WHEN HUMAN ERROR IS GOOD: APPLICATIONS OF BENEFICIAL ERROR SEEDING

Ronald Boring[1], Roger Lew[2], Thomas Ulrich[1,2], Kateryna Savchenko[1,2]

[1] *Idaho National Laboratory, PO Box 1625, Idaho Falls, Idaho, 83415, USA*
[2] *University of Idaho, 709 S Deakin St, Moscow, Idaho 83844, USA*

Human error is usually studied in the context of its negative consequences. It is viewed as something to prevent and mitigate. There are, however, applications of human error in which the process of committing the error becomes beneficial. This paper explores four applications of error seeding in which human errors are deliberately invoked: (1) education and training to enhance human performance, (2) research to model human performance at boundary conditions, (3) entertainment to make gameplay more engaging, and (4) security to maximize critical infrastructure. We demonstrate the field of human error should equally investigate the cases when human error serves a positive outcome.

I. I. BACKGROUND ON HUMAN ERROR

Human error (also sometimes commonly referred to as human *variability*) is described by Reason[1] as, "all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome." In other words, human error is either decision making or behavior that does not turn out as expected. The causes of human error are the subject of several taxonomies[2,3,4] and human reliability analysis methods.[5] External causes such as a random failure of a piece of hardware used by the human are not typically considered human errors. Rather, human errors are considered systematic and the byproduct of discernible root causes. For example, Reason[1] distinguishes errors in carrying out a planned action (e.g., slips and trips) from errors in planning the action (e.g., mistakes).

The inevitable conclusion of these approaches is that the consequences of human error are undesirable and preventable. An accident investigation attempts to isolate the causes of human errors that contributed to the accident and prescribe ways to prevent future occurrences.[6] A human reliability analysis attempts to quantify the human error probability and mitigate the chance of significant consequences.[7] These approaches are valid and necessary, because they address the case where human error has significant safety consequences. But, the literature on human error has largely overlooked another class of human errors—those that may be desirable or beneficial.

Unintended outcomes resulting from human decision making or behavior may in some cases actually be desirable. As with the traditional study of human error, here we rule out cases that are simply random chance or, perhaps better stated colloquially, "dumb luck." The unintended positive consequences of a human error (e.g., accidentally hitting the brakes of an automobile at precisely the right moment before an unseen accident ahead) are not a domain that can be systematically understood. However, there is a category of human errors that serve a beneficial end state. In this paper we explore this class of errors resulting from error seeding,[8] which are errors that are intentionally invoked in a manner where they would not otherwise occur. The term *error seeding* is used in software testing and debugging, e.g., introducing software errors to determine the robustness of the application. In the present case, we refer specifically to human error seeding.

The ATHEANA human reliability analysis method,[9] in its characterization of human errors, introduces the concept of the *error forcing context*, which is defined as, "The situation that arises when particular combinations of performance shaping factors and plant conditions create an environment in which unsafe actions are more likely to occur" (p. xv). Human error seeding is the case of deliberately maximizing the error forcing context with the intent to cause a human error. The intent is not to cause a high consequence event; rather, the goal may serve purposes such as:

- Education and training to enhance human performance in anticipation of possible events,
- Research to model understanding of human performance at boundary conditions,
- Entertainment to make gameplay more engaging, and

- Security to maximize protection of critical infrastructure.

In the following sections, we explore these four uses of beneficial human error seeding and provide examples from each of these applications.

