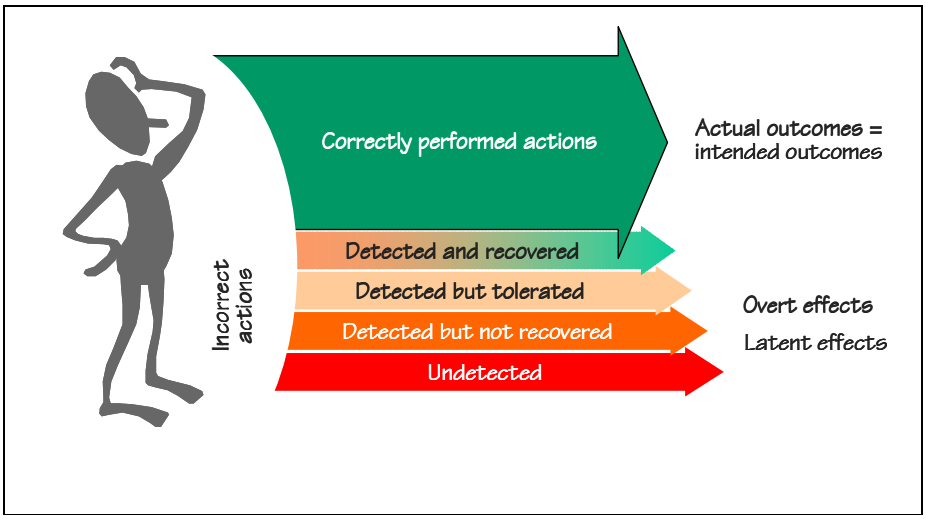

Figure 2: Classification of performance and outcomes.

A common element in the above descriptions is the detection or recognition that the outcome differs from what was expected. In cases where there is an observable outcome, or even some direct feedback from the system, this does not pose any problems. In cases where the recognition happens as the action is carried out, such as in typing or speaking, it raises interesting questions about how such discrepancies can be detected. One explanation is that the brain makes some kind of comparison between actual and intended outcomes on a neural level. Whatever the explanation may be, the fact remains that humans are quite good at detecting that something is wrong.

The existence of these five categories of action makes it clear, that the binary distinction between correct actions and "errors" is an oversimplification and therefore inappropriate. In fact, the whole argument so far leads to the conclusion that it is misleading to consider the specific action as a cause in itself. Furthermore, even if an action was carried out incorrectly, it is not necessarily a bad thing. Failures provide an important opportunity to learn, particularly if the outcome was a near miss or an incident rather than an accident.

3. THE ARGUMENT FROM PHILOSOPHY

The argument from philosophy relates to the metaphysical nature of causation. There is, of course, no reason to doubt the reality of causality, since almost

everything we do bear witness to that. It is enshrined in the laws of physics, at least outside the world of quantum effects, and if further proof is needed it is sufficient to consider the manifest success in building technological systems and, indeed, in being able to survive in a complex world in the first place. Yet even though it is possible to observe two events, call them A and B, and also to infer with more than reasonable certainty that one is the cause of the other, the determination of a causal relation is the result of reasoning rather than of observation. This was clearly pointed out by David Hume, who noted that the necessary conditions for establishing a causal relation between two events were priority in time, meaning that A should happen before B, and contiguity in space and time, meaning that A should be close to B in both respects. The conditions of priority and contiguity are necessary to conclude that a causal relation exists, but they are not sufficient. Indeed, it is generally acknowledged that causality cannot be attributed solely on the basis of a temporal relation.

In the case of "human error", the issue is even more complicated since it refers to the notion of backward causality, i.e., reasoning from effect to cause. In the simple case, we may observe that event A was followed by event B and conclude that B was the effect of A. In the more complex case that is the subject matter here, we observe event B, assume that is was an effect of something and then try to find out which event A was the cause of it. The problem of backward causation is aggravated by two common mistakes. The first is the human tendency to draw conclusions that are not logically valid. Thus, if we know that "If A, then B" is true, we are prone to assume that "B, therefore A" also is true (Wason & Johnson-Laird, 1972). In relation to backward causation this means that we fall into the trap of falsely associating a cause with an effect. This deficiency in the ability to reason in accordance with the rules of logic is exacerbated by tendency to rely on heuristics in reasoning, as described by e.g. Tversky & Kahneman (1984).

