Human Success: Old wine in new bottles, or a shift of mindset for HRA in an automated world?

Andreas Bye
IFE, Halden HTO Project, Norway

Abstract: Words matter. Human reliability analysis (HRA) has been hampered by a negative connotation around the concept of human error. Many people feel that this concept is related to blaming humans. Why doesn't the HRA community redefine the vocabulary and call it human success?
Most HRA practitioners would say that the term human error is neutral and has nothing to do with blaming. After all, HRA analyzes the context and predicts the way in which humans are set up for error and quantifies its probability. In an apriori analysis, this is as much a prediction of the probability of human success. This discussion is of interest for the reputation of HRA, and more for educating the public about what HRA is. However, more automation in process control will change the operator role. Thus, it may be relevant to change the vocabulary also to underline the fact that HRA is looking for other things than before. It is not so relevant anymore to search for errors of omissions of procedure steps if there are no procedure steps.
It is a fallacy to believe that the human role can be excluded from the safe operation of an automated, complex industrial plant, or an airplane for that matter, as seen in the Boeing Max8 accidents. The human will be given a role, and we have to design that role to be adapted to the human strengths and weaknesses. Also, we need to be able to analyze the safety impact of the human role in this automated future.
The human role will be more monitoring and detection, not only of anomalies in the processing plant, but also of automation failures. The analysis of this role must be the task for future HRA. How do we analyze this role regarding safety? A redefinition of some of the HRA vocabulary would better explain what HRA can do for safety analysis of automated systems, and more importantly, improve understanding of the new human role in the future world of automated and autonomous systems.

Keywords: HRA, human success, human error, automation, human role.

1. INTRODUCTION

The words "human error" are used in many different settings. In daily speak, e.g., in newspapers, they are mostly used about causes of accidents or incidents. Examples are numerous on descriptions of how a "human error" caused an accident. This might typically be that somebody has done something outside of prescribed normative behaviour, sometimes even described in procedures or instructions. If it was a big accident in a big company, the immediate follow-up reaction is normally that "we shall improve our routines", or "fix the procedures". The next step is often looking for ways to remove the human from the loop, "so human error cannot happen again". All this is in good intentions to be able to avoid a similar accident to happen again. Although in later years, investigation committees have been much better at analyzing accidents in a systemic way, based on a system view as all human factors experts promote. This is discussed in a vast number of papers, and to change this way of looking at the world has been the motivation for several fields of research, e.g., Resilience Engineering (Hollnagel et al., 2006). The main point is that human variability will always be there, and that it is good, and that we should exploit it to make the systems more robust or resilient.

A misunderstanding partly caused by the "human error" language is that many people, including human factors (HF) professionals, view human error as a blaming term. However, it is not. The definition of human error is failure of a Human Failure Event (HFE). An HFE is any event in which humans are involved in a scenario. HRA is to analyse the probability, i.e., to predict, whether the event (including humans) will succeed or fail. So the choice to call it human error is a choice to focus on the failure branch of a binary tree. In the analysis itself, of course the success branch is as important, and is determined the moment the failure probability is calculated. This will be the fact in all apriori analysis of events.

The use of the terminology of "human error" in HRA may also give the impression that we are studying weaknesses of operators, and the reliability of human actions decoupled from the environment that they are in. This is clearly a misconception, since HRA analyzes the operators' environment and context and how this sets

them up for success or failure. In the long term, such wrong beliefs might build under the common argumentation that the automation and technology should be designed to remove humans from the loop and thus eliminate the opportunity for "human error".

Anyways, the wording on human error has a negative connotation, and searching for human error is seen as a quest for only studying how humans can fail, not how they can save the day. This should be changed. This paper argues for a change to "human success" based on two pillars: The actual practical use of the human error term by HRA practitioners, and the new role of humans in a more automated setting that also should be analyzed. The question is also: Is this old wine in new bottles?

2. REDEFINE THE LANGUAGE DUE TO PRACTICAL USE?

2.1. Definition of Human Error and Human Failure Event

Historically, "Human error" has been defined in many different ways. Reason (1990) discussed human error in a cognitive context, and many has used the framework of Skills, Rules and Knowledge (Rasmussen, 1983). Different ways of failing relates to different context based on the task at hand and the training and experience of the person. HRA has often used the distinction between Errors of Omission (EOO) and Errors of Commission (EOC). Errors of omission is the classic normative view of humans that fits well with the "normal" view of human error, that is: If the human follows the procedures everything will be fine and it is the fault of the human to do something else than prescribed. Errors of Commission though is more complex and involves other actions in various settings.