II. APPLICATIONS OF BENEFICIAL ERROR SEEDING

I.A. Errors for Training: Learning from Your Mistakes

Learning and training philosophy fluctuates between two opposing paradigms in relation to the role of errors.[10] The traditional training approach is referred to as error avoidance and stresses that errors should be avoided entirely. The rationale for avoiding errors stems from a minimalistic philosophy, which aims to provide the trainees with the simplest set of materials and instruction containing the core information needed to be an effective user.[11] There are numerous benefits to this minimalistic approach. First, the training time can be reduced because the material only covers the optimal method for accomplishing a given goal. Second, good habits are formed since the trainee is exposed only to the proper solution paths for a given task. This training approach is quite effective for simple tasks with simple problem structures, because the problem structures and solution path to solve each problem are clearly defined and readily apparent. More elaborate tasks with complex problem structures may suffer from this training approach, however, since multiple solution paths to follow and multiple failure paths to avoid are present but not necessarily apparent to the trainee. Furthermore, the trainee has not been exposed to these failure paths, because they were only exposed to the solution paths or even a single solution path selected for the training. The trainee might not have the proper mental model to navigate the problem space when encountering an error for the first time, which may lead to poor performance and a lack of resilient responses to problems.

The other, opposing error training paradigm stresses errors are important and should be allowed throughout the learning process to improve outcomes.[12] Errors can be allowed during the learning process through two different mechanisms referred to as error tolerance and error seeding. *Error tolerance* results from typical human learning behavior generating naturally occurring errors if the restraints on the training environment are sufficiently relaxed to allow for those errors to occur. Errors can also be explicitly seeded into the training program to ensure that each trainee is exposed to particular or important errors based on desired learning outcomes. For example, in nuclear reactor operator training, all operators undergo a variety of loss of coolant scenarios, since successful mitigation of these faults is crucial to maintaining safety of the plant. The plant simulators provide an error tolerant environment for operators to commit errors while learning.

The rationale for both error tolerance and error seeding is plentiful. Inserting faults into the training program familiarizes the trainee with novel situations in which their current mental model might be incorrect or incomplete, causing them to commit errors. By alerting the trainee to these deficiencies, it promotes the integration or reconciliation of their mental model to accommodate the new information. Indeed, it is not necessary to expose operators to every potential error, but, rather, a sample error from a class of errors can transfer to future novel errors possessing similar structures. Error seeding also confers the benefit of increasing the alertness of the operators during the training and providing an impetus of potential failure that adds an element of realism to the training program.

Error seeding can also be used as a method for assessing performance and predicting future performance, which could prove beneficial for personnel selection. Error seeding provides a means to identify the strengths and weakness of the trainee during the learning process and provide immediate feedback to eliminate incorrect mental models and enforce correct mental models. Nuclear reactor operator training simulations, for example, can benefit from these error seeding advantages to generate accurate mental models within the trainees and ultimately improve their nuclear process monitoring and control task performance.

I.B. Errors for Research: To Err is Human

Training invokes human errors to familiarize users with challenging scenarios beyond normal use cases. A variant on error seeding in training is using a similar experimental configuration for research to discover the underlying principles behind why humans commit errors in the first place. An accurate model of human performance must be able to account for human misperceptions and error tendencies. By systematically controlling the context in which errors occur, it is possible to gain insight into human perception and cognition.

The use of a training simulator for human error research is illustrated in a recent international human reliability analysis benchmark.[13] Using a training simulator comparable to that found in a nuclear power plant, a series of scenarios were

presented to licensed reactor operators with the intent to test performance. One of the difficulties in testing highly trained and skilled operators is the low likelihood of significant errors. As such, the scenarios that were presented to the operators included scenarios deliberately designed to invoke errors. Because the operators are trained on all likely accident scenarios, the way to catch them off guard is by creating scenarios that would be extremely unlikely to occur in the real world. The tradeoff between surprising the operators and maintaining realism favors scenario novelty in this case, by triggering a compound fault in which one fault masks another fault. Even in such cases, the operators are often able to resolve the faults, but the resolution of the faults takes longer and involves more steps than would otherwise be evidenced. In addition, the available operating procedures may not be a good match to the circumstances, forcing the operators to use additional problem solving strategies beyond the rote procedural prescriptions. Such scenarios present the opportunity to measure a range of human performance effects of the error seeding. The types of manipulations may also be controlled to limit the types of error contexts being investigated (e.g., outside procedural scope, symptom masking, or complex event evolution over time). It should be noted that such studies are not tests of operator deficiencies; it is assumed that given the right error forcing context, any skilled person will eventually fail to perform the task correctly. Indeed, the need to resort to contrived (i.e., error loaded) scenarios is a testament to the proficiency of the operators.