The second mistake is the failure to realise that the sequential relation between events to a considerable extent is an artefact of a description based on time. In the search for a cause, such as in accident analysis, it is common practice to represent how the events took place by means of a timeline. While it is certainly very sensible to do so, it should be realised that in such a description events will always follow each other. There will therefore be contiguity in time (and also in the graphical space of the representation) that is fortuitous but which nevertheless may affect the search for a cause. In the case of "human error" this is of some importance since one or more human actions always can be found. The artefactual contiguity in time combined with the tendency to make false logical conclusions therefore strongly predispose people to find

causes where there are none, and in particular to find "human errors" all over the place.

4. THE ARGUMENT FROM LOGIC

The argument from logic addresses the problem of the stop rule in searching for causes. As pointed out by many authors – too many to mention here – the stop rule is always relative rather than absolute. Even though accident investigations ostensibly aim to find the "root cause", the determination of a cause reflects the interests of the stakeholders as much as what actually happened. As Perrow (1986) noted:

> "Formal accident investigations usually start with an assumption that the operator must have failed, and if this attribution can be made, that is the end of serious inquiry. Finding that faulty designs were responsible would entail enormous shutdown and retrofitting costs; finding that management was responsible would threaten those in charge, but finding that operators were responsible preserves the system, with some soporific injunctions about better training" (p. 146).

Finding a cause is thus a case of expediency as much as of logic. There are always practical constraints that limit the search in terms of, e.g., material resources or time. Any analysis must stop at some time, and the criterion is in many cases set by interests that are quite remote from the accident investigation itself. As Woods et al. (1994) have pointed out, a cause is always a judgement made in hindsight and therefore benefits from the common malaise of *besserwissen*. More precisely, a cause – or rather, an acceptable "cause" – usually has the following characteristics:

- It can unequivocally be associated with a system structure or system function (people, components, procedures, etc.).

- It is possible to do something to reduce or eliminate the cause within accepted limits of cost and time. This follows partly from the first characteristic, which in a sense is a necessary condition for the second.

- It conforms to the current "norms" for explanations. This in particular means that the cause corresponds to the most popular theory at the time. For instance, before the 1960s it was uncommon to use "human error" as a cause, while it practically became *de rigueur* during the 1970s and

1980s. Later on, in the 1990s, the notion of organisational accidents became accepted, and the norm for explanations changed once more.

Even if the search for the cause is made as honestly as possible, it is necessary to stop at some point. In the case of hierarchical classification systems, such as the common "error taxonomies", the stop rule is given by the structure of the taxonomy. Not only that, but the analysis always begins in the same place and goes through a pre-determined number of steps, i.e., it has a fixed direction and a fixed depth. The situation is somewhat better for analysis methods that use a flexible or non-hierarchical classification scheme. In these cases the analysis begins at the most appropriate category, and the direction and depth of the analysis is determined by the context rather than by the number of categories. Yet the stop rule problem exists even here, since the search can only continue as long as there is sufficient information.

The logical problem in searching for causes exists because there can be no absolute limit to the depth of the analysis. Even though there in practice always will be some point where it makes little sense to go on, the stop rule is still subject to the accumulated knowledge and experience as it is encapsulated in the commonly accepted classification schemes. This can be illustrated by considering the development in the categories of causes over the last 50 years or so. As shown by Figure 3, the starting point was a distinction between technical failures, "human error" and "other" – the latter being the famous garbage can category for things we either do not know or do not care about. Over the years there has been a proliferation of categories in the "human error" and "other" groups, but less development in the group of technical failures. Figure 3 only illustrates the general principles, and neither the detailed contents nor the organisation of the categories should be taken as absolute. Indeed, it is certainly possible to list a far larger number of specific causes, since the imagination of analysts and psychologists seem to know few bounds. Relative to the present discussion, the development illustrated in Figure 3 shows that the determination of a cause is limited by the categories available to the analyst as well as by the uncertainty of the stop rule. This is not least the case for notion of a "human error", which has undergone several radical changes over the years.

