Proceedings of ICONE12:
12th International Conference on Nuclear Engineering
April 25-29, 2004, Virginia, USA

49287

THE PROBABILITY AND MANAGEMENT OF HUMAN ERROR

Romney B. Duffey
Atomic Energy of Canada Limited, Chalk River Laboratories, Chalk River, ON K0J 1J0
Tel: 613-584-8811 Ext. 6272 Fax: 613-584-8246
E-mail: duffeyr@aecl.ca

John W. Saull
International Federation of Airworthiness
W. Sussex, RH19 3RF, United Kingdom
E-mail: John@ifairworthy.fsnet.co.uk

ABSTRACT

Embedded within modern technological systems, human error is the largest, and indeed dominant contributor to accident cause. The consequences dominate the risk profiles for nuclear power and for many other technologies. We need to quantify the probability of human error for the system as an integral contribution within the overall system failure, as it is generally not separable or predictable for *actual* events. We also need to provide a means to manage and effectively reduce the failure (error) rate.

The fact that humans learn from their mistakes allows a new determination of the dynamic probability and human failure (error) rate in technological systems. The result is consistent with and derived from the available world data for modern technological systems. Comparisons are made to actual data from large technological systems and recent catastrophes. Best estimate values and relationships can be derived for both the human error rate, and for the probability.

We describe the potential for new approaches to the management of human error and safety indicators, based on the principles of error state exclusion and of the systematic effect of learning**.**

INTRODUCTION

Fortunately or unfortunately, “human error” has an overriding (typically >60%) contribution to almost all events. This is true worldwide for the whole spectrum, all the way from transportation crashes, social system and medical errors, to large administrative failures and the whole gamut of industrial accidents. Such event rates generally follow the Universal Learning Curve (ULC) [1]), where a learning pattern is clearly evident, and well over 1000 data sources formed the basis for the estimated value taken for the learning rate “constant”.

The human error probability is indeed *dynamic*, and evolves with experience. It is not a constant as taken in may risk and probabilistic safety analyses. In this difficult arena of coupled human and technological system behavior, these new results show a very reasonable level of concordance between the MERE probability and the human error data, using the typical minimum error interval (100,000 to 200,000 hours). The method and approach we have validated here also allows for predictions to be made.

As errors occur, Managers, Society and Regulators introduce new measures (Rules, Procedures and Systems) to attempt to manage (eliminate) their re-occurrence or reduce their probability. Investigators evaluate in detail the “event sequences” to determine the apparent causes. New Measures are taken to try to control the errors caused by the role of humans themselves. New Requirements are developed that may require different safety measures. New Process changes are implemented in design manufacturing or operation, or using other management means (such as safety programs, retraining and/or the reassignment of staff). Fines may be levied as a means of punishment for allowing something to even happen.

The Federal Aviation Agency (FAA) has noted, “We need to change one of the biggest historical characteristics of safety improvements - our reactive nature. We must get in front of accidents, anticipate them and use hard data to detect problems and the disturbing trends”. This is a very clear message to us all, and to be proactive requires a new theory and understanding of the stochastic nature of error causation and prevention. Our fundamental work on the analysis of recorded events in major technological industries has been published in the book “Know the Risk”.

The adoption of a corporate-wide dictated “safety culture” or a “safety management system” are often seen as effective management responses. The inclusion of these Rules along with so-called “performance indicators” is an example of learning, which should change the observed distribution of outcomes with experience, and hence explains the effects of safety management or intervention in the system.

1. The MERE Solution for the Human Error Rate

1.1 Working Definitions and the learning Hypothesis

We cannot easily separate the contribution of human error to system failures. Instead, we provide a means to analyse the average emergent behavior, not pretending that we can describe or model the human-machine interaction in all its complexity. So we may model the human as introducing a failure (error) rate as an integral part of and occurring within any technological system.