The research on human error mentioned in the previous paragraph focused on using full-scale simulators with licensed operators. Such research is costly and may preclude a statistically significant sample size simply due to the small available population size of qualified personnel. For example, in the international human reliability analysis benchmark, 14 crews were tested across four conditions.[13] The small sample size of 14 crews presented a challenge to conventional inferential statistical techniques, and Bayesian analytic techniques were employed instead. Yet, the sample size of 14 represents the largest published study of its sort in human reliability analysis. A typical single-unit nuclear power plant may, for example, only have a dozen available crews split across all shifts, and the uniqueness of each plant configuration confounds the use of crews from different plants in a single study.

Thus, it is a challenge to find a large enough sample size to arrive at statistically reliable conclusions on human error. To overcome this limitation, recent work on microworlds[14,15] creates a simplified simulation environment that can be used by non-professional users. For example, Liu and Li[16] developed a simplified process control simulation to validate human error probabilities posited by the SPAR-H human reliability analysis method.[17] Liu and Li subjected 75 participants who were not process control experts to a simplified simulation with scenarios for which they were trained and untrained. The untrained condition served as the error seeding condition. Liu and Li tallied errors committed during the scenarios, using the overall number of actions as the denominator. Although the study did not faithfully reproduce the conditions of professional operators using a full-scope power plant, it provided important data toward validating human reliability analysis, and it demonstrated the value of error seeding.

A more generalized form of error seeding for research is any psychological experiment that manipulates a variable that includes difficulty. Such experimental manipulations are common in cognitive psychology experiments. For example, psycholinguistic research to determine lexical processing may use studies contrasting common (i.e., high lexical frequency) words with difficult (i.e., low lexical frequency) words. The point of such research is to determine if similar cognitive processes are used, for example, for word recognition or word naming.[18] The purpose of the difficult words is not specifically error seeding, but the process acknowledges that difficult conditions may elicit different cognitive processes than their easier counterparts. Such experimental conditions extend well beyond psycholinguistics research and frame the importance of error seeding as an important manipulation in research to model cognitive processes.

I.C. Errors for Gaming: Trial and Error

Error seeding for training purposes teaches humans about complex systems, while error seeding for research allows scientists to study human perception and cognition. Error seeding in gaming allows players to learn about the virtual universes in which they are interacting. Gaming represents one of the most common applications of error seeding, although it is rarely framed in the context of errors.

Many computer game players choose to play games in order to fulfill fantasies as a part of escapist relaxation from their daily routine.[19] This type of interface provides an opportunity to create a stronger connection between the player and the in-game character and other and deeper interactions with the use of audio and haptic elements, thereby increasing the immersion and escapism of the players into the fictional world while challenging their abilities on a wide range of skills and social interactions.[20] Many of these virtual universes are completely fictitious—sometimes they strive to accurately represent real world places and circumstances (e.g., edutainment games), while others walk a line between entertainment and history.

For many players, the challenge is what attracts them to the game, and, if the game is too easy, the person would lose interest in progressing through the game.[19] The challenge in the game indicates that the player would not be able to become successful at first. The ability to recognize and analyze the reasons why they failed and what went wrong in the game is important to make sure that the player understands that to succeed and overcome the in-game obstacles one needs to develop skills and make many attempts until the tasks are complete such that progress within the game is possible. The practice areas inside the gaming environment for testing those skills are needed for the player to improve.[20,21] For example, many games have options for creating custom game play against artificial intelligence instead of other real players to test and master the characters' unique in-game mechanics. Games challenge the player's skills, knowledge and resourcefulness, demanding a wide range of cognitive abilities and requiring more visual attention resources, reaction time, and faster processing speed.[22] Beyond simply solving logical problems, players can engage in exploratory learning in complex social, political, and cultural settings in ways that challenge their core-beliefs in "real-world" settings. Self-discovery can occur not only while in-game but also through reflection after leaving the interactive environments.[23]