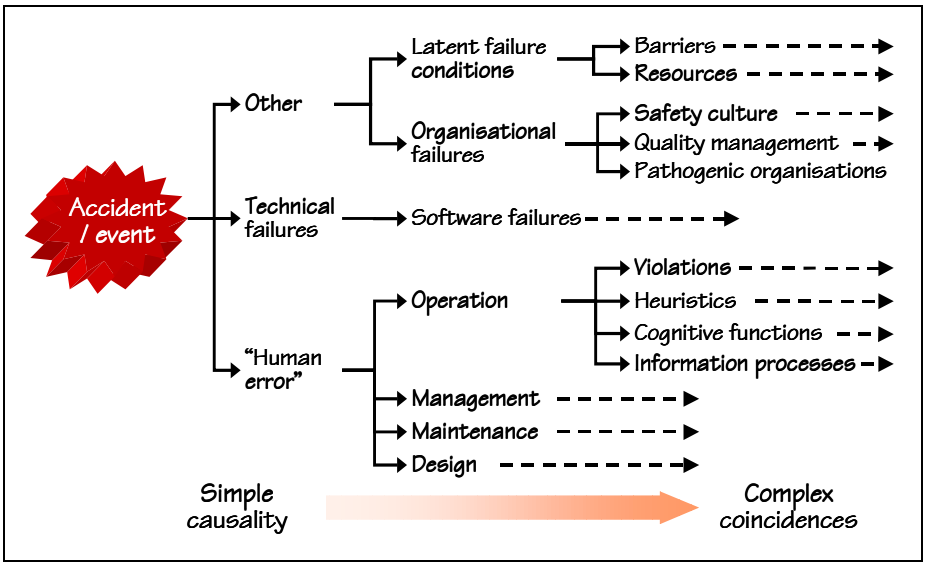

Figure 3: Development in categories of causes since 1950.

5. THE ARGUMENT FROM PRACTICE

The fourth argument comes from the practical problems in using a "human error" classification. This was recently demonstrated in a study undertaken as part of the second phase of the HERA (Human Error in Air Traffic Management) project, carried out by Dedale SA for Eurocontrol. The first phase of the HERA project had developed a model of human performance, together with the types of taxonomies needed to classify "errors" and contextual factors relating to ATM incidents, and the format that these taxonomies should take. There were two purposes of the above-mentioned study:

- To collect data on Air Traffic Controller (ATCO) "errors", to analyse these "errors" both by calculating statistics (primarily number of "errors" per hour), and to characterise and explain the types of "errors" made using the HERA model and taxonomy.

- To determine how confidently the retrospective technique developed in the first phase of HERA could be used for real-time observation.

The data collection took place on three days in September 2000, during an extended period of training with the free route system (FRAP) at the full-scale

simulator facility at Bretigny, France. There were three subjects who were all experienced ATCOs with 14, 31, and 23 years of experience respectively. On each of the three days, one controller was observed for a period of 60 minutes by two people, one being an experienced ATCO and the other a human factors specialist / psychologist. The observers were instructed to take note of all the "errors" they could see, using both a free text description and the list of External Error Modes (EEM) developed in the HERA I project. In addition, recordings were made of the ATCO screen as well as of the ATCO's performance, using a lateral camera and audio. Each observation session was followed by a 45-minute debriefing period, which also included auto-confrontation based on the recordings.

In relation to the present discussion, one interesting issue was the consistency in using the list of EEMs from HERA I. As shown by Table 1, a considerable number of "errors" were observed during each period. For all three sessions there were substantial differences in the number of "errors" noted by the two observers, and more importantly only there was only a very small number of "errors" were noted by both. This finding is striking because it means that two people observing the same performance, and ostensibly using the same classification terms, nevertheless can come to very different results.

Table 1: Number of observations and categorizations for the three sessions.

	Session 1		**Session 2**		**Session 3**	
	N	**%**	**N**	**%**	**N**	**%**
Number of observed actions	50		35		33	
Number of "errors" observed by ATCO	30	60.0	16	45.7	8	24.2
Number of "errors" observed by psychologist	11	22.0	9	25.7	12	36.4
Total number of observed "errors"	35	70.0	24	68.6	20	60.6
Number of "errors" in common between observers	6		1		0	
Number of actions clarified by interview	16	32.0	8	22.9	5	15.2
Number of "errors" clarified by interview	12	34.3	6	25.0	2	10.0

A closer look at the "error types" used by the two observers shows that the ATCO and the psychologist tended to use different subsets of the list that was provided. The ATCO mainly used categories that related to the external working conditions, such as HMI, resource management, and strategy. The psychologist tended to use categories that related to the subject's psychological state, such as slips and attention. In addition, both observers used the

categories of communication and situation awareness. These differences, as well as differences in the number of "error types" that each observer found, shows that the background and experience has a strong effect both on what was observed and on how the observations were classified.

