Essay-3 The human factor

Precis: While the human factor initially was about the shortcomings of human performance relative to the perfection of machines (epitomised by the Fitts' List) it soon became associated with one specific feature, namely the unpredictability and unreliability, that turned into the concern for human error, in the sense that variable and unreliable performance of humans quickly became the favourite candidate for causes of UUOs. This essay will present a broader view of human factors. The particular human error view is the topic of the following fourth essay.

The concern for human factors

Professors Andrew Hale and Jan Hovden (Hale & Hovden,1998) defined the second age of safety thinking as the age of human factors (Figure 1 and Table 1).

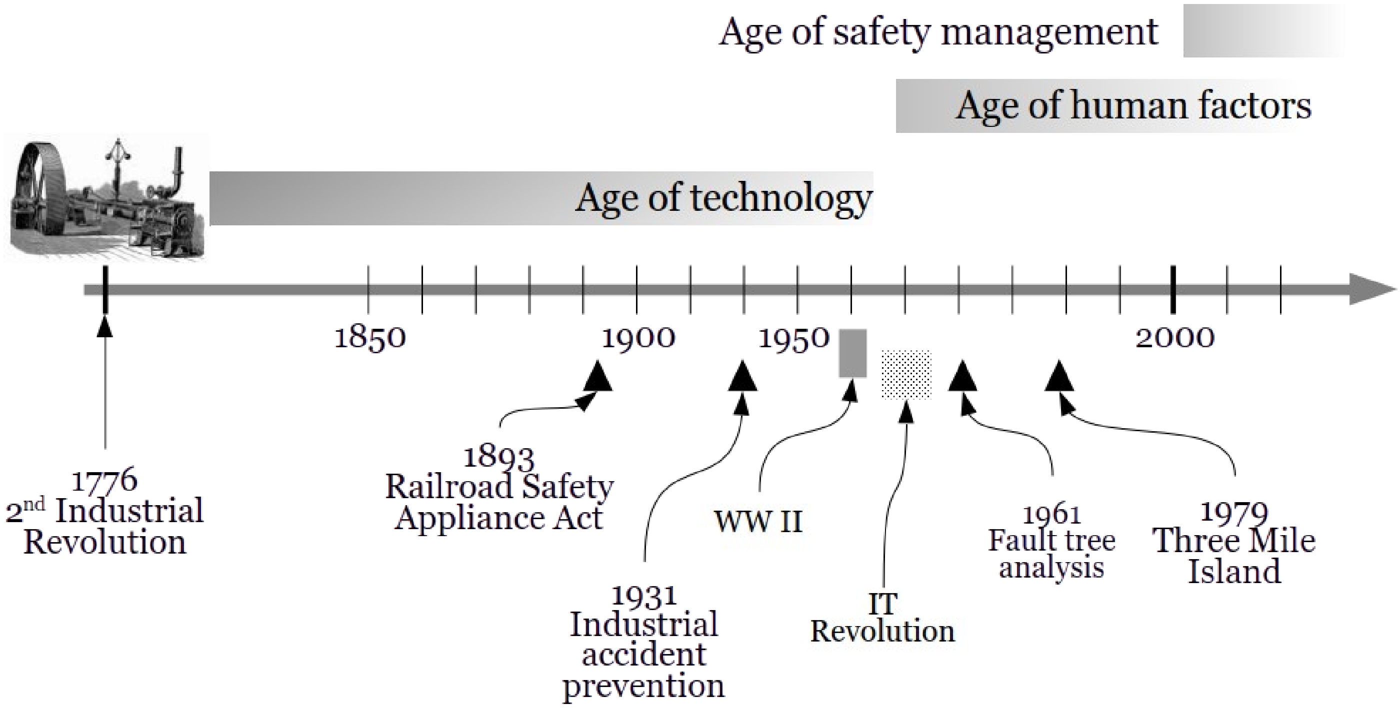

Figure 1: The three ages of safety based on (Hale & Hovden, 1998).

Table 1: Characteristics of the three ages of safety (from Hale &Hovden, 1998).

Age of safety thinking	Typical (default) cause	Typical (default) response	Characteristic mode of causality	Initiating event
I. Technology	Failure of technology	Replace, repair and improve	Linear, Monocausal	Watt's steam engine 1769
II. Human factors	Human factor "human error"	Blame, train, design, automate	Linear, Monocausal	Multiple accidents during World War II
III. Age of safety management	Organisational failure	Audits,standardisation regulation	Linear, Polycausal	Challenger & Chernobyl (1986)

The original version of the domino model Figure 2 (Heinrich, 1959) included Fault of Person as the second domino piece, as defined as in and Table 2:

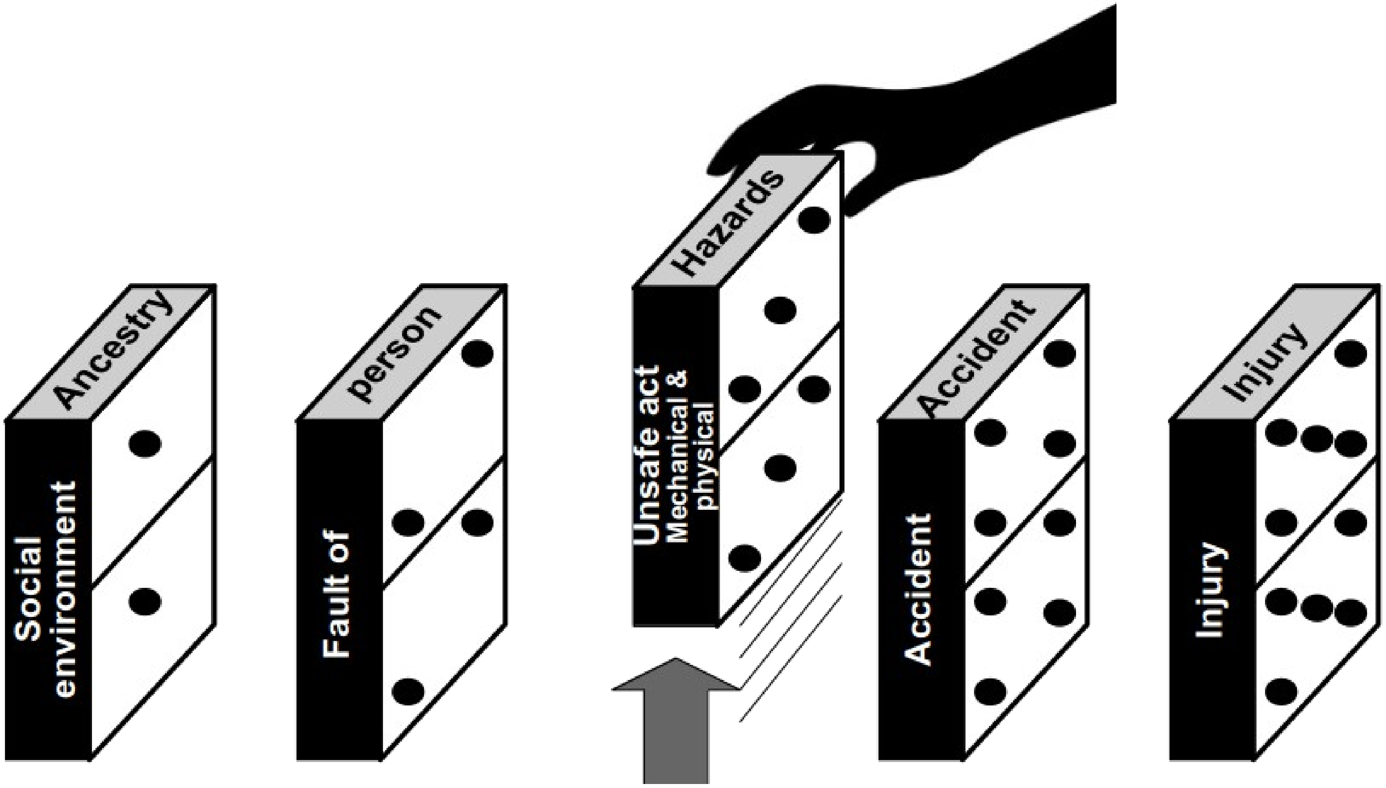

Figure 2: The five factors in the accident sequence (Heinrich, 1959)

The reasoning principle of the model of the accident sequence is simple. Indeed the very name “accident sequence” makes it obvious. Each domino piece that tumbles is the assumed cause of what happens next. The simple physical analogy of tumbling domino pieces has been used in many other contexts During the Cold War, the US foreign policy was influenced by the so-called domino theory (O’Sullivan, 2021), the domino effect was also used to render the financial crisis in 2008 in understandable terms for the general public and some politicians. What is interesting, but not made clear is whether the first domino piece ”Ancestry and social environment” actually is the first; the Pre-socratic philosopher Leucippus of Miletus (c. 480–c. 420 BC), is credited with declaring that ‘Nothing happens in vain (by itself), but everything from reason and of necessity.’ So neither ancestry nor the social environment appeared out of the blue. Indeed, the undesirable traits of character are directly said to be “passed along”, cf. Table 2, which means they must have arisen at a previous, undefined stage. The same uncertainty applies to the last piece “injury” Heinrich himself had introduced the concept of hidden costs, to describe the hard to detect consequences that follow an injury. So the accident sequence neither really begins by the first domino piece, nor ends by the fifth and last. It is possible to continue to apply the same sequential reasoning in either direction for as long as one likes, or as far as the imagination allows.

