International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 244-253, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.015

The Question of Analyzing System Safety with Consideration to Human Factor

Iosif Z. Aronov
Department of Commerce and Trade Regulation,
MGIMO (Moscow State Institute of International Relations) University, Moscow, Russia.
E-mail: izaronov@itandi.ru

Anna M. Rybakova
Department of Commerce and Trade Regulation,
MGIMO (Moscow State Institute of International Relations) University, Moscow, Russia.
E-mail: amrybakova@itandi.ru

Nataliia M. Galkina
Department of Trade Barriers Analysis,
International Trade and Integration (ITI) Research Center, Moscow, Russia.
Corresponding author: nmgalkina@itandi.ru

(Received December 30, 2019; Accepted April 24, 2020)

Abstract
This article presents the research results of the human factor effect on system safety. The authors used historical data on inductive reasoning to obtain the results. The article shows that there is no good reason to believe that there is no human factor that the human factor has no effect on system safety. In other words, it is considered that most disasters and incidents were initiated by human beings, the so-called, human factor. It is concluded that inductive reasoning plays a key role in the human factor, which leads to disasters and incidents.

Keywords- Safety, Human factor, Safety culture, System safety, A harbinger of accidents.

1. Introduction and Statement of the Problem

Each man-made disaster proceeds in its own unique scenario. However, the analysis allows to detect some common reasons, which, as a rule, cause these disasters. This common reason is a human being.

According to the statistics (Aronov and Papic, 2018; Aronov, 2019), human beings are the “weak link” of system safety. For example, motor vehicles lead to incidents up to 90%, railway transport - about 50%; aviation and water transport - up to 83%, about 75% of sea incidents are related to the human factor.

Long-standing experience of operating power-generating equipment has shown that most disasters and incidents were initiated by human beings’ behavior, as well as their attitude to responsibilities and safety. According to individual assessments, in providing energy reliability and safety, human errors cause more than 80% of incidents and technological disasters (Aronov, 2019).

According to the American researcher Davis (2002), the so-called "human factor" of technogenic disasters is limited almost entirely to such human qualities as stupidity, neglect and greed.

Articles (Kopachevsky et al., 2016; Singh et al., 2020) have also highlighted the role of the "human factor" in assurance reliability.

The root cause analysis of this phenomenon is the purpose of this article.

2. Inductive Reasoning

Despite the efforts of specialists in the safety area, there is one fundamental reason that limits the opportunities of such specialists. It is formulated as a philosophical principle of narrow-mindedness, which is inherent to all human beings when there is a shift from fragmented towards a common approach (Taleb, 2008). Every specialist faces this situation when dealing with information on previous disasters or incidents with a reference to a current situation. From the fact that there has been no disaster or incident for a considerable period of time, it doesn't mean that it will not happen in the following period.

The majority of specialists in the safety area usually forget about it. And this is what the inductive reasoning means. The historical record is peppered with such examples. Take the example of the historical disaster, which was caused by inductive reasoning. The Football Cup Semi-Final between Liverpool and Nottingham Forest was played at Hillsborough on 15 April 1989. More than 96 people were killed, and more than 600 people were injured from the stampede. A commission headed by Lord Justice Peter Taylor was charged by Parliament to investigate the disaster. In a Report to Parliament Taylor (1989), he said: "I can give you two good reasons. First: craving for deterrence against bullies. The second- *complacency* that enabled the responsible officials to think that *if before nothing bad happened, it is not going to happen this time*".

Here's another historical example. On 14 April 1912 the most modern as of that time passenger liner Titanic received several ice warnings from various ships in the North Atlantic, but despite this, the captain didn't reduce the speed. When an iceberg was detected, the Titanic could not quickly change the course due to high speed. At 2:20 am on April 15 Titanic went completely underwater. The disaster took between 1400 and 1600 lives. The message is simple: the captain's confidence in the unsinkable passenger liner Titanic led to a disaster.

Generally, complacency and confidence almost always accompany inductive reasoning.