A new and simpler definition is rather to state that "Human Error is the failure of a Human Failure Event". In order to fully define this, we need to define a Human Failure Event (HFE). In this paper I will use the definition: HFE is an event that involves human actions.

2.2. Categories of human actions in HRA

Probabilistic Safety/Risk Analysis (PSA/PRA) uses fault trees and event trees in order to estimate the probability of a scenario ending up in a failure state or in an OK state. In the context of nuclear power plant operation, the most common end state of interest is core damage. This is the typical endpoint of PRA level 1 (although there are others of interest as well, such as release of radioactivity to the environment). PRA level 1 is far more studied than PRA level 2, which is estimating the probability of release to the environment, given core damage, and level 3, which estimates probabilities of number of casualties given release of radioactivity. The scenario in a PRA level 1 is often well understood and well covered with emergency operating procedures (EOPs). This is also why errors of omission are so much used and modeled. HRA model HFEs as part of the event trees in the scenario modeling.

The starting point for PRA level 1 is the Initiating Event (IE). Human actions have traditionally been classified in three types (IAEA, 1995):

- *"Category A actions that cause equipment or systems to be unavailable when required post-fault.*
- *Category B actions that either by themselves or in combination with equipment failures lead directly to initiating events/faults.*
- *Category C actions occurring post-fault. These can either occur in the performance of safety actions or can be actions that aggravate the fault sequence."*

Type C, post-IE human actions, is the most studied field of HRA, as for PRA, maybe simply because it is the most important one. If the goal is proactive safety, one should try to avoid core damage and be able to handle any IE. In a nuclear power plant, the handling of IEs is covered by a set of EOPs that when executed by the operator will mitigate an accident by activating the relevant safety systems designed to support the safety functions. There is a limited amount of safety systems, so the actions feasible are also limited. If the main goal is to cool the core, there are not that many other ways to cool the core than by utilizing the existing cooling and auxiliary cooling systems at hand, at least within the short term. The main purpose of HRA is to estimate the probability of whether the operators, using their EOPs, succeed in mitigating the accident, get the plant into a safe state, and avoid core damage. This human error probability is then direct input to the PRA.

The discussion above is focused on nuclear industry applications. HRA is and should also be used in other industries. These may have other framework conditions than the nuclear industry, especially when it comes to the use of operating procedures. The issue of automation is also even more relevant in other industries, and in some ways these industries have a less confined search space for the possible event sequences than what is described above for post-IE actions in the nuclear field. Such a more open search space may resemble more the field of pre-IE actions in the nuclear field, which is typically maintenance events (type A and B actions) and not supported to the same degree with procedures. This is still an under-explored field of study and needs more attention in the future. However, this is also a field in which the use of human success has been explored more, e.g., by Solberg et al. (2023).

2.3. Practical Use of Human Error in HRA, type C human actions

It is the job of an HRA analyst to model the tasks at hand, to explore the errors that may happen, and to estimate the probability of failure by analyzing the context the operating crew is acting in, typically by evaluating the impact of performance shaping factors on the fulfillment of the tasks or goals. Adequate time to react is always an important factor in such evaluations. It is also a good example of why human error can be a wrong term in such cases. If the time required to mitigate an accident by the procedures at hand is far less than the time available, their chance to succeed is almost zero. However, it is still called human error. We could as well call this case human success, and analyze whether the context (in this case the available time) gives the human any chance whatsoever to succeed. Does the context, the environment and the situation, set the human up for success?

We could then define Human Success as the success of a Human Success Event (HSE). HSE is defined as an event in which humans are involved. As we see, both an HFE and an HSE can end in both success and failure. We are just calculating the probability of both of them. (by definition, if the Human Error Probability (HEP) is 0.1, the Human Success Probability is 0.9). So whether we call this HFE, HSE, human error or human success, is actually irrelevant in the practical use in these cases. An HSE would be used in exactly the same way as an HFE today. The event involving the human action would be the same, only the name would change.

Hence, we may skip the whole use of the term “human error”, and use “human success” instead. For practical use in HRA type C, this would mean exactly the same. For people outside the HRA and PRA communities though, it might give a more positive view on what we are doing. Most importantly, we could get rid of the impression that we are studying actions and situations in which humans are blamed for failures that are happening. They are actually rather part of saving the day, by using the tools at hand (safety systems and procedures).