Table 2: Assigned "error types" used by the two observers.

ATCO	S 1	S 2	S 3	Psychologist	S 1	S 2	S 3
Communication	4	5	0	Communication	2	2	3
HMI	9	1	2	Slips	2	3	4
Resource management	7	2	0	Attention	1	1	1
Situation awareness	4	7	2	Situation awareness	3	1	1
Strategy	6	1	0				

So far the conclusions would seem to be that there is a considerable variability in how a classification scheme is used and that the observers' background and experience may be at least as important as the categories themselves. This outcome would have been of interest even if the study had been limited to making a set of observations and categorising them. But as mentioned above, the study also included a debriefing period, which allowed the observers to go through their observations with the subject. In the cases where observers and subjects had a different view of what had happened, the debriefing gave the subjects an opportunity to explain the reasons for their actions. This provided a clarification which in many cases meant that actions, which had been classified as "errors", were reclassified as "normal" actions. The extent of this reclassification is shown in the last two rows of Table 1. A few examples of the clarification provided by the debriefing are given below.

- Early transfer of planes: The anticipated transfer of planes corresponds to a strategy of prevention: this makes it possible quickly to get rid of planes, which then no longer require surveillance (effectively a kind of resource management).

- Late transfer of planes: When a controller detects a conflict a little late, he/she may want to solve the conflict before transferring the plane to the adjoining sector. In addition, when the workload is heavy (e.g. during the resolution of a conflict), other planes may be momentarily "forgotten". These situations are typical of working "near the limits".

- Inaction: In some cases where a Short Term Conflict Avoidance (STCA) was announced by the system, controllers would decide to do nothing because "these alarms are not always pertinent", i.e. the ATCOs would let time solve the problem.

- Attention failures: The ATCOs justified many "errors" by invoking habits acquired in relation to real work, poor knowledge of the HMI (one week of training) and the "lack of logic" of certain procedures or configurations.

The issue here is, of course, not so much whether the subjects acted correctly but rather that the observers often would classify actions wrongly because they could not see the situation from the subject's point of view. The lesson to be learned is that an action should not be classified as an "error" only based on how it appears to an observer.

5.1 Nominal And Actual Performance Deviations

The outcome of the study may seem confusing, but on second thought the problems are mainly due to the assumption that actions – whether correct or incorrect – can be unambiguously classified by a context-free set of categories. According to this view, the reclassification was an unexpected and, in principle, unwanted adjustment of the observations. It is, however, possible to adopt a completely different perspective, according to which the classification-reclassification process is the norm rather than the exception. The starting point is that when the behaviour of a subject is observed, most of the actions will be recognised as meaningful while a few may be seen as meaningless. (The problems of how a meaningful action can be defined and recognised will not be discussed here.) Of the set of recognised meaningful actions, some will be seen as nominally correct and some as nominally incorrect. The latter we shall call "nominal performance deviations". They are nominal in the sense that although they correspond to "error types" that are defined by the classification scheme, it is acknowledged that the match with an "error type" is not necessary a final one. Put differently, the actions are "errors" relative to the observer's point of view, but may not necessarily be so in a different context. A process of clarification, which effectively enriches the context description, must therefore follow the initial classification. The enriched context must include how the subjects understood the situation and how they reasoned about what they had to do. As a result of this clarification, some of the actions from the group of "nominal performance deviations" may be reclassified as "performance adjustments", while the remaining are classified as "actual

performance deviations". The principles of the classification-reclassification procedure are shown in Figure 4.