Although Heinrich’s use of “fault of person” and “unsafe performance of persons” referred to humans, this was not the beginning of human factors. The contribution of humans at a place of work is absolutely essential, without humans nothing can be done, automated machines can do many things, but they cannot yet by themselves decide what they should do or when to do it. They have to be instructed and activated by humans. Until now we still use machines for our purpose, and woe the day when machines begin to use us for their purposes. Heinrich’s terminology does not correspond to how the term human factors is used today, it merely introduced the contagious suggestion that something done by a person could be part of the explanation of an OOU.

Table 2: Details from the accident sequence (Heinrich, 1959).

Name of domino piece	Explanation
1. Ancestry and social environment.	Recklessness, stubbornness, avariciousness, and other undesirable traits of character may be passed along through inheritance. The Environment may develop undesirable traits of character or may interfere with education. Both inheritance and environment cause faults of person.
2. Fault of person.	Inherited or acquired faults of a person; such as recklessness, violent temper, nervousness, excitability, inconsiderateness, ignorance of safe practice, etc., constitute proximate reasons for committing unsafe acts or for the existence of mechanical or physical hazards.
3. Unsafe act and/or mechanical or physical hazard.	Unsafe performance of persons, such as standing under suspended loads, starting machinery without warning, horseplay, and removal of safeguards; and mechanical or physical hazards, such as unguarded gears, unguarded point of operation, absence of rail guards, and insufficient light, result directly in accidents.

Despite the widespread concern for and ditto importance of what humans do at a place of work, there are no simple definitions of what human factors are. The UK Health and Safety Executive define human factors as" "Thee nvironmental, organisational and job factors, and human and individual characteristics, which influence behaviour at work in a way which can affect health and safety" There are many available accounts of human factors as a discipline as an area of R&D as well as a Handbook of Human Factors and Ergonomics that regularly comes in a new edition, but no definitive authoritative account not even in (Dul et al., 2012), who just wrote that:

> "HFE focuses on systems in which humans interact with their environment. … The focus of HFE is to jointly improve performance and well-being by designing the integrative whole better, and by integrating the human into the system better." (Dul et al., 2012, p. 3).

(Meister & Enderwick, 2001) may be a better place to look.

Few, however, realise that the very term human **factors**, endorses the view that a system can be understood by decomposition, by breaking it into its constituent parts, following the millenia old tradition of atomism, that is central to the safety legacy. Doing so raises the question of what is meant by the term system.

What is a system?

The traditional and widely-shared view of humans as parts or components of work systems, applying a physical mechanical analogy that goes back at least to Julien Offray de la Mettrie (1996) whose treatise "Man a Machine" was published originally in 1748 not forgetting that the Golem (Wiener, 1964) is mentioned already in the Bible, according to this way of thinking humans are intricate mechanisms or machines, but machines nonetheless hence equal parts or components of the larger work system where they can be treated as if they were mechanical parts,This was recognised when (Taylor & Garvey, 1959, p. 187) wrote

> "It is now common to regard the human operator of a machine and the machine itself as two elements in one overall man-machine system. "The pilot and his plane, the helmsman and his craft, and the lathe operator with his lathe are examples of such systems."

Humans are thus described both as a mechanical component themselves and as a genuine part of a larger socio-technical or workplace. Although the Golem was not a machine in the sense we use the term today, but more like Galatea and Pinocchio – a humanoid object made from clay (Golem) marble (Galatea) or wood (Pinocchio) that miraculously became alive. The temptation to think of humans as a machine only grew stronger and became nearly irresistible after digital computing machinery became a reality around the middle of the 19th century, spearheaded by Turing (1950), followed by von Neumann (1958), and Arbib (1987), and concluded most famously by Ashby (1960). These and many other papers beginning by what the mathematician George Boole wrote about the Laws of Thought (Boole, 1854) made it philosophically and scientifically legitimate to describe the human operator as a machine and inadvertently supported the outrageous idea that human information processing could become a replacement for psychology (Lindsay & Norman, 1972), Newell & Simon 1963). The comparison between humans and machines was made on technological premises and therefore rarely advantageous to humans. An early expression of that view was provided by Paul Fitts, the undisputed pioneer of Human Factors Engineering, who wrote that

> ”the final consideration which needs mention is relative fallibility of a man to a machine” (Fitts et al., 1951). p. 6)

Since then it has been commonly agreed that humans at best are fallible machines. Fitts also noted that

> ”We have been very much occupied in perfecting the machines and tools which the worker uses in the economic arts, following which he continued:
>
> “We have hardly attempted to improve the worker himself.” (Fitts, et al., 1951 P. iv). The purpose of human engineering was clearly to improve the worker himself. That is no longer an acceptable purpose.

The generic solution that Paul Fitts was instrumental in developing became known as human factors engineering,today modsly called human factors and Ergonomics (HFE) although Fitts just called it human engineering with the clearly stated purpose to engineer the human to provide a better fit with the machines. The three main remedies which from the beginning were endorsed by human factors were, **training**, **design**, and **automation**. Fitts' concerns led him to propose a rigorous method to compare humans and machines, that we now known as the Fitts' List which he introduced in the following way:

> We begin with a brief analysis of the essential functions ... We then consider the basic question: Which of these functions should be performed by human operators and which by machine elements? (Fitts, 1951 p.x).
>
> Althoug Paul Fitts is the undisputed pioneer of Human Factors Engineering, the concern for what we now call ergonomics goes even further back to (Jastrzębowski, 1857), although it was then called praxiology.

The Fitts List later achieved fame as the MABA-MABA List after the initials in (Men Are Better At) — (Machines Are Better At) (Dekker & Woods, 2002).

System definition

The notion of a human-machine system (HMS) became very popular initially called just a Man-Machine System (MMS). The term referred to the conventional definition of a system as:

> “a set of objects together with relationships between the objects and between their attributes” (Hall & Fagen, 1969, p. 81) –

or even as anything that consists of parts connected together. In this definition, the nature of the whole is arbitrary, and the boundary of the system is therefore also arbitrary.

This view is illustrated by the following delightful quote:

> “It is legitimate to call a pair of scissors a system. But the expanded system of a woman cutting with a pair of scissors is also itself a genuine system. In turn, however, the woman-with-scissors system is part of a larger manufacturing system – and so on. The universe seems to be made up of sets of systems, each contained within a somewhat bigger, like a set of hollow building blocks.”(Beer, 1959, p. 9)

Taking a different approach Cognitive Systems Engineering (CSE) (Hollnagel & Woods, 1983) defined a system by how the boundary depends on the system's functions, leading to the following

Table 3: *Definition of boundaries in Cognitive Systems Engineering (CSE)*

	functions that are essential for the ability of the JCS to maintain control.	Functions that are of no consequence for the ability of the JCS to maintain control.
Objects that can be effectively controlled by the JCS.	1.Objects are included in the JCS	2. Objects may be included in the JCS
Objects that cannot be effectively controlled by the JCS	3.Objects are not included in the JCS	4. Objects are excluded from the description as a whole.

The lack of an authoritative account of what human factors are has consequences also for the history of human factors – when did the concern for the human factor begin. Some date the beginning to 1898 and the efforts of Frederick W. Taylor, a mechanical engineer who promoted the idea of Scientific Management (Taylor, 1911).

Taylor's Scientific Management, today known as Taylorism is interesting because the brazen purpose was to increase productivity, to make humans more efficient components in a place of work. This is obvious from the four main principles Taylor proposed (Taylor, 1911 p. 14):

1. Analyse tasks to determine most efficient performance.
2. Select people to achieve best match between task requirements and capabilities

3. Train people to ensure the specified and required performance but nothing more and nothing less than that,
4. Insure compliance by economic incentives

The first principle alone, makes clear that the primary concern was productivity rather than the well-being of the workers.

Some pretentiously date the beginning of human factors several centuries earlier to Leonardo da Vinci's famous drawing of the Vitruvian man from 1490. (Included at the bottom right in Figure 3. It is, however, unreasonable to assume that Leonardo had the conditions of an industrial worker in mind, since industrial work did not exist at the time. It is most likely just a study of the ideal human proportions.

Others realistically put the start much later, after the end of the Second World War, (Hale & Hovden, 1998) in their proposal of three ages of safety, for instance wrote:

> "Just as the second age of human factors was ushered in by increasing realizations that technical risk assessment and prevention measures could not solve all problems, so were the 1980s characterized by an increasing dissatisfaction with the idea that health and safety could be captured simply by matching the individual to technology. [which otherwise was the foundation of human factors engineering]. The 1990s are already well into the third age of safety, where management systems are the focus of development and research." (Hale & Hovden, 1998, p. 130).

Since the start of an age of safety usually is due to the presence of UUOs that defy the hitherto habitual explanations, cf. Table 1, it is far more reasonable to date human factors to the period after the end of the Second World War, because the discrepancy between the many systems (weapons) invented during the war and designed to make optimum use of technological capabilities on the one hand and the innate or natural human ability to fit into these newly designed systems on the other became so large that it constituted a problem both for the efficiency by which these systems performed, and for their reliability, that it became necessary to do something about it. (Hale & Hovden, 1998) in their proposal of the three ages of safety, saw it primarily as a safety problem, recognisable by the tendency to accept the human factor in general and human error in particular as the default socially accepted causes. Human error is itself the subject of the next essay. Yet an explicit concern for the human factor became a problem already at the start of the first age of safety thinking, that (Hale & Hovden, 1998) called the age of technology marked by the widespread use of the steam engine that ushered in the second índustrial revolution. (It is actually possible and meaningful to distinguish among even more ages or stages in safety thinking. (Hollnagel, 2026, for instance proposes no less than seven stages or eras, including, of course the three described by (Hale & Hovden, 1998). Much of today's fashionable discussion about a sequence of fictitious industry generations, for instance, Industry 5.0, without justifying the use of decimals in the numbering, miss that the first industrial revolution was the change to agriculture about 12.000 years ago, and that industry 5.0, therefore ought to be industry 6.0. The transition to agriculture required as much use of novel technology as the proper industrial revolution in 1769, although as farming equipment, flails and ploughs pulled by oxen rather than large, noisy steam engines. The second industrial revolution in the 18th century logically required a revised focus on the human factor, because people from then on had to work in conditions (factories) designed by others rather than as artisans in conditions gradually developed by themselves to suit their own needs. The change to agriculture did not change the farmer's place of work in the same way

Training as a human factors solution

Human factors from the very beginning clearly needed ways to engineer and improve the human factor, mostly to ensure a better fit between humans and the machines they had to work with but also to make humans perform more reliably like machines or technological artefacts, as addressed by the Fitts' List. Training, design, and automation were adopted from the start as solutions and have been used ever since, although in various proportions since they are not mutually independent.