To proceed, we may adapt conventional reliability terminology to all systems that include humans somewhere in their design, operation, maintenance, and performance [2]. We define an *error* as the unintended action or the combination or confluence of actions, equipment and circumstances that causes an observed or measurable *outcome* (an accident, event or injury) in or by a human-technological system.

The *failure rate* is the measure of the rate of outcomes for the human-technological system, and the failure (error) *probability* is the chance or likelihood of the observed outcome actually happening.

We simply observe the outcomes of the system including whatever human behavior occurs; and as we Humans make mistakes, we expect and try to make the rate of errors decrease as and if we learn.

To determine the failure probability for any system using reliability analysis, it is conventional to define a so-called hazard function, h(t), varying with some elapsed time [2,3]. Human errors and reliability are coupled to and embedded within all technological systems, and although different psychologically, there is no difference mathematically. We cannot prescribe all the workings of the human mind. So, we use a general *emergent theory*, where the failures are invisible inside the entire human-technological system until we observe the errors as outcomes, and record and try to correct for them. The basic and sole assumption that we make is the "learning hypothesis" as a physical model for human behavior when coupled to any system. Simply and directly, we postulate that humans learn from their mistakes (outcomes) as experience is gained. So, the rate of reduction of outcomes (observed in the technology or activity as accidents, errors and events) is proportional to the number of outcomes that are occurring (Duffey and Saull[1]).

1.2 The Probability of Human Error

Conventionally, reliability engineering is associated with the failure rate of components, or systems, or mechanisms [2], not human beings in and interacting with a technological system We invoke the simplest idea, namely the learning hypothesis that is a function of experience.

Thus, the human error or failure rate, λ, within the system, is equivalent to a dynamic hazard function $h(\varepsilon)$ which varies with experience, ε, as given by the Minimum Error Rate Equation (MERE) [1],

$$d\lambda/d\varepsilon = - k \ (\lambda - \lambda m) \qquad (1)$$

where k is the learning rate constant, and λ_m the minimum obtainable rate. The failure rate, $\lambda(\varepsilon)$, is then obtained by straightforward integration as,

$$\lambda(\varepsilon) = \lambda_m + (\lambda_0 - \lambda_m) e^{-k\varepsilon} \qquad (2)$$

where the failure rate $\lambda \equiv h(\varepsilon)$; λ_m is the minimum obtainable rate at large experience; and λ_0 is the initial rate at some initial experience, ε_0.

We have found the value of the learning rate constant, $k \approx 3$, as determined from all the available world data [1]. The value for $1/ \lambda_m$ from observation and data is in the range, one in 100,000 to 200,000 hours, which corresponds to the maximum time between *human* failures when coupled to, with or in a technological system (so that MaxMTTF $\equiv (1/ \lambda_m) \sim 2.10^5$ hours of experience).

The probability of the outcome or error occurring in or taking less than ε, is just the cumulative distribution function, CDF, conventionally written as $F(\varepsilon)$ so [2,3]:

$$p(\varepsilon) \equiv F(\varepsilon) = 1- e^{-\int \lambda d\varepsilon} \qquad (3)$$

Hence, the probability is a double exponential due to the exponential form of the failure rate Equation (2) itself imposed on Equation (3). This form is related to or may be considered as an "extreme value distribution" function that has arisen quite naturally from the learning hypothesis.

Substituting the MERE human failure rate in Equation (2) into Equation (3) and carrying out the integration from an initial experience, $\tau=\varepsilon_0$, to any interval, ε, we obtain the probability as the double exponential:

$$p(\varepsilon) = 1 - \exp \{(\lambda- \lambda m)/k - \lambda(\varepsilon_0 - \varepsilon)\} \qquad (4)$$

where, from Equation (2),

$$\lambda(\varepsilon) = \lambda_m + (\lambda_0 - \lambda_m) \exp - k(\varepsilon-\varepsilon_0) \qquad (5)$$

and $\lambda (\varepsilon_0) = \lambda_0$ at the initial experience, ε_0, accumulated up to or at the initial outcome(s).