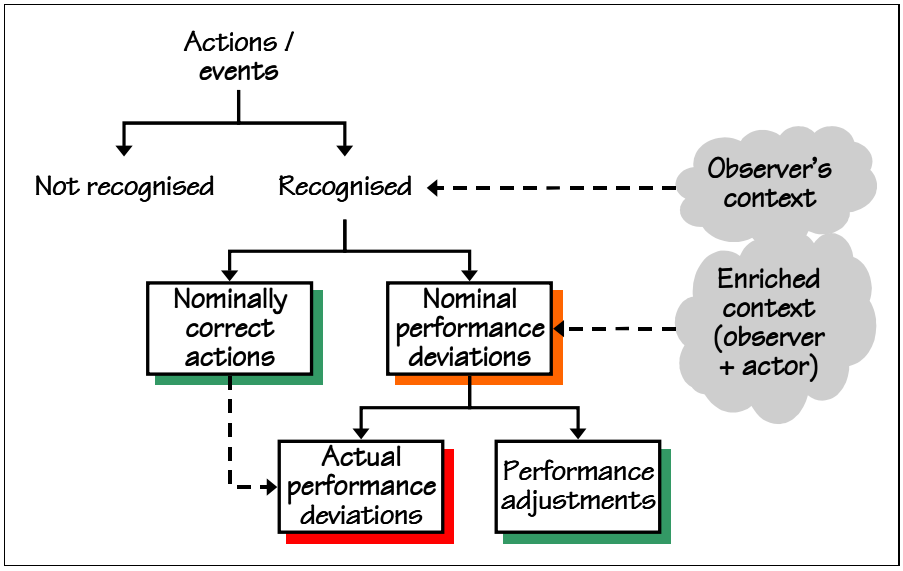

Figure 4: The classification-reclassification procedure.

5.2 A Note On Terminology

The description above has introduced a number of new terms in order to avoid the problems in talking about "errors". While most of the terms speak for themselves, the notion of "performance adjustments" probably requires a few comments. It is a widely acknowledged fact that people, as a rule, never carry out a task exactly as taught or as prescribed (to the ever lasting dismay of instructors and procedure writers). What they do instead is to adjust (rather than adapt) their actions to meet the perceived demands and constraints of the work situation and to avoid problems in the short or medium term. In short, performance is adjusted – to the best of the acting person's ability – in order to remain in control of the situation. Such performance adjustments are furthermore part of the natural and irreducible variability of human performance and therefore represent normal performance rather than deviations.

The description of the reclassification focused on how actions might change from being "nominal performance deviations" to become "performance

adjustments". As Figure 4 suggests, it is also possible that "nominally correct actions" come to be seen as "actual performance deviations". In practice this means that even if there are good reasons for doing something, it may still lead to unwanted consequences, hence in retrospect be seen as an "error". The terminology proposed here thus reinforces the view that an "error" is a judgment in hindsight. At the time of the action it can, in principle, not be known whether the action will succeed or fail. Since furthermore the classification of an action, as being either nominally correct, a nominal deviation, an adjustment, etc., is a relative (social) judgement rather than an absolute distinction, it means that unwanted consequences may follow from any kind of action. When something goes wrong we can, with hindsight, begin to look for the "error" or the action that was a likely main contributor to the outcome, but it may turn out to be a "nominal performance deviation" as often as a "nominally correct action".

To conclude, an important lesson from this study is that it is very difficult to classify actions without a context. (This obviously goes for "correct" actions as well as "errors", and for human observers as well as for automated classification systems.) Or rather, that a classification always implies a context, but that the context of one observer may be quite different from that of another, and different again from that of the person who is acting. It is furthermore impossible to define an absolute or reference context relative to which actions can unequivocally be classified as right or wrong. Since the context implied by most "error" taxonomies is very sparse and highly speculative, it follows that it is both principally and practically impossible to use such taxonomies to classify "errors".

6. DISCUSSION

In this paper we have summarised four main arguments against the use of the term "human error" with the intention of showing that it creates more problems than it solves.

- The semantic argument pointed out that there is a fundamental ambiguity in the use of the term "human error", specifically that it fails to distinguish between "error as action" and "error as cause".

- The philosophical argument referred to the metaphysical status of causation, and to the problems in backward causation. To these were added the problems humans have in reasoning according to the rules of

logic, and the fact that a sequential relation may be an artefact of a temporal description.

- The logical argument discussed the arbitrariness of the stop rule in searching for causes, as well as the fact that accident analysis is a matter of expediency as much as of logic.