2.4. “Human Success” for Type A and B Human Actions

Can “human success” be used also for type A and B human actions? These types of actions are not that much modelled and thus studied in HRA. The reasons are many. One is that mitigation actions (type C) are the most important for the safety analysis in PRA. Another reason is that it is more difficult to study type A and B actions. As stated above, the type C actions are more confined and limited by the available safety systems. Type A, e.g., maintenance errors, does not necessarily have the confined frames. Although there may be a set of procedures, the ways in which to make failures so the goal is not achieved may be more. There may be more ways to do errors, e.g., to inhibit a valve by wrongly painting it, or by leaving an inhibition mechanism in the wrong position. So if the search space for errors will be more open, the number of possibilities are more and thus it may be difficult to know when enough is enough.

Can we use the term “human success” on such actions? The use of human error is useful in HRA since one may reduce the search space by looking for ways to make errors compared to a guideline or procedure. Searching and counting for thousand ways of succeeding (given the probabilities 10E-3 for failing) might require more resources than searching for a few ways to fail. However, one may keep the same way of analyzing or studying the issue, what we call it in the end depends on the relation to the final goal. Valves that are left in the wrong position or tags that are painted over are examples of errors that will become latent errors for operational scenarios at a later stage. Whether one calls this success or error is maybe irrelevant. So related to type A and B, it should also be possible to use the term human success. However, one should maybe keep the same methodologies, and search for error opportunities.

2.5. Predictive and Reactive Analysis

The concept of success and error may be different in a predictive safety analysis use (HRA) and in event analysis. In investigations of accidents the events have already happened, and there are no probabilistic predictions. The practice of only looking for errors in event analysis is long criticized (Hollnagel et al., 2006; Solberg et al., 2023), and one drawback is that it can easily lead to blaming, as well as not being able to prevent anything else than the exact same accident happening again. This has led authors to promote looking into human variability and adaptability (Hollnagel et al., 2006) and human success when investigating accidents and incidents (Solberg et al., 2023). The idea is the same, that by only investigating failures, we can avoid the exact same accident to replicate, but we won't be able to avoid new types of accidents. Also, another point is that many human errors are unavoidable, and we should learn to live with them, and build joint cognitive systems that can handle variability both in human behaviour and in system behaviour. However, the search for human errors in predictive safety and HRA is fundamentally different. HRA actually searches for the variability in performance that is discussed in e.g., Resilience Engineering (Hollnagel et al., 2006), and evaluates the context for human events and thus their probability for error, or success.

Sometimes HRA methods are used for post-incident analysis of events. This kind of reactive use of HRA methods is different from the apriori, proactive kind of analysis. In the case of after-the-fact analysis the event has already happened, errors have been made (although maybe no errors that anyone can be blamed for), and that is the focus of the analysis. HRA methods may then be used in order to explore what else could have gone wrong, and in that case the HRA method is still used in kind of a predictive way, in searching with what-if kinds of questions and expanding the context to explore. Also in a reactive analysis, one may find contexts in which the persons involved had a very difficult environment to handle. They were in a way "set up for error". Another way to view this is that they were *not* set up for success.

2.6. In Practical Use in HRA we can use Human Success instead of Human Error

The conclusion so far is: When we are describing human error in the context of practical HRA, predictive analysis of human actions' impact on safety, we may as well use human success instead of human error. This is especially the case within post-initiator mitigation actions, i.e., type C human actions.

However, this would be "old wine in new bottles". There is no new meaning in this that would lead to changed purpose for HRA practitioners. Whether it might be worthwhile to change the vocabulary only to get a better public understanding of what HRA is about, is a discussion the HRA community should initiate.

3. IS HUMAN ERROR DIFFERENT WITH AUTOMATION?

3.1. The Human Role with Automation

The role of operators in a more automated world is by many described to be more monitoring, and less action oriented. An example is that instead of executing procedures manually, procedures or sequences will be executed automatically, or even adaptively (Fernandes et al., 2024). Also, with the passive safety systems, safety will be based on more physical and simpler things happening, not even automated or based on active events such as starting pumps or closing valves.

The basic question with automation is: Can you at the time of design predict all situations that are going to happen and their corresponding end states? This does not necessarily have to be on exact event sequence level, but on a level covering critical events. Will black swan situations occur, and how are they handled? If it is possible to predict all things that are going to happen, one might as well make a completely autonomous system. However, if this cannot be guaranteed, one must include a human to collaborate in some way with the automatic system (in this paper I don't separate automation and autonomy or classify these levels). And in this case, how should the relation between the human and automation be designed and analyzed?