Video game designers and game developers are trying to cater to a general audience by implementing different types of user interfaces. The non-diegetic or points-based user interface is the traditional and most commonly used type of interface with a head-up display (HUD) showing information like a health bar and map.[24] The alternative to points-based interfaces is diegetic or story-based user interfaces. Diegetic user interfaces create the greatest immersive experience for game users by enabling the player and avatar to interact with the elements that exist within the virtual universe through visual, auditory, and haptic modalities. All the interactions take place in fictional worlds that are sometimes based on reality, and the normal player HUD can be a part of the character's equipment, like the car dashboard in many vehicle simulation games.[19,24,25] The diegetic game increases the sense of immersion and makes the players believe that they are playing the character in real-world situations.[19] A good example of an immersive diegetic environment is a game where users must rely on the use of in-game maps and a global position system to obtain the information needed for the character to proceed toward the goal in the game world.

Demands on the player's attention create an opportunity for human error that makes the game itself more challenging and keeps the player interested in the story and accomplishing tasks.[19] The immersion experience makes the player forget that they are playing the game, and they instead become absorbed by the in-game action. Games can become dull without players making mistakes and errors. Vigilance pays off. A few human errors arise because of lack of attention. Some players just do not pay enough attention to their actions, making an error of omission or inattention. Such a slip happens, for example, when the player places an item in a certain equipment slot and, during the distracting event, the player automatically activates the item that is not useful in that situation.

Game designers and developers create challenging in-game mechanics and seeding opportunities for human errors to destroy the established patterns of responding and to create engaging situations. Some perceptual confusion can occur when two things appear that are similar to each other, which is very similar to masking as discussed in the previous section, resulting in the player making the wrong decision. The appearance of similar or ambiguous items requires time for discernment by the player, but the distraction of fighting bad characters is created while the player has to choose the right option. A secondary task serves as an error seed for an already difficult task. Highlighting and making the important objects or items in the game pop up can assist the player in picking the right thing. The game developer must strike the right balance—a game that is too easy will not be perceived as entertaining, but a game that is too difficult will not be pursued by the gamer. Error seeding—making selective parts of the game challenging—is the key to active gameplay.

I.D. Errors for Security: To Err on the Side of Caution

Analog and early digital control systems could be physically secured. However, modern control systems connect field devices, controllers, historian databases, and human-machine interfaces (HMIs) in a control system local area network (LAN) segmented behind corporate networks.[26,27] Connecting the control room of a critical infrastructure to the corporate network provides analytics for engineers and business analysts as well as a means for vendors to provide remote support to control systems, but may also provide a path for a bad actor to engage in a remote cyber-attack. If a bad agent gains "backdoor" access to a computer on the corporate network, the compromised machine then provides a platform for gaining access to the control system network and staging an attack on the control systems and operators. Bad actors may try to mislead operators into taking control actions by modifying HMI displays, or use HMI displays to make operators believe the system is acting normally while it is under control of the attacker.

As a prime example, the Stuxnet Worm is a sophisticated "weaponized" computer virus designed to attack and destroy Iranian uranium enrichment centrifuges. Stuxnet replicated itself searching for specific Siemens programmable logic controllers. It destroyed its target by ramping the speed close to the mechanical breaking point. While doing so it provided spoofed signals to mislead operators. When the equipment failed it would appear as though it was caused by a mechanical failure, delaying the discovery of the worm.[28]

According to a recent report by IBM, human error was a contributing factor in 95% of security incidents.[29] Information technology professionals will carry most of the responsibility for monitoring networks, but operators will have to be mindful of cyber-security issues as well. Control system networks may contain mandatory intrusion detection systems (IDS) that will monitor network traffic for suspicious activity and alert system administrators.[27] Operators would then need to work closely with administrators to further assess the threat and take appropriate action.