- **Training** was used to shape humans to meet the requirements of technology and was initially seen as an ideal and inexpensive solution, which unlike design and automation had a pedigree stretching back thousands of years, which was already widely practiced by most industries and therefore familiar.
- **Design** was used to ensure that the requirements of machines corresponded to the innate abilities of humans, as determined e.g., by the Fitts' List.
- **automation** was the ultimate solution of replacing humans by technology, thereby enabling, if not ensuring a perfect match between system components.

It did, however, not take long before it became clear that training was not the perfect solution to engineer the human factor and overcome the problems, due to performance variability, and insufficient reliability. Training became known as a “Procrustean” approach referring to the legendary ***Procrustes***, a robber of Attica, who had an inn on the sacred road between Athens and Eleusis (home of the Eleusian mysteries). The inn only had a one iron bed, where Procrustes invited everyone who passed by to spend the night. Procrustes became famous for making the visitors fit the bed, rather than the other way around. If visitors were shorter than the bed, Procrustes would stretch them until they were long enough, and if they were too long he would make them fit the bed by having parts of their legs cut off. (Taylor and Garvey, 1959) used the Procrustes analogy to criticise training when they pointed out that:

> “Two rather different human factor approaches may be distinguished in efforts to optimise the performance of man-machine systems.
>
> - One approach tries to standardise performance by “Making people shorter” , i.e., limit or constrain what people should do, to limit and constrain their performance and to use less than their full potential.
> - The other effort approach is to “make people longer" to extend or stretch human capabilities to meet task demands through additional specialised training, and or technological support

The Achilles heel of human factors, as a discipline, is the blind acceptance of the machine analogy as an explanatory, analytical principle for what a human is, which meant that humans from the very beginning of human factors were seen as inefficient variable, and unreliable, a view that led to the perceived need of human factors engineering cf. the Fitts quote above. The machine analogy is still widely accepted and remains an essential part of the safety legacy humans are, therefore, seen as a liability in consequence of which the human factor, and specifically human performance variability became problems that had to be solved. Fitts provided the justification when he wrote that:

> “We have been very much occupied in perfecting the machines and tools which the worker uses in the economic arts. We have hardly attempted to improve the worker himself.” (Fitts, 1951, p. iv).

It was this postulated need to improve or engineer the worker that gave rise to human factors engineering which at some point became just human factors. The starting point was then as now the widely accepted view that the human operator was unreliable and inefficient compared to machines and that these imperfections made it impossible fully to exploit the advantages new technologies seemed to offer, hence necessary to engineer the human component. The Main driving forces behind the ever-growing complexity of human-machine systems and work environments, in the 1940s and today, are summarised in Figure 3.

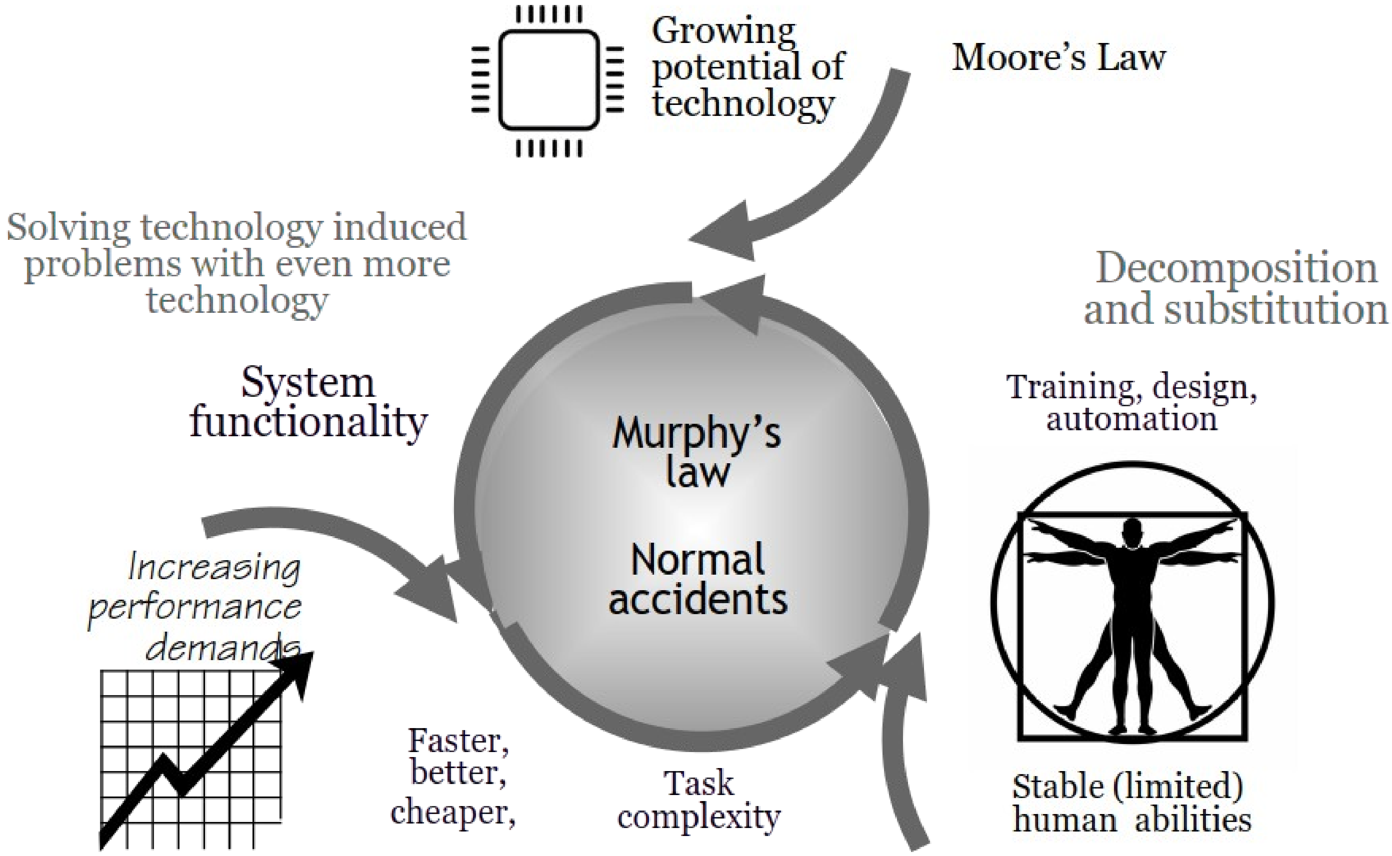

Figure 3: The forces driving the increasing complexity of socio-technical systems

In Figure 3, the growing potential of technology is mostly a euphemism for saying that someone has spotted an opportunity to save money. In recent years often seen by the introduction of remote work in many industries such as remote control of offshore oil rigs, because transporting operators by helicopter from land to an off-shore platform is both risky and expensive. Other examples are the use of remote towers in air traffic management (Papenfuss & Friedrich, 2016), remote control of ships sailing on regular routes, and even robotic surgery (Nyssen & Blavier, 2019). In relation to NASA's famous slogan of Faster Better Cheaper slogan (Paxton, 2007), cheaper is practically always the highest priority for industries (whereas for NASA faster was most important.

Technology is in most cases a solution looking for a problem and is usually adopted because it seems advantageous to do so rather than because there is a well-defined need. Most managers are what the father of Cybernetics Norbert Wiener (1964), called gadget worshippers, defined as people who:

> "regard(ed) with impatience the limitations of mankind, and in particular the limitation consisting in man's undependability and unpredictability"

Today Wieners' concept of gadget worshipping has practically become a design philosophy of its own, called solutionism described as:

> An intellectual pathology that recognizes problems as problems based on just one criterion: whether they are "solvable" with a nice and clean technological solution at our disposal. (Morozow, 2013a & b).

New technology may well be introduced with a specific and reasonable purposes in mind, even if these are more likely to be of an economic rather than an ergonomic nature, but any change soon falls prey to what Larry Hirschhorn called *The Law Of Stretched Systems* which simply states:

> Every system is stretched to operate at its full capacity; whenever improvements are made, for whatever reason, they will be used to achieve a new intensity and tempo of activity — to stretch the system a bit further. (Hirschhorn, 1997)

Training used to limit performance to match demands

The conventional idea that people should have just the competence needed to do the job, but no more because it would be a waste of effort, got a new lease on life with the introduction of Scientific Management (Taylor, 1911). Scientific management hoped that performance variability could be reduced and work efficiency improved by ensuring that people did exactly what they should do, neither with more nor less requirements of the system. The first solution of “making people short” resembles the third principle of Taylorism, reproduced above .