Inductive reasoning affects the entire system of system safety. What is the reason? Let's imagine a potentially dangerous object that has N hazardous factors, where the implementation of each can lead to an incident during system operation. In fact, it is important to consider that, in practice, these causes can occur either on their own or in various combinations. The total number of such combinations are easily seen as equal to 2^N. And this is quite a slump number, even if N is insignificant. For example, according to recent data (Vasilyev and Salnikov, 2014) about 27 reasons can lead to building's and construction's incidents. However, considering that incidents often arise at an unfavorable combination of various defects, theoretically, at the safety analysis it would be necessary to consider already $2^{27} = 134217728$ combinations of factors that is delusory.

In practice, not all causes can occur simultaneously, but even the exclusion of extraordinary combinations of causes leaves a big number of possible incident causes that need to be analyzed.
How does this analysis go? First, the most probable scenarios of incidents should be analyzed. This is obvious because, as mentioned above, it is impossible to analyze all possible incident scenarios due to various combinations of reasons. The basis of any safety analysis is the information on the

credibility of system elements contained in the corresponding databases (DB). If the information in the database is accurate, then it can be assumed that the results of the analysis are correct, and if the data is "fake”, then how should the results of the analysis be qualified? In this relation, the American physicist and mathematician of Hungarian origin Cornelius Lanczos said that "the lack of information cannot be filled with any mathematical tricks."

What happens to the analysis of low-probability incident scenarios? They are waiting for “better” days. And, unfortunately, these days are coming. The reports of the International Atomic Energy Agency (IAEA 2015a, 2015b) devoted to the analysis of incident causes on the nuclear power station Fukushima Daiichi (Japan), provide that when the station was constructed, the design engineers proceeded based on the analysis of historical data for 400 years (!): the force of earthquakes in this prefecture has never exceeded 7 points, the maximum tsunami wave height was 3 m. However, an earthquake of magnitude 9 and a tsunami wave above 15 m had to be encountered during operation, as it happened.

The technical system design engineers also face the problem of inductive reasoning limitations when applying the Failure Mode and Effects Analysis (FMEA) and Failure Modes, Effects and Criticality Analysis (FMECA) methodologies postulating the possible initiating events of potential accidents *F1, F2,…,Fn*, to offer then the necessary means of parry. What is the basis for the indicated list? As a rule, it is information about previous failures of similar objects. But is it possible to assert that this list is exhaustive? Of course, it is not possible. In that respect English writer Douglas Adams made a caustic statement: “A common mistake that people make when trying to design something completely foolproof is to underestimate the ingenuity of complete fools.”

Thus, design engineers of systems should remember about the limitation of such approach and not to be indulged in complacency, believing that the new methods of safety analysis in conjunction with modern computer systems guarantee reliability and safety of designed systems.

And there are other factors caused by inductive reasoning. As a simple example (Alexandrovskaya et al., 2001), an assessment was made regarding the degree of reliability reduction (probability of non-failure operation) of the system safety, consisting of two similar sensors, which transmit a signal to the actuators. In terms of reliability sensors are connected in parallel. The data about the probability of non-failure operation of the sensors of this type are contained in the database.

Suppose that the probability of failure of each *Pi* sensor defined by the database is equal to 0.999; i = 1.2. That is the probability of a failure-free operation will be defined as

$$P_{c1} = 1- (1- P_1) (1- P_2) = 0.999999 \quad (1)$$

In the designer engineer opinion, considering the high reliability of this system, the scenario of the accident caused by the failure of this system safety, cannot be considered.

However, let it be clear that two sensors of the same type are used in this system. Therefore, under certain conditions, the system may fail as a result of sensor failure due to a common cause, such as a manufacturing defect peculiar to both products. In this case, the failure-free rate of such system must consider the probability of failure due to a common cause. Suppose that this probability is Q $=10^{-3}$. Then the probability of a system failure will be equal to

$$P_{c2}= [1- (1- P_1) (1- P_2)] (1-Q) = 0.9989 \qquad (2)$$

By comparing the values of Pc1 and Pc2, it can be concluded that due to a common cause of failure, the positive safety rate has decreased by three (!) times. In the considered situation, the exclusion of an incident scenario caused by system safety failure is incorrect.