It is clear on either a log –log or linear –linear plot, the rates for a given experience evolve as a double exponential. We illustrate the failure probability, $p(\varepsilon)$ in Figure (1), using the standard learning rate constant value, k=3, with an illustrative range of the minimum or asymptotic error rates, λ_m. The initial failure rates taken for λ_0 are completely arbitrary (1 and 10), so that the curves can be scaled to or for any initial value.

It is evident that the double exponential form for the human error component plays a key role. We learn as we gain experience, and then reach a region of essentially no decrease, in rate or in probability, and hence in likelihood. It is easy to obtain the first decrease in rates or probabilities but harder to proceed any lower. This is exactly what is observed in transport, manufacturing, medical, industrial and other accident, death and injury data [1].

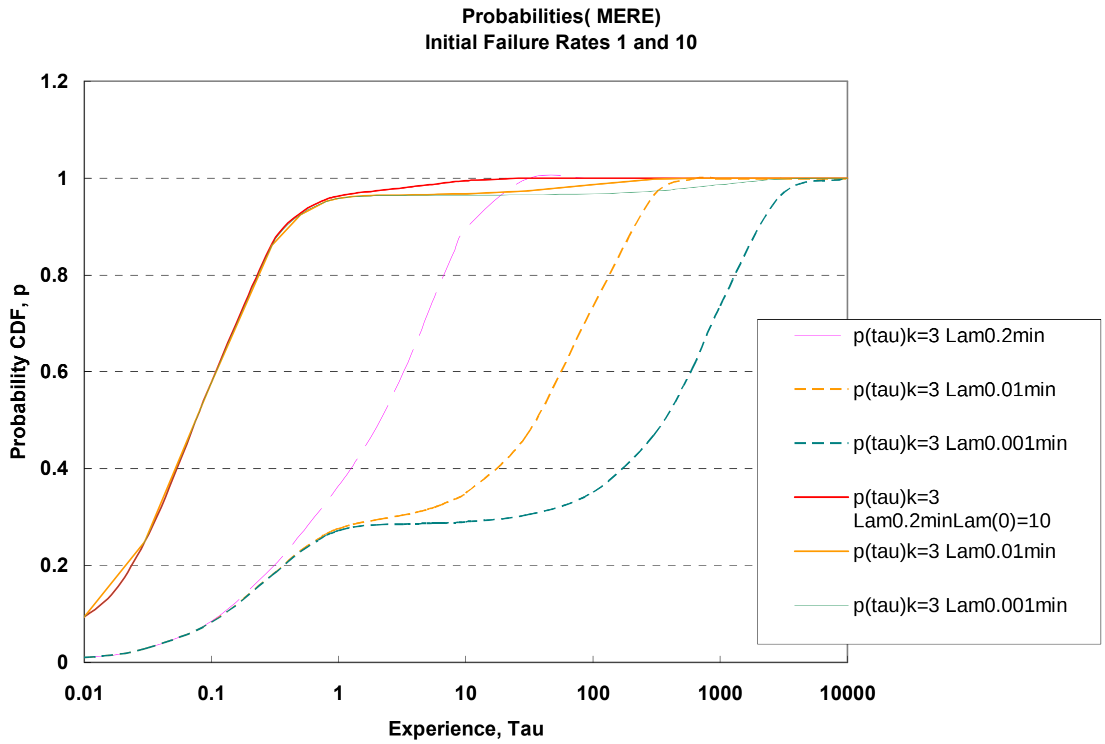

Figure 1: The MERE Probability function, p=h(ε), for arbitrary initial rates in Equation (5)

2. THE INITIAL FAILURE RATE AND ITS VARIATION WITH EXPERIENCE

Having established the learning trend, we need to determine the actual parameters and values using data and insight. Now, in reality, the initial rate, λ_0, is *not* a constant as assumed so far since the outcomes are stochastic in experience "state space". Hence, $\lambda_0 = \lambda(\varepsilon_0)$, and it is not known when exactly in our experience we may have an error initially observed (and we might be lucky or not), and the initial value we ascribe to the initial rate observe is an arbitrary value.