Training used to extend performance to meet demands

In the Procrustean analogy, training can also be used to “Make people longer " i.e., to extend or stretch their capabilities and skills beyond what they would naturally or normally do by means of long and specialised training, This has unfortunately has become increasingly necessary as we relentlessly build human-machine systems and work environments that no longer are intuitive despite ambitious attempts such as ecological interface design (EID) (Vicente & Rasmussen, 1992) and State, Operator, And Result (SOAR)(Laird, Newell & Rosenbloom,1987), cf. Figure 3. Few people can naturally or intuitively fly a modern aircraft, or control a nuclear power plant, and perhaps not even drive a present day EV. On the other hand few of us would presumably be able to drive a Ford T. It would be too simple and the instrumentation and controls would be unrecognisable to us.

Limitations of training

Either solution provides only a temporary relief at best, The problem with both is that people inevitably and involuntarily will revert to their “natural” level of skills and performance whenever something unexpected happens, as it is bound to do according to Murphy’s Law (Bloch, 2003) because even though the *probability* of any specific type of disturbance always can be made acceptably low, the *possibility*, that something will happen will still be equal to 1. This, however, does not mean that the probability is also 1, because there is a crucial, but mostly unacknowledged difference between possibility (Zadeh, 2014) and probability. (Dubois & Prade, 1989), and when the unexpected happens people will either neglect or forget the constraints and therefore do more than they were supposed to do (e.g., intervene, manage by exception (Dekker & Woods, 1999) , hence do too much or do it too early), or revert to their former level of comfort and competence and forget what they have been taught and therefore do less than they were supposed to do (e.g,regress to habitual, intuitive or default actions, wrong procedure, do too little and do it too late). Such reactions will inevitably increase the difference between what people actually do. (Work-as-Done, or WAD) and what they were supposed to do (Work-as-Imagined or WAI), which is the opposite of the intended outcome In either case the predictable and involuntary return to the “normal or “natural" way of work renders the system incapable of functioning as intended. Despite the allure of the machine analogy and the promises of Skinnerian behaviorism, humans cannot be programmed as mindless robots, Taylorism notwithstanding. Despite the longevity of the machine analogyHumans are emphatically not machines. It was presumably not a problem for Procrustes if people regained their original length during the night. The stretching might well cease to have an effect, but legs would not spontaneously begin to grow. Training has since the early years been an essential solution for human factors engineering as such and the popularity has hardly diminished over time, despite insufficient evidence that it actually works. Training is, of course, often supplemented by recruiting, by selecting people who naturally have the required competencies as already suggested by the second principle of Scientific Management.

The automaton analogy

Since the systematic study of Human-Machine Interaction (HMI) arose from the need to solve practical, technological problems, the basis for the description was the engineering view of men and machines. The technical and engineering fields had developed a powerful vocabulary to describe how machines worked and it was therefore natural to use the same vocabulary to describe how people worked, hence as a basis for modeling human performance e.g. (Stassen, 1986).

The automaton analogy denotes how one can think of or describe a human being as an automaton or a machine. A particular case is the use of the information processing metaphor (Newell & Simon, 1972, Simon (1972 and Newell, 1990). But the automaton analogy can be found in practically every explanation of human performance , e.g. by behaviourism or psychoanalytic theory (cf. e.g. Weizenbaum, 1976). Although it seemed convenient to do so, the negative consequence was that it forced designers to think of people as if they were machines, a practice that can only be called **the forced automaton analogy**.

An automaton can be described by a set of inputs, outputs, internal states and the corresponding state transitions; a classical example of this is the Turing machine. More formally, a finite automaton is a quintuple: (e.g. Arbib, 1964; Lerner, 1975):

$$A = (I, O, S, \lambda, \delta)$$

where I is the set of inputs,
O is the set of outputs
S is the set of internal states

λ: $S \times I \rightarrow S$ is the set of rules for determining the next state, and
δ: $S \times I \rightarrow O$ is the set of rules for determining the next output.

A machine, or a program, can be described as a finite state automaton or a state machine in terms of the quintuple defined above (a set of inputs, outputs, internal states, and state transitions: all programming languages are actually based on that assumption). In order for the machine to work and to produce a predefined output, it must get a recognisable input. This means that the human user must respond in a way that corresponds to the predefined categories of inputs that the machine can recognise. If not, the machine, hence the Joint Cognitive System (JCS) (Hollnagel & Woods, 1983) cannot function appropriately; it will become ineffective or even unreliable.

Worker-as-imagined

Training and design both imply an idea not just of Work-as-Imagined – but also of Worker-as-Imagined, just as design implies the idea of work-as-imagined (to be discussed later). Scientific Management was refreshingly honest about a worker-as imagined and the extreme case is, of course, the way Procrustes tried to make the physical size of his visitors fit the iron bed. (He might also have posted a sign saying "bed requires people who are 178 cm long") or whichever unit was used at the time. No sane person would, of course, dream of using a Procrustean philosophy today, instead we expect that people do it themselves by bending and stretching leading to work-related stress and physical effects such as back pain. Ergonomics has made a Procrustean philosophy obsolete for physical sizes, but not for cognitive sizes or mental capacity, forcing people to adjust themselves to the many wonderful services the current technology offers. We learn to navigate complicated websites and follow the reasoning of the ubiquitous chatbots that replace human customer service agents.

The allure of standardisation

Training can also be used to "Make people longer" i.e., to extend or stretch their capabilities and skills beyond what they would naturally or normally do through long and specialised training, This has unfortunately become increasingly necessary as we relentlessly build human-machine systems and work environments that no longer are intuitive despite the recognised need and pretentious attempts such as ecological interface design (EID) (Vicente & Rasmussen, 1982) and State, Operator, And Result (SOAR) (Laird, Newell & Rosenbloom,1987), cf. Figure 3. Standardisation is generally attractive because it improves predictability. At least in the designer's imagination, if something conforms to a standard in size, the measure of functioning uncertainty is in principle reduced. If physical components are standardised, it becomes easier to assemble them, cf. (Womack, Jones & Roos, 1990). In terms of efficiency-thoroughness trade-off efficiency can be increased locally because thoroughness presumably has been taken care of already.

Design as a human factors solution

If training is unable to ensure that human performance corresponds to the requirements of the technology of the Human-Machine System (HMS), the seemingly logical alternative, at least until we find a way to design humans or build effective humanoid robots, is to design the technology so that any requirements correspond to what people naturally are able to do, requiring neither too much, nor too little. Design therefore became the second main element to ensure effective human-machine interaction (HMI) and the second tool in the human factor toolbox. Cf. (Hollnagel,1993).

> "To improve the performance characteristics of these man-machine combinations, one has the choice of either trying to alter the man so that he fits the machine better, or to modify the mechanisms to fit the man. The latter approach is that of the engineer and the new specialist called the human engineer, while efforts to adapt the man to the machine characterise the work of the training expert." (Taylor & Garvey, 1959, p. 187.)

In order for the designer "to modify the mechanisms to fit the man", the designer must have a realistic assessment of what a person, an operator is able to do, that surpasses what the Fitts List is able to provide. The designer must find a way to describe the user's capabilities and how the user performs that is reasonably accurate. Human factors required this because of the assumptions that people were inefficient, variable, and potentially unreliable, and therefore had to develop the corresponding solutions. The attractiveness of considering the human as just a part of a system has already been mentioned. But design as a human factors solution did not just consider the user as a simple part of a system, but as an automaton.

The automaton analogy

An automaton can in general be described by a set of inputs, outputs, internal states and the corresponding state transitions; a classical example of this is the universal Turing machine (Turing, 1938). More formally, a finite automaton is a quintuple (e.g. Arbib, 1964; Lerner, 1975):

$A = (I, O, S, \lambda, \delta)$

where I is the set of inputs,

O is the set of outputs

S is the set of internal states

$\lambda: S \times I \rightarrow S$ is the set of rules for determining the next state, and

$\delta: S \times I \rightarrow O$ is the set of rules for determining the next output.

A machine, or a program, can in general be described as a finite state automaton or a state machine, in terms of the quintuple defined above (a set of inputs, outputs, internal states, and state transitions; all programming languages are actually based on that assumption. In order for a machine to work and to produce a predefined output, it must get a correct input. For a Human-Machine System (HMS) this means that the human must respond in a way that corresponds to the predefined categories of input. If that is not done, the machine, hence HMS will not function appropriately; the HMS will become unreliable or worse ineffective. A HMS can more generally be understood as a Joint Cognitive System (JCS) defined as a system that controls its behavior on the basis of experience to achieve its goals. The term Joint Cognitive System specifically implies that control is accomplished by an ensemble of Cognitive Systems.and (physical or social) artefacts that are capable of goal-directed behavior. Humans are natural cognitive systems, while some hope and others fear that an artificial system some day can become functionally equivalent to humans. Cognitive Systems Engineering (CSE) (Hollnagel & Woods, 1983) was proposed as an alternative to traditional human-machine systems to study how typically one or several people (operators or controllers) and one or several support systems as parts of a Joint Cognitive System, together could control or regulate a process or an activity in an incompletely understood and partly unpredictable environment. The specific automaton analogy alluded to here is useless and even potentially misleading for HMI, HCI studies, (Hollnagel, 1992). I will not argue that the automaton analogy is ineffectual as a basis for describing human performance per se; I simply take that for granted. (This point of view has certainly not always been generally accepted and often not even clearly stated even when it was used extensively, for instance, by the mainstream of American Cognitive Science in the 1970s; it is nevertheless a view which is fairly easy to support.) I even believe that the automaton analogy is useless for machines in the context of human-machine systems where the functioning of the machine must be seen together with the functioning of a person.

The forced automaton analogy

The automaton analogy also exists in a stronger version, named The **forced** automaton analogy. The explanation follows.