A classic example is the Boeing Max8 accidents. The actual automation algorithm worked as designed in this case. However, the input was based on a flawed instrument, and with wrong input, the algorithm gave wrong output. One way to fix this problem is of course to think about this at design stage and to think about all other

events or erroneous inputs to the automation system, and then design a more robust system. However, is this possible? Another way of approaching this challenge is to design a collaborative joint system in which an operator (pilot in this case) could be warned or just easily take over without having to fight with the automatic system. There are also several challenges with this approach: Should the operator just monitor the automatic system, and should there be some sort of self-reporting of problems from the system? Or should the operator be more on-line collaborating with the automation and do some tasks? There are a number of questions here, including long-term skill retention. One important challenge to avoid, is making the human a pure backup for the automation. This might be seen in some of the cars these days. If the auto-pilot is not designed for the situation, it simply hands over the control to the driver. One must avoid situations in which this is done at critical points in time. So one should analyze when and in which such hand-overs are done, and put focus on ample time for the human to act and other dimensions giving the joint system a fair chance to succeed. A case of handing over control from automation to a human seconds before impact, is a situation one should avoid. The main point will be to analyze such issues from the human's perspective. How are the humans set up to solve the situation together with the automation, in a joint cognitive systems view?

3.2. Does automation imply a shift of mind-set for safety analysis?

Back to the question of whether a more automated world would imply a shift of mindset for HRA analysts: In order to dive into this, we should discuss the need for procedures. If higher levels of automation are included, how should the human role be? One may think of various extremes: One thing is a clearly defined procedure, in which one has concrete actions with checkpoints to monitor the total system, including the automation. Another extreme would be no procedures, just a free role of an operator to monitor the automation. In this case it would at least be very difficult to talk about errors of omissions (EOOs) from procedures, since there are none or very few to relate to. One might also think of configurations somewhere in between, e.g., operators may intervene and do some manual actions just to get the feel for it. Errors of Commission (EOCs) could then be defined according to the outcome of the actions, which states do the plant/process/plane end up in. In some of these cases, if operators are given a free role, it would be difficult to brand any events "human errors". It would more be a role to intervene if necessary and save the day in various ways. Human success would in that case be more proper wording. However, would such configurations be allowed in the nuclear industry? Probably not, but that depends on the safety relevance. In any case, if humans are given a role in monitoring, their safety relevance must also be studied.

3.3 Beyond Design Basis Events (BDBE)

One thing to consider when talking about more automated systems, are beyond design basis events (BDBE). One thing is technical limits for design basis events based on physical parameters. Another thing is evaluation of automatic systems and what the design basis for them are. Will black swans enter the playground from completely different angles? Angles that are not based on physics, but rather based on assumptions that are forgotten or misjudged. In some way this is analysis of actions that are not on the sharp end such as actions by operators handling an event. Such an analysis will be focused on latent errors that might be compared to maintenance errors: Design errors done in the automation system earlier in time might create a difficult on-line operational problem.

EOPs mostly cover single failure scenarios, not all combinations of events. Random combination of internal events is normally considered beyond design basis. In experiments in HAMMLAB, we have tested operating crews in difficult scenarios, often combinations of scenarios such as a Loss of Feedwater (LOFW) immediately following a Steam Generator Tube Rupture (SGTR). Such combined events may be judged to be beyond design basis. These situations then create difficulties such as choosing the right procedure, or getting out of one when they know they should be in another. We have seen crews succeed in such difficult events, so that is another case to the argument that we should use the vocabulary human success, not human error. Another interesting thing about this, is that in future more automated plants, the claim is that there will be fewer actions to do for the operator. Is the idea that the designers will have thought of everything, and the rest is BDBEs? This is a classic, and not a good, way of developing technology: Engineer a best possible system and leave the rest to the human operators. However, if this should be the case, if the operator has the role of monitoring, diagnosing, and checking the automation system as well, this will probably not be based on detailed procedures. Then probably most deviations might be classified as BDBEs, and there will be no EOOs, only EOCs, and in

addition, cases where the operators save the day due to innovative diagnosis capabilities. In that case, we should definitely use the vocabulary human success.

4. DISCUSSION

How should we define and model HFEs in joint human-automation systems? Will an automation failure be defined as "failure of the human to detect automation failure"? In that case, we might fall into the same trap as earlier, to use the blaming term on humans when humans and automation co-exist. One should maybe try to identify failures of the automation, and similar success opportunities for humans monitoring the automation. This would be a similar approach as in type C human actions, studying mitigation actions and whether operators are given the context to manage to save the day. In that case, the vocabulary should be human success.

As discussed in Arigi & Bye (2024), new advanced reactors are based on a higher level of automation and more passive safety. Even in such configurations the operator will play a role, but probably in a much more monitoring role. Monitoring is different since it is based on a diagnostic role. Thus, which HRA methods will be able to analyze this new human-automation role? Diagnosis is included in many of the HRA methods. The methods that are based on models of cognition and by that setting the humans in the centre should be able to analyze most of the human-automation collaborations. The key is that the humans must be the centre point of the analysis, as this paper argues that changing the vocabulary would help doing.