- Finally, the empirical argument illustrated the problems in making observations of "human error", due to the fact that it is necessary to know the context in order to make a reasonably correct classification.

The conclusion is that the term "human error" should be used carefully and sparingly – if it is to be used at all. In the long term it may be prudent to refrain from considering actions as being either correct or incorrect, firstly because these distinctions rarely apply to the action in itself but rather to the outcome, and secondly because they imply a differentiation that is hard to make in practice. The alternative is to acknowledge that human performance (as well as the performance of technological systems) is always variable.

The consequence of acknowledging the existence of this variability is that many so-called "human errors" can be seen as the outcome of successful performance adjustments, which include ways of saving attention, managing workload, making decisions based on heuristics (in the sense of naturalistic decision making), etc. As long as these adjustments meet the socio-technical expectations to acceptable results, they are seen as being goal-oriented, effective, and reflecting the intelligence of human beings. Moreover, "errors" or "poor decision-making" resulting from such intentionally sub-optimal actions are often detected and recovered in time. Because these adjustments usually are successful they become the norm, and are therefore also used when the conditions – in retrospect – are unfavourable. It is thus only when the detection and/or recovery for some reason fails, that they become "human errors".

The re-evaluation of "human error" furthermore does not stop at the individual level. The output of one operator is often the input to another within the total sequence of work. This means that non-recovered sub-optimal actions may propagate through the system. The team and systemic levels learn to adjust their own functioning to absorb this propagated sub-optimality, which thereby becomes the rule of normality throughout the system. One consequence of this is that when accidents occur they should not be seen as the logical consequence of multiple "local errors" or performance adjustments,

as described above, since such adjustments are part of the ecology of normal work. Any attempt to apply backward causation to find the origin of an accident in the numerous preceding performance adjustments corresponds to a naive view of safety. In practice it may lead to a counter-productive set of responses, such as chasing "human errors" with no relation to safety, and thereby hinder normal adaptive system performance. The accident should rather be seen as a singularity, as the concomitant breakdown of control on one or more levels within the system, which occurs despite of, rather than because of, the many instances of non-optimal behaviour.

The story of accidents and loss of control is basically the story of local sub-optimal compromises, with little or no relation with the past "errors" of the system. This approach clearly goes against the simplistic view of the sequential accident models and diminishes the importance of chasing every little "human error" as a way to enhance safety. Indeed, the only heritage that is relevant for the prediction of future accidents is the history of how individuals and teams learn to make compromises. We know that, depending on the successes achieved in the past, these compromises indicate how fast the system moves simultaneously towards better performance and greater risks. The management of this learning process is consequently one of the keys for keeping safety under control (Amalberti, 2001).

The inherent variability may at times become so large that it leads to unexpected and unwanted consequences, which then are called "errors". Yet regardless of what the outcome is, the basis for the performance variability is the same, and classifying one case as "error" and the other as not makes little sense. Instead of trying to look for "human errors" as either causes or events, we should try to find where performance may vary, how it may vary, and how the variations may be detected and – eventually – controlled (Hollnagel, 2000).

7. ACKNOWLEDGEMENT

The authors would like to thank the Dedale team for providing access to data from the study mentioned in the text, as well as to the people involved in both HERA I and HERA II for many stimulating discussions.

8. REFERENCES

Amalberti, R. (1996). *La conduite des systèmes à risques*, Paris: PUF.

Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety Science*, 37, 109-126.

Hollnagel, E. (1983). *Position paper on human error*. Responses to Queries from the Program Committee for the NATO Conference on Human Error, Bellagio, Italy, September 5-9.
(URL http://www.ida.liu.se/~eriho/EH_HumanError.htm)

Hollnagel, E. *Performance variability management*. Symposia at Shrivenham: People in digitized command and control. 12-14 December, 2000.

Perrow, C. (1984). *Normal accidents: Living with High-Risk Technologies*. New York: Basic Books.

Sellen, A. J. (1994). Detection of everyday errors. *Applied Psychology: An International Review*, *43*(4), 475-498.

Senders, J. W. & Moray, N. P. (1991). *Human error. Cause, prediction, and reduction*. Hillsdale, NJ.: Lawrence Erlbaum.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Wason, P. C. & Johnson-Laird, P. N. (1972). *Psychology of reasoning*. London: B. T. Batsford.

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