If a Joint Cognitive System (JCS) or a HMS is to function properly the input to the machine, which is the output from the user, must fall within a limited set of inputs that can be recognised by the machine. But the determination of the user's response is at least partly determined by the information available. The output from the machine constitutes the input to the user and is partly determined by the previous input i.e., what the user did before, which means the JCS is a non-trivial system (von Főrster & Pőrksen, 2002). The user must be able correctly to interpret and map the input provided onto one of the predefined answer options. In order to do that the designer is forced to consider the user as a finite state automaton, hence as a machine. We know from practice that people may interpret information in many different ways depending on their experience, on habits and the context. We also know that it is impossible to account for this infinity of interpretations (Work-as-Imagined or WAI will inevitably be different from Work-as-Done or WAD) we are therefore forced to assume that the user interprets the information only in the limited number of ways the designers have been able to imagine. Designers are therefore forced to think of the user as a finite automaton and therefore think of the user as a finite automaton. A good example of that is the ubiquitous graphical user interface, which basically serves to restrict the user's degrees of freedom, by limiting the number of responses the user can choose from.

The forced automaton analogy suggests that it is legitimate to think of or describe a human as an automaton or a machine. A particular case is the information processing metaphor used by (Newell & Simon, 1972, Simon (1972) and (Newel, 1990) —- or even worse, preposterously claiming that a human is an information processing system (as for instance (Simon, 1972) and (Newell, 1990) both did.)

The forced automaton analogy means that the starting point to think of the machine as a finite automaton willy-nilly forces the designer to think of the user as a finite automaton as well because there is no other way to think of the user that will fit the requirements of the design. Yet doing so is clearly unacceptable, no matter how sophisticated we assume the automaton to be (even if the number of states is exceedingly large and the state transitions are stochastic or multi-valued). It probably is the case that whatever analogy we use for one part of a Human-Machine System must also be used for the other. Because we want to retain some distinct human elements in the description of the user we are forced by this argument to apply the same elements to the description of the machine (a kind of forced anthropomorphism). One proposal for an alternative description is the notion of a cognitive system as defined by Hollnagel & Woods (1983, p. 345):

> “A cognitive system produces intelligent action, that is, its behaviour is goal oriented, based on symbol manipulation and uses knowledge of the world (heuristic knowledge) for guidance Furthermore a cognitive system is adaptive and able to view a problem in more than one way. A cognitive system operates using knowledge about itself and the environment in the sense that it is able to *plan* and *modify* its actions on the basis of that knowledge. (Hollnagel & Woods, 1983, p. 345)

With regard to design as a human factors solution, in the field of Human-Machine Systems HMS, and in particular Human-Computer Interaction (HCI) the automaton analogy is useless and even misleading. One may go even further and argue that the automaton analogy is useless, and possibly harmful even for machines when the context is Human-Machine Systems, where the functioning of the machine must be seen together with the functioning of a human. (Hollnagel, 1992).

Automation as a human factors solution

If neither training nor design are able to ensure that human performance corresponds to the requirements of the technology of the Human-Machine System (HMS), then the logical alternative is to replace the human by automation, thereby presumably ensuring a perfect match between capabilities and demands.

Automation is usually defined as the execution by a machine, usually a computer, of a function that was previously carried out by a human (Parasuraman & Riley, 1997). At present automation is practically synonymous with the use of computers and computing technology, increasingly in the form of vanishingly small and specialised components. Yet the history of automation goes back to the water clock invented by Ctesibius (c. 285–222 BCE) a brilliant Greek inventor and mathematician from Alexandria, more than 2.000 years before the advent of computers – even if we take Joseph Jacquard’s loom controlled by punched cards (1801) as an early example of the latter.

It is possible to distinguish three main automation philosophies

- The left-over, or proto human factors principle (ca. 1910)

Functions that cannot be assigned to machines are left for operators to carry out. This corresponds to the second of the three ironies of automation, described by (Bainbridge, 1983). The main concern is efficiency. Despite its age, this philosophy is still very much alive.

- The compensatory principle (ca. 1946)

Functions are assigned based on juxtaposing human-machine capabilities (for instance using the Fitts’ List). The main concern is the usability of the Human-machine system.

- Complementarity principle (ca. 1985).

Function allocation aims to sustain and strengthen human ability to perform efficiently. The main concern is to ensure that the joint human-machine system can remain in control of the situation.

Professor Thomas Sheridan later defined 10 possible levels of automation, as shown in Table 4.

Table 4: Levels of human and computer control (Sheridan, 1982).

Level of automation	Characterisation
1	Computers offer no assistance, humans must take all decisions and actions.
2	Computer offers a complete set of decision/action alternatives
3	Computer narrows selection down to a few alternatives
4	Computer suggests one alternative
5	Computer executes suggested alternative if human approves
6	Computer allows human restricted time to veto before automatic execution
7	Computer executes automatically, then necessarily informs humans
8	Computer informs human only if asked
9	The computer informs humans only if it decides to.
10	Computers decide everything and act autonomously, ignoring humans.

Professor Sheridan developed his characterisation in the context of human and computer control of undersea teleoperations. Most people will probably prefer that the system design does not go above level 5, which leaves the human in control.There are, unfortunately, examples of automation at higher levels, especially in aviation, which have resulted in UUOs. In the context of undersea teleoperations that Sheridan considered, it is, of course, possible to stop the process for a while to allow the human operator time to think, while the underwater robot just waits on the ocean floor. For an airborne aircraft it is not possible to halt the process for even a short while. Available time and time pressure are important conditions to consider in the design of joint human machine systems.

- Identify the general essential constraints that must be met by the design, e.g.
- Identify the specific essential constraints that hold for specific conditions that may occur, e.g.
- Describe the options for design and the consequences of each option vis-a-vis the essential functional requirements,
- Identify the specific supplementary constraints that may apply to each situation (i.e., constraints which comply with the functional requirements, but relate to additional requirements that only need be fulfilled if the conditions allow), (quoted from Booth, 1989)

The automaton analogy

An automaton can in general be described by a set of inputs, outputs, internal states and the corresponding state transitions; a classical example of this is the Turing machine (Turing, 1938). More formally, a finite automaton is a quintuple (e.g. Arbib, 1964; Lerner, 1975):

$$A = (I, O, S, \lambda, \delta)$$

where I is the set of inputs,
O is the set of outputs
S is the set of internal states
$\lambda: S \times I \rightarrow S$ is the set of rules for determining the next state, and
$\delta: S \times I \rightarrow O$ is the set of rules for determining the next output.

A machine, or a program, can in general be described as a finite state automaton or a state machine, in terms of the quintuple defined above (a set of inputs, outputs, internal states, and state transitions; all programming languages are actually based on that assumption. In order for the machine to work and to produce a predefined output, it must get a correct input. For HCI this means that the human user must respond in a way that corresponds to the predefined categories of input. If not, the machine, hence the joint cognitive system (Hollnagel & Woods, 1983), will not function appropriately; the JCS will become unreliable or worse ineffective. A Joint Cognitive System (JCS) can control its behavior on the basis of experience to achieve its goals. The term joint cognitive system means that control is accomplished by an ensemble of cognitive systems, humans are natural cognitive systems, and (physical and social) artefacts that exhibit goal-directed behavior. In the areas of interest to Cognitive Systems Engineering, typically one or several people (operators or controllers) and one or several support systems are part of a Joint Cognitive System, which must control or regulate a process in an incompletely understood and partly unpredictable environment.

Assume, for instance, that the automaton is in a state S_j. From that state it can progress to a predefined set of other states (S_k, ..., S_m) or produce a predefined output (O_j) only if it gets the correct input (I_j). If the system in question is a joint human-machine system then the input to the automaton is provided by the user. Therefore, the user must therefore provide an input that the automaton can interpret. Any other response by the user will lead to one of two cases:

- The input may not be understood by the automaton, i.e., transition functions λ and δ are not defined for the input. In this case the automaton may either do nothing or move to a default state.
- The input may be misunderstood by the automaton, for instance if the semantics or the syntax of the response have not been rigorously defined. In this case the automaton may possibly malfunction, i.e., be forced to a state which has not been anticipated. This happens easily if the input was a physical control action or manipulation, e.g. like switching something on or off, opening a valve, etc.

If the Joint Cognitive System (JCS) therefore is to function properly the user must provide a response that falls within a limited set of possible responses the designer has been able to imagine (the set of inputs that can be recognised by the machine). But the determination of the user's response is at least partly determined by the information available. The output from the machine constitutes the input to the user and is partly determined by the previous input i.e., what the user did before, which means the JCS is a non-trivial system (von Főrster & Pőrksen, 2002). The user must be able correctly to interpret and map the input provided onto one of the predefined answer options. In order to do that the designer is forced to consider the user as a finite state automaton, hence as a machine. We know from practice that people may interpret information in many different ways depending on their experience, their habits and the context. We also know that it is impossible to account for this infinity of interpretations (Work-as-Imagined or WAI will inevitably be different from Work-as-Done or WAD) we therefore unrealistically assume that the user interprets the information only in the limited number of ways the designers have been able to imagine. Designers are therefore forced to think of the user as a finite automaton and therefore think of the user as a finite automaton. A good example of that is the graphical user interfaces, which basically serve to restrict the user's degrees of freedom.