We could take the range $0.000005 < \lambda_0 < 0.00005$, but only as bounding estimates. To establish the initial rate, key data are available from commercial aircraft outcomes (fatal crashes) throughout the world. The major contributor is human error not equipment failure, although these latter can also be ascribed to the root cause of human failings. Fatal crashes and accidents for the thirty years between 1970 and 2000 are known (e.g. from [4]), for 114 major airlines with ~ 725 million hours (Mh) of total flying experience.

The major influence of when and whether one occurs is actually human errors *within the system* coinciding or surfacing with throughout the technological system (in design, manufacture, operation, maintenance, regulation or flying). In addition to airlines with both large and small experience, and to cover *six orders of magnitude* of experience, we also examined the data points for two other key events, the crash of the supersonic *Concorde* with a rate, λ_0, of one in about 90,000 flights; and the explosion and disintegration of the space shuttles, *Challenger* and *Columbia,* with a rate, λ_0, of two out of 113 total missions [5].

For all these data and experience, there is a remarkable constancy of risk, as given by the straight line of slope –1, which is given by the equation:

$$\lambda\,\varepsilon = \text{constant, n} \qquad (6)$$

The observed rate is strictly a function of whatever experience it happened to occur at, any value being possible. Thus, in the limit for rare events, the initial rate should be the *purely Bayesian estimate* from the prior experience at whenever the initial outcome occurred, which gives, with n=1 for the initial outcome in Equation (5):

$$\lambda_0 \approx (1/\varepsilon) \qquad (7)$$

This result can be inserted into Equation (5), and is both the simplest and a statement of common sense, in that the initial rate is whatever it may be at the beginning.

What the data are telling us is that the limiting initial rate is exactly what it is for the experience at which the first outcome occurs, no more and no less.

From the analysis of many millions of data points that include human error in the outcomes, we have been able to derive the key quantities that dominate human error in current technological systems. These now include commercial air, road, ship and rail transport accidents; near misses and events; chemical, nuclear and industrial injuries; mining injuries and manufacturing defects; general aviation events; medical

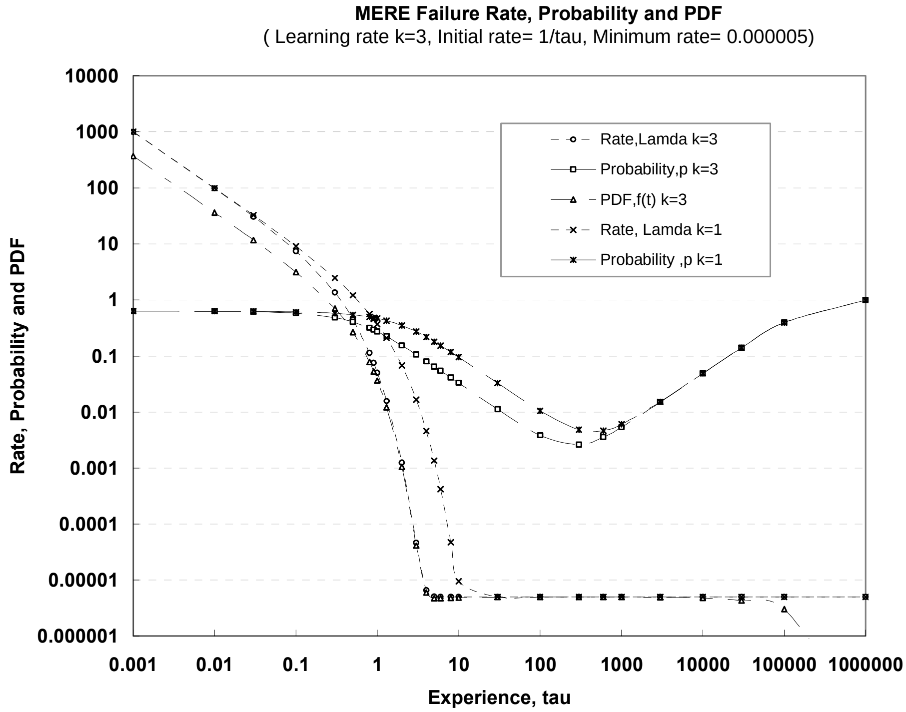

Figure 2: The best MERE values (from Equation (9))

misadministration and misdiagnoses; pressure vessel and piping component failures; and office paperwork and quality management systems [1].