With that said, the inductive reasoning of designer engineers led to negative consequences. The problem was caused largely due to the "faith" of the specialists in the database information and negligence to the system approach, which to some extent can counterbalance inductive reasoning. Consider another situation related to inductive reasoning. Famous historian and philosopher (Harari, 2018) said: “Most of our views are shaped by communal groupthink rather than individual rationality, and we cling to these views because of group loyalty.” Therefore, what is a “groupthink”?

Groupthink is the willingness of group members to decide in the group without critically assessing each of the presented alternatives. The pioneer of the group thinking analysis phenomenon was an American psychologist (Irving, 1982). He emphasized, among the other eight signs, the illusion of invulnerability. This means that group members due to their privileged position develop inaccurate model about the degree of their capabilities. Therefore, these group members are characterized by complacency and belief in their own superiority over all others. This can push such groups to make unreasonable, risky, irresponsible decisions.

For example, this approach prevailed in NASA and led to the crash of the Space Shuttle "Colombia" on 1 February 2003. Seven crew members got killed. The cause of the accident was a piece of foam insulation which broke off from the Space Shuttle external tank and struck the left wing of the orbiter. NASA management was aware of the problem but ignored the repeated statements by experts: the report (Report of Columbia Accident Investigation Board, 2003) refers to seven cases of foam insulation, which broke off and did not cause an incident. The experts stated that they would like NASA management to listen to their advice, but they do not count on it because "they are always convinced that they are right and do not want to listen to the external advice". In this case, complacency played a negative role.

3. Safety Culture

What can be opposed to inductive reasoning, which seems to be global in nature? It is obvious that inductive reasoning is something that is inherent in human nature. Russian psychologist and cultural philosopher P. Gurevich defined the concept of "human nature" in such a way - stable, invariable traits, common characteristics that express its features as a living being that are always inherent in Homo sapiens, regardless of biological evolution and historical process (Gurevich, 2011). At the same time, he stated that human being is a model of cultural conditions that shape him.

If this is the case, a safety culture can be a "tool" for shaping a "new" human being in relation to systems safety (SMM, 2013).

At this point, safety culture is a set of values, beliefs, habits of behavior inherent in employees and which reflect the security policy being implemented in the enterprise. In production activities, safety culture is first and foremost manifested in:

- the attitude of the employee to the error: to hide it or to speak frankly about what happened, being sure that the specialists will deal with it objectively;
- team relations: problems are set and solved or dismissed;
- communication between hierarchical structures: whether the voice of an employee who does not hold a high position can be heard or not;
- the attitude towards the specialist who made an error: to discipline or find the factors that triggered an error, etc.

It is primarily up to employees to identify hazards in a timely manner and take prompt action to address them. Also, the employee is always an information holder of errors or violations, as well as other officially unregistered deficiencies. Bringing this information to management allows to determine the causes and develop effective preventive measures. An employee who is in an atmosphere of fear of punishment for an error or violation will not inform anyone of any dangerous causes.

It is obvious that for every employee to be able to become an active participant of a system safety, it should be warranted that the "non-punitive" production environment has been formed at the company, where errors and violations are not penalized. This will enable employees to develop a new attitude to safety and to realize their role in addressing the issues.

However, if the safety culture is limited to reacting to an incident, it is clearly not enough to make a difference. It is important that employees actively identify safety and security concerns and focus on continuous research on risk factors.

The authors of the research (Reiman and Rollenhagen, 2013) believe that the safety culture represents a “systematic view” on safety, which is rarely explicitly stated. This article argues that the "new" contribution to safety management based on a safety culture has not yet been integrated with classical engineering principles and concepts. However, such integration is necessary to develop a real system-oriented view of safety; when human, technological, organizational and cultural factors are understood as mutually interacting elements. When a safety culture is decontextualized, it is not embedded in safety technology, which violates the principle of systemacity.

It should be noted that unsuccessful motivation or incentive programmes in an enterprise may have a negative impact on safety. For example, if employees are rewarded for meeting company standards for injury indicators, they may not report incidents. This may lead to a reduction in safety within the enterprise. Therefore, incentive programmes must be carefully planned and validated before they are implemented as part of a safety culture.

There are cases when the company uses the method of anonymous observation, where each employee is obliged to inform the management about the unsafe behavior of his colleagues. This can lead to conflict among employees and concerns that safety issues are not being addressed. It is therefore crucial to encourage employees to report concerns without fear of negative consequences, as well as to encourage them to report positive actions and share ideas to improve workplace safety. Responsibility for building a safety culture begins with the company's management and extends to each employee. Everyone must protect themselves and others.