This means that the starting point to think of the machine as a finite automaton willy-nilly forces the designer to think of the user as a finite automaton as well because there is no other conceptualization or model of the user that will fit the requirements of the design. Yet this is clearly unacceptable, no matter how sophisticated we assume the automaton to be (even if the number of states is exceedingly large and the state transitions are stochastic or multi-valued). It probably is the case that whatever analogy we use for one part of a Human-Machine System must also be used for the other (man or machine as a kind of forced anthropomorphism). Because we want to retain some distinct human elements in the description of the user we are forced by this argument to apply the same elements to the description of the machine (a kind of forced anthropomorphism). One proposal for an alternative description is the notion of a cognitive system as defined by Hollnagel & Woods (1983, p. 345):

“A cognitive system produces intelligent action, that is, its behaviour is goal oriented, based on symbol manipulation and uses knowledge of the world (heuristic knowledge) for guidance Furthermore a cognitive system is adaptive and able to view a problem in more than one way. A cognitive system operates using knowledge about itself and the environment in the sense that it is able to *plan* and *modify* its actions on the basis of that knowledge. (Hollnagel & Woods, 1983, p. 345)

Design is telling stories about the future

A further concern with design as a solution, is that design tells stories about the future (Roesler, et al., (2001). Design is about shaping something, an artefact, a work process, a task or job, a machine, or a tool, that does not yet exist, in particular to shape how it is going to be used or how it should be used to perform as intended. To do so it is necessary to make assumptions, not only about Work-as-Imagined. But also about the larger context or World-as-imagined. And telling stories about the future invariably changes the future, including how working conditions will be, and how people respond to the artefact (Merton,1936). What kind of adjustments and trade-offs will people employ to ensure the artefact functions as they expect sometimes in spite of the way it has been designed. Designing an artefact cannot avoid also being Cognitive Task Design (Hollnagel, 2003). Telling stories about the future again highlights the potential and actual differences between Work-as-Imagined and Work-as-Done. In summary, the irony of design is that it is telling stories about the future which thereby affects or changes the future. Changing the premises for design decisions introduces a recursion that creates very difficult problems, as recursion usually does. The specific problem here is that it proposes improvements to a situation and set of working conditions that are imprecisely known. If design fails as an effective tool, another problem to be solved is the difference between Work-as-Imagined (WAI) and Work-as Done (WAD).

- There is first of all a need to consider and describe work as it should be done during the design of machines and tools, not least since human-computer interaction nearly always is part of it. It may be easier to describe for a robot.
- There is also a practical need to think about how work should be done as part of managing and scheduling actual work under real conditions, for instance, to ensure that enough people with the requisite competencies are available when needed to meet the expectations of employers, customers and clients, and the needs of the task.
- And there is finally a need to think about how work should have been done when events are being analysed post hoc — which usually means some kind of accident investigation.

Unfortunately, this is often reduced to inventing stories about the past or proposing explanations in terms of counterfactual conditionals — such as “if only they had done X, then Y would not have happened, or if only we had had more of, or Z (safety culture/situation awareness/trust /communication, then Y would not have happened”. From a scientific and practical perspective it is regrettable that the need to explain and understand WAD is so obvious and sometimes even mandatory, when something has gone wrong , while it is practically non-existent when “nothing“ happened (the dynamic non-events proposed byWeick (1987) because everything and everyone just worked as it/they should. (Hollnagel, Laursen & Sørensen, 2022)

Conclusions about design as a solution

Is the irreducible difference between WAD and WAI really irreducible? The consequence of that is the designed systems function not because of the design, but because of human performance variability that compensates for design deficiencies. The irony of human factors and even more of safety science (Hollnagel, 2013) is eliminating human performance variability if it actually was possible would not solve the problems of imprecision, variability and lack of speed, but only make them larger, and confront us with the fact that we cannot do without that the human factor which we so desperately want to get rid of it.

Conclusions

Human factors engineering from the beginning relied on three main tools namely, training, design, and automation and each of these harbor at least one irony. In her (1983) paper Lisanne Bainbridge presented and discussed three ways, called ironies, in which automation of industrial processes could expand rather than reduce the problems of inefficiency and insufficient reliability. For each of these three ironies the outcome may at times be so different from what was intended, that any hoped for improvements never arrive.

Can we ever imagine how work is done?

Returning to the question above, the answer is the typical human factors reply of “Yes, but ...” The answer is on the one hand affirmative, because we certainly can imagine how work is (to be) done if we try, especially if we pay attention to what actually happens instead of relying on what we imagine or wish should happen or counterfactual hypotheses of what should have happened. On the other hand,the provisory “but” signifies that we should not expect ever to achieve a perfect match. The solution is neither to force WAD to comply with WAI — as in the Zero Accident Vision and (Zwetsloot et al., 2001) and Lean — nor to constrain WAI so that it corresponds to WAD as Scientific Management unsuccessfully tried to do. Work-as-Done is a moving target because internal and external working conditions (demands, and resources) rarely are stable or predictable. The solution is rather to try to understand what determines how work is done and to find effective ways of managing that to keep the variability of WAD within acceptable limits. (but not by using constraints and compliance to make people “shorter”).

not by constraints and compliance, by making people “shorter”.

Performance variability as a liability

Human performance variability, individually and/or collectively is necessary to adjust the ways and means to fit the current conditions as they actually are, and not as they have been imagined at design-time. Yet it is normally described by negatively loaded terms for example, as deviations, violations or non-compliance (of which there are several types, cf., Hudson et al., 2008)., or even as just the utterly useless “human error” (Le Coze, 2022). The negative descriptions of performance variability conveniently disregards that it can be found at both the sharp-end and the blunt-end, and that resorting to such categories usually means that the analyst has made an efficiency-thoroughness trade-off. There is a further small irony here, namely that performance variability by management usually is seen as positive (at least by management itself), and called flexibility or innovation, even though it by its nature does not differ from the performance variability at the sharp end.

Performance variability as an asset

The greatest problem in describing performance variability as an asset is that there are few if any terms that can be used to do it. We are blessed with nearly countless error taxonomies, but have practically no words or terms to characterise and therefore talk about that which goes well, the dynamic non-events that Weick (1987)referred to. The most notable exceptions are the concept of efficiency-thoroughness trade-off (ETTO, Hollnagel, 2009). Another candidate term is serendipity (Merton & Barber, 2011)

The Ironies of automation

In the 1983 paper Bainbridge described three ironies of automation

- The first irony is that the more advanced a control system is, the more crucial the contribution of the human operator will be
- The first irony is that the designer who tries to eliminate the operator still leaves the operator to do the tasks which the designer cannot think how to automate. This is another consequence of the limitations of Work-as- Imagined
- The third irony is that the automatic control system has been put in because it can do the job better than the operator, but yet the operator is being asked to monitor that it is working effectively

A further issue of automation that Bainbridge did not address at the time, because it had not yet been recognised as a problem, is due to **the substitution principle.**

The Fourth irony of automation

The substitution principle expresses the common assumption that artefacts are neutral in their effects and that their introduction into a system therefore only has intended and no unintended consequences. The basis for this principle is the concept of interchangeability, which of course has proved its value as the basis for large scale industrialisation. Thus if there are a number of identical parts, such as light bulbs or pumps, it is possible to replace one by another without unwanted side-effects. In general, however, substitutability only works when parts are not interacting and when there is no appreciable tear and wear. If parts are interacting, they constitute a system with dependencies, which almost by definition invalidates the substitution assumption.

Automation was seen as a possible replacement for humans because it is based on components with known failure rates. Humans, on the other hand, are generally seen as fallible and unreliable, as 'proved' by countless examples of 'human error'. The fallacy of this argument should by now be so obvious that it hardly needs to be belaboured. The substitution principle expresses the common assumption that artefacts are neutral in their effects and that their introduction into a system therefore only has intended and no unintended consequences. The basis for this principle is the concept of interchangeability, which of course has proved its value as the basis for large scale industrialisation (Womack, Jones & Roos, 1990) Thus, if there are a number of identical parts, such as light bulbs or pumps, it is possible to replace one by another without unwanted side-effects. In general, however, substitutability only works when parts are not interacting and when there is no appreciable tear and wear. If parts are interacting, they constitute a system with dependencies, which almost by definition invalidates the substitution assumption.

If the substitution principle is dubious in the case of technological components, it is even more suspect in the case of a substitution of functionality of which the extreme is replacing humans by automation. Carroll & Campbell (1988, p. 4), for instance, noted that

> "(n)ew tools alter the tasks for which they were designed, indeed alter the situations in which the tasks occur and even the conditions that cause people to want to engage in the tasks"

similar comments were made by (Sarter, Woods & Billings, 1997) and (Hollnagel, 2003). In relation to both risk reduction and function allocation, a substitution of functions fundamentally changes the basis for both. It is consequently not warranted to claim that the substitution is value neutral (Besnard & Hollnagel, 2014) unless both short and long term consequences of the change are fully taken into account.

When training is unable to meet its stipulated purpose, namely to ensure that human performance, or what the human operator is required to do corresponds exactly to the requirements of the technology of the human-machine system, then the logical alternative is to design the technology so that the requirements correspond to how people naturally perform, requiring neither too much, nor too little. This is why **design** became the second main element to ensure effective human-machine interaction (HMI) and the second tool in the human factors' toolbox. cf. (Hollnagel,1993)

The design of HCI is based on a number of assumptions. Some of these are explicitly stated in design guidelines. Others are hidden in the design and possibly concealed even for the designer. A particular instance is the assumptions about the reliability of human performance --- and cognition --- and about which things can go wrong. It is important for designers of HCI to know more about human reliability, regardless of whether the HCI is for industrial or academic applications. Lack of knowledge may deceive designers to rely on their personal experience or intuition. That is, however, not a valid basis for the design and the resulting system is therefore likely to be inadequate. This caution is pertinent for human reliability as well as for the, unfortunately, many other aspects of man-machine interaction that do not receive the attention they rightly deserve.

The design of Human-Computer Interaction (HCI) is based on a number of assumptions. Some of these are explicitly stated in design guidelines. Others are hidden in the design and possibly concealed even for the designer. A particular instance is the assumptions about the reliability of human performance – and cognition – and about which things can go wrong (Swain and Guttman, 1983), (Cojazzi, Pedrali & Cacciabue, 1993), (Kirwan, 1994), (Hollnagel, 1998) and (Boring, 2012). It is important for designers of HCI to know more about human reliability, regardless of whether the HCI is for industrial or academic applications. Lack of knowledge may tempt designers to rely on their imagination and/or personal experience. That is, however, unlikely to be a valid basis for the design and the resulting system is therefore likely to be inadequate. This caution is pertinent for human reliability as well as for the many other aspects of man-machine interaction that do not receive the attention they rightly deserve.