One issue that could come up if we redefine human error to human success: Will a search for "success modes" be more normative than search for error modes? Is it rather so that the search for failure modes actually includes the variability needed? A danger would be to move into an even more normative and "adhering to procedures" way of analyzing work. Also, there are probably more ways to succeed than to fail. Will the search space be too vast? It may be easier to describe one error dimension than hundreds of successes. Will the number of success modes be more than the number of error modes? Probably, we need to keep the term human error probability. It gives more meaning to talk about probability numbers as 1E-4 rather than promoting 0.9999 as a meaningful number.

For the analysis of automation failures, one might find it useful with a change of mindset for analysts, from searching for possible errors to searching for success. What is needed for support to humans to "save the day"? The dimensions could be the same as for the performance shaping factors (PSFs). For example, how much time is needed after automation stops working to get a proper situation awareness, for again to be able to do the proper actions? This would be analogous to analyzing PSFs though, but it could be linked to a more general "overview measure", than being linked to a specific action or event. It could also be linked to other goals, not directly task execution which would be linked to events and scenarios, but it could be linked to the goal of retaining skills. In this way a more long-term collaboration with the system could be analyzed. It would be important to find the right goals and dimensions of such an analysis. Alternatively, one could do the analysis for PSFs for classical scenario goals and operational safety, and then a further analysis towards organizational support and safety culture.

4.1 Analyzing Automation, we can use Human Success instead of Human Error

Given the discussion above, the term Human Success describes Human-Automation collaboration as good as the term Human Error. However, does it contribute with something new, and it is needed?

It may contribute with something new, in focusing on slightly other dimensions than earlier. The focus would more naturally be on EOCs, which is more relevant in an automated setting as described above. However, again, this would to some extent be old wine in new bottles.

The more important and newer thing is rather that it would enable a broader type of analysis than just focusing on PSFs related directly to the event sequence. In that way, the analysis could be broadened to cover other goals than the pure safety goals as well. It would also cover better the new way in which automation is linked with human behaviour.

5. CONCLUSION

We should change the vocabulary of HRA from "human error" to "human success". The first argument is that it would better explain what HRA is for people outside the HRA community, as it would better describe what the analysis is about. It would also be easier to avoid misconceptions such as that HRA is linked to blaming. Human error has for some gotten a bad connotation. For HRA practitioners, this interpretation has not been relevant, especially not in practical use. So it is old wine in new bottles.

The second argument is that it may better describe the new automated future, and thus help analysts and the community in a shift of mindset required for a broader safety analysis. It could help rethink the factors in play in more highly automated systems, so we better understand the new human role in the future world of automated and autonomous systems and to be able to determine this role's impact on the safety of the total system. This is of crucial importance since if we follow the technology drive and think automated systems are so much safer since they "eliminate human error", we will still have serious accidents in the future. The human will play a role in the future automated systems, and we should rather see automation and I&C as part of the technology in a joint cognitive system, where the role of the human is defined and evaluated from the beginning. Then HRA can evaluate the degree to which operators can actually detect and determine what is going on and have a possibility to save the day.

Acknowledgements

This work was supported by the OECD Nuclear Energy Agency (NEA) Halden Human-Technology-Organization (HTO) Project. Thanks to the members of the Programme Review Group of the Halden HTO Project for review and constructive comments.

References

Arigi, A.M., Bye, A. (2024, *in press*). Needs for change of human reliability analysis for new advanced reactors. PSAM17 & ASRAM2024, Sendai, Japan.

Fernandes, A., Bisio, R., Bye, A. (2024, *in press*). Adaptive Automation in Control Rooms: Discussing safety challenges in computerized procedures. PSAM17 & ASRAM2024, Sendai, Japan.

Hollnagel, E., Woods, D. D. & Leveson, N. C. (Eds.) (2006). *Resilience engineering: Concepts and precepts.* Aldershot, UK: Ashgate.

IAEA. (1995). *Human Reliability Analysis in Probabilistic Safety Analysis for Nuclear Power Plants*. IAEA Safety series 50-P-10, International Atomic Energy Agency, Vienna.

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernethics.* Smc-13, No 3. 257-266.

Reason, J. (1990). *Human Error.* Cambridge University Press.

Solberg, E., Kwei-Narh, P., Bisio, R. (2023). Including successful performance in the scope of the event's causal analysis. HTOR-046. Halden HTO Project.