One of the most significant challenges to implementing and maintaining an effective safety culture is the constant tension between safety efforts and the need to improve productivity. Indeed, if the focus is on improving safety, it can unwittingly lead to productivity losses. It may appear that the enterprise is designed to "produce safety" rather than produce a material product, transport goods and passengers, etc. On the other hand, if safety aspects are not considered properly, it inevitably leads to incidents. Thus, there must be a reasonable balance between safety efforts and efforts to increase productivity, that is, profit. That balance is exactly what a safety culture provides.

4. Human Reliability Analysis (HRA) Issues

In the 60's of the last centuries, the methodology for analyzing the reliability of human-machine systems was rapidly developing, which led to the creation of human reliability analysis (HRA). Within HRA, researchers focused on the development and study of mathematical models of human - operator reliability. An overview of such research can be found in various research papers (Dragan and Isaic-Maniu, 2014). This approach implicitly ignored personal characteristics of operators associated with the indistinct nature of human properties: personality, health, age, education, qualification, experience, work capacity, workload, fatigue, man-made effects and other properties. For example, it is known from numerous studies that pilots' resources differ by 3-5 times, which is based on the results of their selection for a profession and subsequent professional activity (Plotnikov, 2015).

In economic-mathematical models in which the person participates, the problem of the human factor is solved by postulating that in economic system the "rational human being" always makes decisions, is guided by selfish interests (Von Neumann and Morgenstern, 1953) and that allows to neglect features of human being behaviour.

However, in reliability and safety calculations of technical systems ignoring human being variability - the operator can lead to considerable uncertainty of forecasts.

In this situation, it seems advisable to take into account possible differences in the behavior of a human being – operator (Dragan and Isaic-Maniu, 2014), which allows reducing bombshells in the operation of systems.

5. Early Predictions

When reading N.N. Taleb, it is obvious that the prediction of "black swans" is unpromising. At the same time, futurologists, who have largely absorbed Taleb's ideas, believe that one can observe weak signals or minor events that carry information about incidents (Harris and Zeisler, 2002; Barber, 2003). By analyzing these signals, some incidents can be predicted.

One major study (Lewis, 1990) claims that throughout human history, disasters have been a primary focus of attention, whereas "behind the scenes", the smaller-scale disasters (mini-incidents), are much more numerous and often more informative. Small incidents occur literally every second, and often only by happy coincidence, they do not turn into disasters. The same aspect was addressed in another research (Aleksandrovskaya et al., 2001) and several documents regulating safety, for example, the documents of the International Civil Aviation Organization (ICAO) and the International Atomic Energy Agency (IAEA). It is important to use this information for incidents prognosis.

It is reasonable to estimate the severity of each mini-incident: the higher the probability of this incident, the higher the severity of it. In the research (Aleksandrovskaya et al., 2001) this indicator is called the "rank of breakdown".

The implemented indicator makes it possible to compare the failures (mini-incidents) of the same dangerous object or failures of the same technical systems by the degree of their severity and to evaluate the effectiveness of corrective measures to improve the safety of the object. Mini-incidents with the highest rating values for some fixed period of operation are called precursors of an incident. It is the analysis of incident precursors in conditions of limited resources that allows to focus on a deeper analysis of their causes in order to identify physical characteristics and parameters of the system, which could be further used to predict major man-made incidents.

6. Looking Ahead

As stated above, a human being is a "weak link" in a complex system, which is tempted to be replaced by automation technologies such as artificial intelligence (AI). Although definitions of AI vary, for the purposes of this article we will consider AI technologies and systems to include software and/or hardware that can learn to solve complex problems, make predictions, or perform tasks that require human perception (such as vision), cognition, planning, learning, communication, or physical action.

In an interesting study by McKinsey (Chu et al., 2015) on the subject, the following conditions (criteria) for the possible replacement of a person by AI were formulated:

- Technical feasibility, which is a necessary precondition for automation;
- The cost of developing and introducing both hardware and software for automation;
- The labour cost and the associated supply and demand dynamics: if there is an oversupply by employees, and this is much cheaper than automation, then this factor may be a decisive argument against replacement;
- Higher productivity, better quality and less errors;
- Addressing the social issues of AI application, which must be considered.