Thus the fourth irony is that the a simple replacement of the human operator with automation fundamentally changes the human-machine system. and the work that the operator does, hence the premise for what should be automated. The fourth irony is therefore a paradox and as such far more serious than an irony.

Performance variability or performance agility?

The term performance variability was introduced partly to avoid the negative connotations of deviance , violations, and non-compliance. But even variability is tainted, since it clashes with the ideal of steady machine-like functioning. If something is variable it is also not completely predictable, which is not a desired quality. An alternative to performance variability could be agility: the state or quality of being agile.It is human agility that allows people to overcome the temporary glitches and problems that every tool use involves

Performance agility as putty?

The ideal way to build a house, a bridge or a wall, is to make sure that every element or stone fits exactly to its neighbour. This can be seen for instance from the Pyramids, Roman bridges, or the walls around Japanese castles, in e.g., in Osaka or Himeji. But this a slow process that requires enormous work and infinite care.The alternative to use in stones that each is carefully shaped to fit with neighbouring stones, is to use prefabricated elements that do not necessarily fit exactly, such as the bricks used to build houses today, and then to fill in the gaps with putty. In this case the components only have to fit reasonably well to each other, but not precisely so. The putty will cover any gaps and help to provide the appearance of a whole, especially when some time has passed. Since it is impossible to fit humans exactly to match the machines, lest we resort to the Procrustean solution, and also impossible to fit the machines exactly to the individual human tailor made tools like tailor made clothes are far too expensive. Instead some kind of putty is needed to ensure that the human-machine system works. In the case of human factors the putty is human agility or performance variability. The deep irony of human factors is that we attempt to remove this putty, because it mistakenly has been seen as a liability rather than the asset it really is. There are interesting historical reasons why this is so, as this paper has tried to explain. But there are even more compelling reasons to stop doing so.

When WAI is compared to WAD, there is always a small gap or discrepancy. The two never match completely (Hollnagel, 2017). So just as when bricks are put together to build a wall or a house, something is needed to fill out and cover the gaps. The alternative would be to ensure that the bricks fit perfectly The design of HCI has generally suffered from the lack of a methodology or discipline corresponding to what can be found e.g. in the field of software engineering, although some attempts have been made (e.g. Dowell & Long, 1989). Design can roughly be defined as the art (or skill) of matching constraints with possibilities, as shown in Figure 1. If there are no constraints, the design can be anything that the system or the technological options allow. If there are too many constraints, then possibly the system capabilities are insufficient, i.e., there are no options that match the constraints. And if either constraints or possibilities are ill-defined, design becomes an unwieldy process.

The designer's problem can be considered in the situation where there are two (or more) options available to solve a specific problem, e.g. displaying information about the state of a sub-system. If both of these options match the requirements (e.g. with regard to information requirements or ease of diagnosis) and constraints (e.g. with regard to maintaining safety or control), then both alternatives are equally good. However, the designer would usually prefer that there was a further constraint or requirement, so that a choice could be made between the two options. One often used solution to that problem is design guidelines, which can assist the designer in the following ways:

whatever the mode of operation. because it is the conceptual putty that fills out the cracks between WAI and WAD. We have been too obsessed with the old view of humans as a liability and source of error and variability. The truth is that without this variability, no system would be able to function. It is the putty for the inevitable discrepancy between WAI and WAD !

If neither training, nor design can ensure that the human becomes the required and desired reliable component in the workplace,then the only remaining solution, and also the third tool in the human factors toolbox is automation, i.e., having a (reliable) machine take over part of the human's work and activities. In this case the ironies have already been described by Bainbridge (1983), and can therefore simply be summarised here:

Conclusions

The irony of training is

The irony of design is

the ironies of automation are, and the overarching irony of human factors is

The irony of human factors is that by trying to eliminate or at least neutralise the human factor, meaning the variability of human performance we expand the problem we try to solve. Because human agility is the putty that hides the cracks between Work - as Imagined and Work - as Done. The irony of that is that human factors are doing themselves a disservice. It is reducing and if possible eliminating human performance agility it is eliminating the quality that is necessary for human-machine systems to work.

Conclusions

This essay has provided a retrospective overview of the role of human factors in safety. Human factors became a recognised problem after then end of the Second World War, partly as a consequence of the many hitherto unknown types of UUOs that the War provided, and partly as a consequence of the rapid technological advances brought about by the War, and eagerly transferred into civilian applications soon after.Humans played a larger and more important role in the new technology-driven systems, and since humans can only develop at an evolutionary pace, they soon were seen as inefficient and unreliable compared to the technological artefatcs. In that sense WW-II also became the first illustration o f the WAI-WAD discrepancy.Since the human factor was perceived as a problem, solutions have to be found, following the safety legacy reasoning, that problems could be solved by identifying and eliminating their causes. The cause was in all cases the perceived inefficiency and unreliability of the humans at the sharp ends of the systems, and human factors proposed three different solutions: training, design, and automation. Unfortunately none of these were really effective as this essay has demonstrated.

- **Training** was an imperfect solution because it represented the procrustean approach of fitting the human to the machine and the demands of the technology.
- Although **design** tried the opposite, namely to fit the machine and technology to the human, it was also an imperfect solution.
- And so finally was **automation**, because of the three ironies that Lisanne Bainbridge hertil identified:First,. Second, and third,

Fitts List today.

It is interesting to consider what a Fitts List would look like today, if anyone would dare use the same principle of comparison. The original, there is surprisingly little that must be changed, which is remarkable, because digital computers were practically unknown when Fitts produced the List.The ENIAC had only started to operate in 1945, and it was a mainframe it contained 18,000 vacuum tubes, 7,200 crystal diodes, 6,000 relays, 70,000 resistors, was roughly 10 ft (3 m) tall, 3 ft (1 m) deep, and 100 ft (30 m) long, and occupied 300 sq ft (28 m^2) and far less powerful than a modernd ay PC, or even a smartphone. Fitts' assessment of the machine's capabilities is therefore nothing short of impressive.

Table 5: Details of the Fitts List (1951).

Criterion	Explanation, definition.	
	Humans surpass machines in terms of	Machines surpass humans in terms of
1 (Detection)	*Ability to detect small amounts of visual and acoustic energy*	Ability to respond quickly to control signals and to apply great force smoothly and precisely.
2	Ability to perceive patterns of light and sound.	Ability to perform repetitive routine tasks.
3	Ability to improvise and use flexible procedures.	
4	Ability to store very large amounts of information for long periods of time and to recall relevant facts at the appropriate time.	Ability to store information briefly, and then to erase it completely.
5	Ability to reason inductively	Ability to reason deductively, including computational ability
6	Ability to exercise judgment	Ability to handle highly complex operations i.e., to do many different things at once.

References

Arbib, M. A. (1987). *Brains, machines and mathematics*. New York: McGraw-Hill.

Ashby, W. R. (1960) Design for a Brain. The origin of adaptive behavior. (Second Edition). London: Chapman & Hall, Ltd.

Bainbridge, L. (1983). Ironies of automation *Automatica*, *19*(6) pp. 775 779,

Besnard, D., & Hollnagel, E. (2014). I want to believe: some myths about the management of industrial safety. *Cognition, Technology & Work, 16*, 13-23.

Bloch, A. (2003). Murphy's law. Harmondsworth UK: Penguin.

Boole, G. (1854). *An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities (Vol. 2)*. London, UK: Walton and Maberly.

Booth, P. (1989). An introduction to human-computer interaction. London: Erlbaum.

Boring, R. L. (2012). Fifty years of THERP and human reliability analysis INL/ CON-12-25623).Idaho Falls, ID; IdahoNational Lab.(INL), (United States).

Carroll, J.M., Campbell, R.L. (1988). Artefacts as Psychological Theories: The Case of Human–Computer Interaction. User Interface Institute, IBM T.J. Watson Research Centre, Yorktown Heights, NY.

Cojazzi, G., Pedrali, M. & Cacciabue, P. C. (1993). Human performance study: Paradigms of human behaviour and error taxonomies ISEI/TE/2443/93). JRC Ispra, Italy: Institute for Systems Engineering and Informatics.

Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human—automation co-ordination. Cognition, Technology ¢» Work, 4, 240-244.

Dekker, S. W., & Woods, D. D. (1999). To intervene or not to intervene: The dilemma of management by exception. Cognition, Technology & Work, 1, 86-96.

Dowell, J. & Long, J. (1989). Towards a conception for an engineering discipline of human factors. Ergonomics, 32(11), 1513-1535.

Dul, J. Et al.(2012): A strategy for human factors/ergonomics: developing the discipline and profession, *Ergonomics*, DOI:10.1080/00140139.2012.661087

Dubois, D. & Prade, H. (1989). Handling uncertainty in expert systems – pitfalls, difficulties, remedies. In E. Hollnagel (Ed.) The reliability of expert systems. Chichester: Ellis Horwood Ltd. (64 – 99).

Hall, A. D. & Fagen, R. E. (1968). Definition of system. In W. Buckley (Ed.), *Modern systems research for the* *behavioural scientist*. Chicago: Aldine Publishing Company

Feigenbaum, E. A., & Feldman, J. (1963). Computers and thought (Vol. 7). McGraw-Hill: New York.