From all these data, and many more, we have estimated the minimum failure rate or error interval, the typical initial error interval, and the learning rate constant as follows as our "best" available estimate [1]:

$$\lambda(\varepsilon) = \lambda_m + (\lambda_0 - \lambda_m) e^{-k\varepsilon}, \qquad (8)$$

which becomes, for $\lambda_0 = (n/\varepsilon)$, with n = 1 for the initial outcome,

$$\lambda = 5.10^{-5} + (1/\varepsilon - 5.10^{-5}) e^{-3\varepsilon} \qquad (9)$$

The human failure (error) rate, λ, can be evaluated numerically, as well as the probability, $p(\varepsilon)$, and the PDF, $f(\varepsilon)$, as shown in Figure (2), where $\varepsilon \equiv \tau$ units in order to represent the accumulated experience scale.

It is evident that for k>0 the human error probability is a classic "bathtub" shape, being just under near unity at the inexperienced start (Figure 2), and then falling with the lowering of error rates with increasing learning with experience. After falling to a low of about one in a hundred "chance" due to learning, it inevitably rises when the experience is $\varepsilon > 1000$ tau units, and becomes a near certainty again by a million tau units of experience as failures re-accumulate, since the minimum rate exists of $\lambda_m \sim 5.10^{-5}$ per experience tau unit. The importance of learning is evident, since for k<0 forgetting causes a rapid increase to unity probability with no minimum.

The future estimate is once again derivable from its unchanged prior value, and thus prior knowledge from *the past predicts the future* [6].

3. THE MANAGEMENT OF FAILURE AND HUMAN ERROR

For the human error case, we have the results of the probability of non-detection (i.e. human error) of latent faults for nuclear plant transients [7], which are also fairly regular event. The data and the MERE error rate were normalized to fitting the curves using a maximum experience of ~100,000 h (1 tau ≡ 1 transient hour), close to the MERE minimum error interval. The initial probability was taken as unity, that is $p(\varepsilon_0) = 1$, for comparison purposes. The data and the theory are in reasonable accord, despite the necessity of having to be renormalized.

Error states exist as part of all technologies, and the challenge is therefore to eliminate them, not just partially correct them. It is evident that they are generally as a result of differing combinations of circumstances. The concept of error states we describe is new, and is derived from the observed characteristics and statistics of actual accidents, errors and events in all of which humans play a major part.

Changing the interpretation and use of large data sets (e.g. as derived from Flight Operations QA, accidents statistics, Risk Management reporting, Performance Indicators, and Significant Event Reports and the like), from being reactive of historical in nature (tracking the known), to being proactive in making predictions (about the unknown) for yet unobserved Error States is, therefore, the major challenge. Just recording the outcome data is not enough.

Learning forms the important aspect of corrective measures and we have shown the importance of increasing experience. However, Forgetting is also evident and must be addressed aggressively if we are to cut the accident and serious incidents worldwide. Human error being the major causal factor of the outcomes we observe (as Error States), characterized by the number or rates of accidents, events, lapses and incidents.

From the results, we infer that error rate reduction (learning) is proportional to the rate of errors being made, which is derived from the total observed number of stochastic distributions of errors. Both Learning and Forgetting are also naturally included via the experience shaping function and/or learning rate constant. The microscopic distributions of error states now appear manifested as the observed macroscopic outcomes that form the systematic trends evident in the ULC.