Error seeding in this arena has multiple applications. The most obvious is in training operators to recognize and respond to cyber events. Cyber events can be malicious or accidental due to misconfiguration, equipment failure, and other potential causes. Event symptoms may appear as benign, subtle, or seemingly unrelated to one another. Operators must learn to recognize these cues, such as slow or unresponsive interfaces, intermittent glitches, etc., and respond appropriately. Counterintuitively, taking immediate corrective actions, such as rebooting a system or initiating a control action, may actually be a sub-optimal response. Rebooting a system can erase critical logs related to the event. Taking control actions could play into the hands of a bad actor. Presenting operators with unexpected cyber-related errors during training and possibly during everyday operations could increase the competency of operator response to cyber events.

A second application is to assess the capability of operators to detect malicious intent. If knowledgeable security professionals can detect intrusions that mimic known weaknesses of the system, then the ability of operators to detect those weaknesses can be identified. Ultimately, known weaknesses should be safeguarded, but the reality of industrial control is that many plants contain old and insecure code or older vendor systems that are more exploitable than modern systems. Identifying these weaknesses and their potential risk of inducing human error is the first step to more robust fixes.

Lastly, cyber-warfare has the potential to neutralize enemy critical infrastructure while reducing the risk of collateral damage. For example, destroying a power plant may accomplish the short-term goal of neutralizing the enemy, but it has long-term consequences to the civilian population. Introducing selective problems (i.e., seeding process control errors) could disrupt the function of the plant but leave it otherwise physically intact. Cybersecurity intrusions are one way to do that, but there are other technologies, including introducing grid disturbances. Error seeding disrupts operations without the need to destroy the system being operated. Understanding human operational parameters becomes the basis for beneficial error such seeding.

III. CONCLUSIONS

These four applications of error seeding are diverse, but they have a common thread. Whether training, research, gaming, or security, each domain relies on human users. Safety research has identified many of the causes of human error but always as a method to reduce human error. Indeed, the consequences of missteps can be great, and the tools produced to date have thankfully helped prevent accidents. Yet, the focus on the safety consequences of human error omits an entirely broader field of study—the use of human errors for positive outcomes. Seeding human errors can help raise awareness of possible safety and security concerns, can provide insights into error and general cognitive processes, can ensure and maintain engagement, and, by being used to disrupt normal operations, can even minimize the need for destructive military practices.

Human error is not always a bad thing; it can instruct, engage, and alert. Ultimately, seeding human error can increase human performance. Perhaps it has been an error to focus only on the negative consequences of human error. The science of using human errors in a beneficial manner remains novel, but there is already evidence to suggest it is worthy of further pursuit.