Fitts, P. M. et al. (1951). Human engineering for an effective air-navigation and traffic-control system. Washington D. C: National Research Council

Hirschhorn L. (1997). Quoted in R.I. Cook, D.D. Woods and C. Miller, (Eds.), (1998). *A Tale of Two Stories: Contrasting Views on Patient Safety*. Chicago IL: National Patient Safety Foundation.

Hollnagel, E. (1983). The design of reliable HCI: The hunt for hidden assumptions. In J. L. Alty, D. Diaper & S. Guest (Eds.), *People and computers VIII*. Cambridge: Cambridge University Press.

Hollnagel, E. (1992). The art of efficient man-machine interaction: Improving the coupling between man and machine. Invited paper for CADES workshop, St. Prix, France, August 1992. (Also in: J.-M. Hoc, P. C. Cacciabue & E. Hollnagel (Eds.), *Expertise and technology: Cognition & human-computer cooperation*. Lawrence Erlbaum.

Hollnagel, E. (1998) *Cognitive Reliability and Error Analysis Method (CREAM)*. Oxford: Elsevier Science Ltd.

Hollnagel, E. (2013). Is safety a subject for science? *Safety Science*, *67*, 21-24.

Hollnagel, E. (Ed.), (2003). *Handbook of cognitive task design*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Hollnagel, E.(2009). *The ETTO principle: Why things that go right sometimes go wrong*. Farnham, UK: Ashgate.

Hollnagel. E (2017) Can we ever imagine how work is done? *HindSight 25* 10-13

Hollnagel, E. (2025). *From Safety to Safely: Principles and practice of Systemic Potentials Monitoring*. Abingdon, Oxon, UK: Routledge.

Hudson, P. T. W. et al. (2008). Meeting expectations: A new model for a just and fair culture. In SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production. OnePetro.

Jastrzębowski, W. (1857). Rys ergonomiji czyli Nauki o Pracy, opartej naprawdach poczerpnietych z Nauki Przyrody [An outline of ergonomics or the science of work based on the truths drawn from the science of nature]. *Przyoda i Przemysl*, *29*, 227-231

Kirwan, B. (1994).A practical guide to human reliability assessment. London: Taylor & Francis.

Laird, J., Newell, A. & Rosenbloom, P. (1987) SOAR: An architecture for general intelligence. Artificial Intelligence 33:1-64.

Le Coze, J. C. (2022). The 'new view' of human error. Origins, ambiguities, successes and critiques. Safety science, 154, 105853.

Lindsay, P. H. & Norman, D. A. (1972). Human information processing, An introduction to psychology. NewYork:Academic Press.Lerner, A. Y. (1975). Fundamentals of cybernetics. London: Plenum.

Meister, D. & Enderwick, T. P. (2001). Human Factors in System Design, Development, and Testing. CRC Press, Inc.

Merton, R. K. (1936). The unanticipated consequences of purposive social action. American sociological review,1(6), 894-904.

Merton, R. K., & Barber, E. (2011). The Travels and Adventures of Serendipity. Princeton University Press.

Moehlenbrink, C., & Papenfuss, A. (2011, September). ATC-monitoring when one controller operates two airports: Research for remote tower centres. In P*roceedings of the Human Factors and Ergonomics Society Annual Meeting 55*(1), pp. 76-80). Sage CA: Los Angeles, CA: Sage Publications.

Morozow, E. (2013a). *The perils of perfection*. The New York Times, 2 March 2013.

Motozow, E (2013b). *To save everything, click here*. New York: The Perseus books group

Newell, A. (1990). *Unified theory of cognition*. Cambridge, MA.: Harvard University Press.

Newell, A. & Simon, H. A. (1963). GPS, A program that simulates human thought. In: E. A. Feigenbaum & J. Feldman (Eds.) *Computers and thought*, 279-293. New York: McGraw-Hill Book Company.

Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nyssen, A. S., & Blavier, A. (2019). Investigating expertise, flexibility and resilience in socio-technical environments: a case study in robotic surgery. InE. Hollnagel, J. Braithwaite & R. L. Wears (Eds.) *Resilient health care* (pp. 97-110). CRC Press.

O'Sullivan, F.(2021). *Domino theory and U.S. foreign policy*. https://historyguild.org/domino-theory-and-U.S.foreign-policy/(accessed December 1, 2024).

Papenfuss, A., & Friedrich, M. (2016, September). Head up only—a design concept to enable multiple remote tower operations. In 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC) (pp. 1-10). IEEE.

Paxton, L. J. (2007). “Faster, better, and cheaper at NASA: Lessons learned in managing and accepting risk. *Acta Astronautica, 61*(10), 954-963.

Reason, J. T. (1990). Human error. Cambridge university press.

Roesler, A., et al., (2001). Design is telling stories about the future.Columbus, OH: Cognitive Systems Engineering Laboratory, Ohio State University,.

Sarter, N.B., Woods, D.D., Billings, C.E., 1997. Automation surprises. In: Salvendy, G. (Ed.), Handbook of Human Factors and Ergonomics, 2ndEd. New York: Wiley.

Sheridan, T. B. (1982). *Supervisory control: Problems, theory, and experiment for application to human-computer interaction in undersea remote systems*. Boston, MA., MIT. Dept. of Mechanical Engineering.

Shorrock. S. (2020, October 28). Proxies for work-as-done: 1. Work-as-imagined. Humanistic Systems.https://humanisticsystems.com/2020/10/28/proxies-for-work-as-done-1-work-as-imagined/

Simon, H. A. (1972). The sciences of the artificial. Cambridge, MA.: The M. I. T. Press.

Stassen, H. G. (1986). Decision demands and task requirements in work environments: What can be learned from human operator modeling. In E. Hollnagel, G. Mancini & D. D. Woods (Eds.), Intelligent decision support in process environments. Berlin: Springer Verlag.

Strauch, B. (2017). Ironies of automation: Still unresolved after all these years. IEEE Transactions on Human-Machine Systems, 48(5), 419-433.

Swain, A. D. & Guttman, H. E. (1983). Handbook of human reliability analysis with emphasis on nuclear power plant applications (NUREG CR-1278). Washington, DC: NRC.

Swain, A. D. & Guttman, H. E. (1983). Handbook of human reliability analysis with emphasis on nuclear power plant applications (NUREG CR-1278). Washington, DC: NRC.

Taylor E W.(1919). *The principles of scientific management*. New York: Harper & Brothers Publishers

Taylor, F V., & Garvey, W. D. (1959). The limitations of a ‘Procrustean’ approach to the optimization of man-machine systems. Ergonomics, 2(2), 187-194.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. J. of Math, 58(345-363), 5.

Turing, A. M. (1960). Computing machinery and intelligence. *Mind*, *LIX* (236),433-460.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. IEEE Transactions on systems, man, and cybernetics, 22(4), 589-606.

Von Főrster, H., & Pőrksen, B. (2002). Understanding systems: Conversations on epistemology and ethics. Heidelberg, Germany: Carl-Auer Verlag.

von Neumann, J. (1958). The computer and the brain. Yale University Press

Weick, K. E. 1987. Organizational culture as a source of high reliability. California Management Review 29(2), 112-128.

Weizenbaum, J. (1976). Computer power and human reason. From judgment to calculation. San Francisco: W. H. Freeman

Wiener, N. (1966). God & Golem, Inc.: A comment on certain points where cybernetics impinges on religion (Vol. 42). MIT press.

Wienet, N. (1988). The human use of human beings: Cybernetics and society (No. 320). Da Capo Press.

Womack J, Jones D, Roos D. The machine that changed the world. New York: Simon & Schuster; 1990.

Zadeh, L. A. (2014). A note on modal logic and possibility theory. Information Sciences, 279, 908-913.

Zwetsloot, G. I. J. M., et al., (2013). The case for research into the zero accident vision. *Safety Science, 58*, 41-48.

making longer: They do less than they should do, hence may render the system incapable of functioning as planned (e.g. default actions, wrong procedure, too little and too late). The fourth irony is that we try to limit or eliminate the variability of human performance that is needed for systems to function as they were intended to — or at all. This is an irony of automation, but also of human factors and HSE management in general. The ETTO-principle only referenced

The irony is that by trying to constrain human performance and introducing barriers, you actually increase variability rather than reduce or limit it! The effect is therefore the opposite of what was intended, this is more than an irony it is a paradox

Bainbridge on performance variability;

Long-term knowledge . An operator who finds out how to control the plant for himself, without explicit training, uses a set of propositions about possible process behavior, from which he generates strategies to try (e.g, Bainbridge. 1981). Similarly an operator will only be able to generate successful new strategies for unusual situations if he has an adequate knowledge of the process. There are two problems with this for machine-minding operators. One is that efficient retrieval of knowledge from long-term memory depends on frequency of use (consider any subject which you passed an examination at school and have not thought about since). The other is that this type of knowledge

Performance variability as putty

When WAI is compared to WAD, there is always a small gap or discrepancy. The two never match completely (Hollnagel, 2017). So just as when bricks are put together to build a wall or a house, something is needed to fill out and cover the gaps. The alternative would be to ensure that the bricks fit perfectly together like Lego blocks or the stones in the walls of the pyramids, Roman bridges or Japanese castles. But it is far easier to use putty to fill out the gaps and in this way disguise the differences. In the context of human-machine systems, human performance variability plays the role of putty and helps to ensure that the wall works as it should and looks complete particularly when it has been in place for some years.

Conclusion: The irony of human factors

As this paper hopefully has demonstrated,human factors as such therefore comprise five ironies, one for each of training and design plus three for automation. The overall irony for human factors is that it considers human performance variability as a liability and, even though it is an asset without which few systems would work.

Wiener, N. (1964). *God & Golem, Inc.: A comment on certain points where cybernetics impinges on religion (Vol. 42*).Boston, MA: The MIT press.