The implications are very profound. The error distribution with experience is such that the errors decrease to a minimum to define a learning curve. Learning cannot eliminate the lowest or zeroth error rate: there are no “zero defects”, simply a trend towards the minimum state. The zeroth microstate now represents the “built in” contribution from human error.

Can we eliminate errors? After all we observe outcomes, and the traditional approach is to investigate prior events, identify root causes, allocate blame, and devise ways to stop them “ever happening again”. Now, in reality, we cannot distinguish among the many error states from observation as they have random behaviour: they simply follow a learning curve. One error state distribution is not identical to another, and never will be statistically because the distribution of states is always different. This simple fact explains why (different) errors still occur while we reactively try to eliminate known or previously observed causes (or microstates).

The inclusion of rules and systems is an example of learning, which should change the observed distribution of outcomes with experience (the learning rate), and hence explains the effects of safety management or intervention in the system.

We note that imposing such Rules arbitrarily can clearly be a piecemeal and totally reactive approach, which is trying to eliminate the observed outcomes microstate-by-microstate, case-by-case, and accident-by-accident. The Rules should be optimized so as to address the multitude of other possible distributions and Observations made to determine their effect on the distribution.

To put it clearly, management, laws, inquiries, regulations and safety systems try to impose order on disorder. Hence, they actually try to achieve the impossible: to reduce the “entropy” of the system [6]. Implicitly, they try to reach a minimum state of disorder, or a maximum state of order, as effectively as possible using Goals and incentives (both rewards and punishments). But changing the distribution is not the same as the desired error state elimination.

CONCLUSIONS

Analysis of failure rates due to human error and the rate of learning allow a new determination of the dynamic human error rate and probability in technological systems, consistent with and derived from the available world data. The basis for the analysis is the “learning hypothesis” that humans learn from experience, and consequently the accumulated experience defines the failure rate. The exponential failure rate solution of the MERE defines the probability of human error as a function of experience.

A new “best” equation is given for the human error or failure rate, λ, which allows for calculation and prediction of the dynamic probability of human error in technological systems. The expression combines the influences of early inexperience, learning from experience, and stochastic occurrences with having a finite minimum rate and is:

$$\lambda = 5.10^{-5} + \{(1/\varepsilon) - 5.10^{-5}\}\, e^{-3\varepsilon} \qquad (9)$$

In reality, after the initial failure rate given by the purely Bayesian estimate, the subsequent failure rate is of a double exponential form, where the probability naturally follows a classic bathtub curve. The future failure rate is entirely determined by the experience: thus the past defines the future.

The results demonstrate that the human error probability is dynamic, and that it may be predicted using the MERE learning hypothesis and the minimum failure rate.

The implications of this error state elimination using safety management systems show we must adopt new goals and event tracking systems.

REFERENCES

[1] Duffey, R.B., and Saull, J.W., 2002, “Know the Risk”, First Edition, Butterworth and Heinemann, Boston, USA.

[2] Lewis, E.E., 1994, “Introduction to Reliability Engineering”, Second Edition, JohnWiley and Sons, New York, USA, pp.61, 112, and 141.

[3] Tobias, P.A., and Trindade, D.C., 1986, “Applied Reliability”, First Edition,Van Nostrand Reinhold, New York, USA, pp.64, 70, and 117.

[4] Airsafe, 2000, http://www.airsafe.com, “Fatal Events and Fatal Event Rates from 1970-2000”, September.

[5] Columbia Accident Investigation Board (CAIB), 2003, “Report, Volume 1”, August, and “Working Scenario”, CAIB/NASA/NAIT, July, and “Volumes II through VII”, October (available at http://www.caib.us).

[6] Jaynes, E.T., 2003, “Probability Theory: The Logic of Science”, First Edition, Cambridge University Press, Cambridge, U.K, Edited by G.L. Bretthorst.

[7] Baumont, G., Bardou, S., and Matahri, N., 2000, “Indicators of Plant Performance During Events Identified by Recuperare Method”, paper presented at Specialist Meeting on Safety Performance Indicators, Madrid, Spain, October 17-19.