DISCLAIMER

REFERENCES

1. J. Reason, *Human Error.* Cambridge University Press, Cambridge (1990).
2. A. D. Swain, and H. E. Guttman, *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, Final Report, NUREG/CR-1278,* US Nuclear Regulatory Commission, Washington, DC (1983).
3. J. Rasmussen, Skills, rules, knowledge: Signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions on Systems, Man & Cybernetics, 13*, 257-267 (1983).
4. J. Reason, Lapses of attention in everyday life. In R. Parasuraman and R. Davies Eds., *Varieties of* Attention, pp. 515-549, Academic Press, New York (1984).
5. F. T. Chandler, Y. H. J. Chang, A. Mosleh, J. L. Marble, R. L. Boring, and D. I. Gertman, *Human Reliability Analysis Methods: Selection Guidance for NASA*, NASA Office of Safety and Mission Assurance Technical Report, Washington, DC, (2006).
6. S. Dekker, *The Field Guide to Understanding 'Human Error,' Third Edition,* Ashgate Publishing Ltd, Aldershot (2014).
7. B. Kirwan, *A Guide to Practical Human Reliability Assessment*, Taylor & Francis, London (1994).
8. T.Q. Tran, R. L. Boring, J. C. Joe, and C. D. Griffith, Extracting and converting quantitative data into human error probabilities. *Official Proceedings of the Joint 8th IEEE Conference on Human Factors and Power Plants and the 13th Annual Workshop on Human Performance / Root Cause / Trending / Operating Experience / Self Assessment*, 164-169 (2007).
9. J. Forester, A. Kolaczkowski, S. Cooper, D. Bley, and E. Lois, *ATHEANA User's Guide, NUREG-1880*, US Nuclear Regulatory Commission, Washington, DC (2007).
10. R. Quiñonez, T. Ryan, and L. Olfman, Designing CBT systems with errors in mind: Avoidance, seeding, and tolerance. *Journal of Information Technology Education*, *6*, 65-80 (2007).
11. J. M. Carroll, Reconstructing minimalism. In J. M. Carroll Ed., *Minimalism beyond the Nurnberg funnel*, pp. 1-17, MIT Press, London (1998).
12. M. Frese, F. Brodbeck, T. Heinbokel, C. Mooser, E. Schleiffenbaum, and P. Thiemann, Errors in training computer skills: On the positive function of errors. *Human-Computer Interaction, 6*, 77-93 (1991).
13. J. Forester, V. N. Dang, A. Bye, E. Lois, S. Massaiu, H. Broberg, P. Ø. Braarud, R. Boring, I. Männistö, H. Liao, J. Julius, G. Parry, and P. Nelson, *The International HRA Empirical Study: Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data, NUREG-2127*, U.S. Nuclear Regulatory Commission, Washington, DC (2014).
14. R. Boring, D. Kelly, C. Smidts, A. Mosleh, and B. Dyre, Microworlds, simulators, and simulation: Framework for a benchmark of human reliability data sources. *Joint Probabilistic Safety Assessment and Management and European Safety and Reliability Conference* (2012).
15. B. P. Dyre, E. J. Adamic, S. Werner, R. Lew, D. I. Gertman, and R. L. Boring, A microworld simulator for process control research and training. *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting*, 1367-1371 (2013).
16. P. Liu, and Z. Li, Human error data collection and comparison with predictions by SPAR-H. *Risk Analysis, 34,* 1706-1719 (2014).
17. D. Gertman, H. Blackman, J. Marble, J. Byers, and C. Smith, *The SPAR-H Human Reliability Analysis Method*, *NUREG/CR-6883*, U.S. Nuclear Regulatory Commission, Washington, DC (2005).
18. K. R. Paap and R. W. Noel, Dual-route models of print to sound: Still a good horse race. *Psychological Research, 53,* 13-24 (1991).
19. R. Rouse and S. Ogden, *Game Design Theory & Practice* (2nd ed.), Wordware Pub, Plano, Texas (2004).
20. Rosario, R. A. Munoz, and G. R. Widmeyer, An Exploratory Review of Design Principles in Constructivist Gaming Learning Environments. *Journal of Information Systems Education, 20*(3), 289-300 (2009).
21. E. A. Boyle, T. M. Connolly, T. Hainey, and J. M. Boyle, Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior, 28*, 771-780 (2012).

22. P. Dobrowolski, K. Hanusz, B. Sobczyk, M. Skorko, and A. Wiatrow, Cognitive enhancement in video game players: The role of video game genre. *Computers in Human Behav*ior, *44*, 59–63 (2015).
23. Informational Resources Management Association, USA, *Gaming and Simulations: Concepts, Methodologies, Tools, and Applications*, IGI Global (2010).
24. N. R. Prestopnik and J. Tang, Points, stories, worlds, and diegesis: Comparing player experiences in two citizen science games. *Computers in Human Behavior, 52,* 492-506 (2015).
25. M. Angelides, Immersion in Digital Games: Review of Gaming Experience Research. In *Handbook of Digital Games*, pp. 337-361, Wiley, Hoboken (2014).
26. V.R. Segovia and A. Theorin, *History of Control: History of PLC and DCS,* Lund University Department of Automatic Control, Sweden (2013).
27. Department of Homeland Security, Recommended Practice: Improving Industrial Control Systems Cybersecurity with Defense-In-Depth Strategies (2009).
28. D. M. Berry, Critical Theory and the Digital. Bloomsbury, USA (2014).
29. IBM Global Technology Services, *IBM Security Services 2014 Cyber Security Intelligence Index* (2014).