In addition, it should be noted that the ability to replace is significantly affected by the environment in which the work is performed: the more unpredictable the environment, the harder it is to replace. Thus, the high reliability of AI is an important factor providing the replacement. These findings are based on a detailed analysis of more than 2000 jobs for more than 800 professions. The ethical problems associated with the AI are discussed in detail by Harari (2018). Underscoring the importance of the AI technology for the future of the U.S. economy and national security, the President issued Executive Order (EO 13859) on 11 February 2019, requiring federal agencies to take various measures to ensure that U.S. leadership in the field of AI is maintained. The document states: "It is necessary to ensure that technical standards reflect federal priorities for innovation, public understanding, and public confidence in systems using AI technology and develop international standards to promote and protect these priorities".

National Institute of Standards and Technology (NIST) plays a leading role in the standardization of AI in the United States, which coordinates U.S. government and private sector activities in this area. AI is a strategic priority for NIST, and its research program includes fundamental research to measure and improve the reliability of the AI systems, for example, by examining ways to measure the safety and transparency of the AI systems. NIST promotes the use of AI in research programs ranging from offline detection of advanced materials to robotic systems in manufacturing environments.

NIST was assigned to develop an “A Plan for Federal Engagement in Developing Technical Standards and Related Tools”. This task was completed, and the Plan was published on 10 August 2019 and is available on the NIST website.

The Plan highlights that increasing confidence in AI technologies is a key element in accelerating their adoption for economic growth and improved safety. Accuracy, reliability, safety and confidentiality are among the characteristics that ensure confidence in the AI technologies. These factors are in addition to those listed in the McKinsey study. Ideally, they should be considered early in the design process and verified in the development and use of AI technologies.

Their development has raised many legal, ethical and social issues that developers, policy makers and users should consider. The development of this Federal Plan emphasized the importance of establishing desirable principles and limitations in the development of AI standards. The principles governing the introduction of AI are being developed by several organizations, including the Organization for Economic Cooperation and Development (OECD), whose member countries have recently adopted these principles.

There are currently two areas where there is some consensus:

- The relationship of ethical requirements included in the Standards to the risk degree of causing harm to human beings;
- Confidentiality requirements to be included in any standards governing the collection, processing, exchange, storage and disposal of personal data.

Recognizing the importance of standardization in the field of AI, ISO and IEC international standardization organizations formed the ISO/IEC JTC 1/SC 42 Artificial Intelligence Technical Committee in 2017, which has developed several international standards:

ISO/IEC 20546:2019 Information technology — Big data — Overview and vocabulary

ISO/IEC TR 20547-2:2018 Information technology — Big data reference architecture – Part 2: Use cases and derived requirements

ISO/IEC TR 20547-5:2018 Information technology — Big data reference architecture – Part 5: Standards roadmap.

However, many standards are under development.

In addition, the Institute of Electrical and Electronics Engineers (IEEE) has been involved in standard development. In particular, the following standards IEEE P7000™ series should be mentioned:

P7000 - Model Process for Addressing Ethical Concerns during System Design
P7001 - Transparency of Autonomous Systems
P7002 - Data Privacy Process

P7003 - Algorithmic Bias Considerations
P7004 - Standard for Child and Student Data Governance
P7005 - Standard for Transparent Employer Data Governance
P7006 - Standard for Personal Data Artificial Intelligence (AI) Agent
P7007 - Ontological Standard for Ethically Driven Robotics and Automation Systems.

7. Conclusion

With the increasing reliability of systems and their complexity, the role of the “human factor” increases, including inductive reasoning. Overall, this kind of reasoning often dominates. Nevertheless, the "antidote" to it would be a safety culture formation. Therefore, developing a safety culture and creation of work environment “without fear” remains a crucial task of the theory and practice of systems safety.

Replacing human beings with artificial intelligence still does not solve the problem of the “human factor” completely, as the problem of human reliability is replaced by the problem of the AI reliability and complemented by ethical problems. Besides, it is necessary to consider the limitations typical to the inductive reasoning of artificial intelligence developers.

Conflict of Interests

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgements

The authors would like to express thanks to the reviewers for their valuable suggestions towards the improvement of the paper.

References

Aleksandrovskaya, L.N., Aronov, I.Z., & Elizarov, A.I. (2001). *Statisticheskie metodi analiza bezopasnosti slozhnih tehnicheskich sistem*. M: Logos, 232, Moscow.

Aronov, I. (2019). Safety theory of technical systems: «black swans» and other animals. In *10th DQM International Conference, Life Cycle Engineering and Management Proceedings*, June 27-28, 2019, pp. 3-19, Prijevor, Serbia.

Aronov, I., & Papic, L. (2018). Reliability and Safety management of engineering systems through the prism of black swan theory. In: Anand, A., Ram, M. (eds.) *System Reliability Management. Solutions and Technologies*. CRC Press. Taylor and Francis Group. pp. 103-112.

Barber, M. (2003). Documentation: columbia accident investigation board report. *1*, *Astropolitics*, *1*:3, 127-135, DOI: 10.1080/14777620312331270069.

Chu, M., Manyika, J., & Miremadi, M. (2015). Four fundamentals of workplace automation. *McKinsey Quarterly*, November. Retrieved from https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/four-fundamentals-of-workplace-automation#

Davis, L. (2002). *Man-made catastrophes: from the burning of Rome to the Lockerbie Crash*. Facts on File, 402, New York.

Dragan, I.-M., & Isaic-Maniu, A. (2014). The reliability of human factor. *Procedia Economics and Finance*, *15*, 1486-1494.

Gurevich, P.S. (2011). *Philosophskoe postizhenie cheloveka: problem, tendencii I novie temi philosophskoy antropologii*. Saarbrucken: LAP LAMBER Academic Publishing, Saarbrücken, 654. Germany.

Harari, Y.N. (2018). *21 Lessons for the 21st Century*. Spiegel & Grau, Jonathan Cape.

Harris, S.D., & Zeisler, S. (2002). Weak signals: detecting the next big thing. *The Futurist*, *36*(6), 21-28.

Irving, J. (1982). *Groupthink: psychological studies of policy decisions and fiascoes*. Boston: Houghton Mifflin.

Kopachevsky, I., Kostyuchenko, Y.V., & Stoyka, O. (2016). Land use drivers of population dynamics in tasks of security management and risk assessment. *International Journal of Mathematical, Engineering and Management Sciences*, *1*(1), 18-25.

Lewis, H.W. (1990). Technological risk. *N.Y.: W.W. Norton*, *13*(1), 45-45.

Plotnikov, N. (2015). Osnovania teorii nadeznosti cheloveka-operatora (pilota). *Nadeznost*, *2*, 90-97.

Reiman, T., & Rollenhagen, C. (2013). Does the concept of safety culture help or hinder systems thinking in safety? *Accident; Analysis and Prevention*, *68*, 5-15.

Safety Management Manual (SMM) (2013). *Doc 9859 AN/474, ICAO*, 300.

Singh, K., Rajput, A., & Sharma, S. (2020). Human fall detection using machine learning methods: a survey. *International Journal of Mathematical, Engineering and Management Sciences*, *5*(1), 161-180.

Taleb, N.N. (2008). *The black swan: the impact of the highly improbable*. Penguin Group. **ISBN:** 9780141034591

The Fukushima Daiichi Accident. (2015a). *IAEA Report. Description and Context of Accident*, 238.

The Fukushima Daiichi Accident. (2015b). *IAEA Report. Safety Assessment*. 2, 186.

The Rt Hon Lord Justice Taylor (1989). The Hillsborough Stadium Disaster. *Interim Report. Home Office London, UK*.

U.S. Columbia Accident Investigation Board (CAIB). *Report of Columbia Accident Investigation Board*: *United States, 2003* (Report No 1). Retrieved from https://www.nasa.gov/columbia/home/CAIB_Vol1.html

Vasilyev, G.G., & Salnikov, A.P. (2014). Analiz prichin avariy vertikalnich stalnich reservuarov. Neftyanoe Khoziaystvo, *2*, 106-108.

Von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton University Press.