Evaluation of a Software Implementation of the Cognitive Reliability and Error Analysis Method (CREAM)

Roger D. Serwy and Esa M. Rantanen
University of Illinois at Urbana-Champaign

The Cognitive Reliability and Error Analysis Method (CREAM) represents a second-generation approach to human reliability analysis (HRA). The method, however, is very tedious to apply manually and not yet in widespread use and therefore largely untested. To allow for rapid and systematic evaluation of the CREAM method, a software tool for its application was developed. Results from several analyses undertaken to evaluate the method and the tool are also presented. The simplicity of our CREAM software allowed novices to analyze events in much detail but also revealed some critical shortcomings in both the method and the software tool. Several conclusions could be drawn from this allow for drafting of specific design guidelines for future upgrades of the CREAM software. These conclusions and recommendations can be further generalized to all software applications of HRA methods

INTRODUCTION

Human reliability analysis (HRA) refers to techniques that seek to model human error in the context of complex systems in similar terms as probabilistic risk assessment (PRA), which in turn refers to systematic and comprehensive methodology to evaluate risks associated with complex engineered systems (Kolaczkowski et al., 2005). Serious criticism has, however, been leveled against many current HRA techniques. These critiques can be briefly summarized in the following points: Human performance is typically too complex to be represented by models used for component and system reliability, that is, human actions and performance cannot be decomposed in a mechanical fashion. Description of actions in terms of binary success or failure states (e.g., by event trees) also cannot capture the full range of human performance, and consideration of contextual factors, a.k.a. performance shaping factors (PSFs), is often difficult within simplified (and mathematically tractable) models of human performance.

These shortcomings are shared by many different HRA techniques that Hollnagel (1998) has labeled collectively as ‘first-generation HRA’. Hollnagel (1998) also proposed an alternative approach to HRA, a ‘second-generation’ technique, which explicitly and specifically considers the context of human performance and impact of contextual factors on human reliability, the Cognitive Reliability and Error Analysis Method (CREAM). CREAM is meant to be both a predictive and retrospective analysis tool, that is, fully bidirectional, applying the same principles for either direction of analysis. The logic of this approach is that the same theory that goes into predicting an event must be consistent with explaining a past event. CREAM represents a departure from hierarchical classification for human error causes, which often fail to produce consistent hierarchies of human performance due to insufficient knowledge about the causes of human error, and which tend to be sequential (Hollnagel, 1998).

Instead of the hierarchical and sequential approach, CREAM uses a recursive approach. Several tables interlink into each other through a set of rules for analysis, and each step offers several possibilities for proceeding. The context in which the erroneous action occurs helps simplify the analysis by helping indicate the more likely paths. This is an important consideration, for human actions never occur in a void but always within the context of the situation. CREAM organizes interactions between the human and the environment using the Man–Technology–Organization (MTO) triad. The conditions that shape context are called ‘Common Performance Conditions’ (CPC), and resemble the traditional PSFs of other methods, such as the Technique of Human Error Rate Prediction (THERP; cf. Swain & Guttman, 1980).

The CREAM technique, too, suffers from various shortcomings that limit its applicability in real-world, operational settings (e.g., power plants). The method is quite tedious and time-consuming when done by hand because more than 15 tables need to be searched at each step and because of the rapid growth of branching points in the analysis. Application of the method also requires a copy of the book (Hollnagel, 1998), which is not written to provide clear procedural guidance to the analyst. Consequently, there are few published human reliability analyses or accident investigations that have used CREAM, or evaluations of the method using well-documented case studies; rather, CREAM appears in the HRA literature merely on a conceptual level (e.g., Kirwan, 1998). Another obstacle for widespread use of CREAM and other HRA methods—in parallel with the inherent tediousness of performing HRA using these methods—is the lack of usable software to track the steps in the analysis and calculate probabilities. Our work sought to directly address the latter concern.

The purpose of this research was to create a software tool to help automate much of the CREAM process. As far as we know, this software is the first of its kind, and it is under continual development. It can, however, in it most recent form be tested using various types of data from various sources. In this paper we describe the development of the CREAM software tool and report results from several HRA analyses undertaken to evaluate both the method and its software implementation.

OVERVIEW OF THE METHOD

CREAM stresses the separation of manifestations from causes when classifying erroneous actions. The manifestation of an erroneous action is physically observable and measurable. CREAM calls this manifestation a phenotype and each of its possible causes a genotype. The reason for separating phenotype and genotypes is simple: it leads to better consistency in analysis among different analysts. If the phenotypes and genotypes are mixed, as when making intuitive judgments of human behavior to describe an action in terms of its causes, then it is difficult to guard the consistency and reliability of the analysis (Hollnagel, 1998). In addition, reversion and revision of the analysis becomes impossible. Furthermore, to avoid a sense of causality that the terms 'cause' and 'effect' imply, CREAM uses the terms 'antecedent' and 'consequent' to describe before and after events. In the following we review both retrospective and prospective analysis methods using CREAM.

Retrospective Analysis

For retrospective analysis, that is, analysis of accidents or event that have already occurred, context is first defined to help identify CPC's, which in turn limit the analysis to probable paths rather than all possible paths. Next, probable error modes (phenotypes) are determined. Hollnagel (1998) provided a listing of error modes (Table 19, p. 179) as general consequents, each having several general antecedents. After choosing a general antecedent, the analyst looks for it as a general consequent in other tables. If the general consequent has no general antecedents, the analysis stops for this branch. If enough information is available to choose a specific antecedent, then the analysis can stop for that branch. Clearly, many branching points become tedious to track by hand.

The tables in CREAM are formally called 'groups', as the listed consequents and antecedents fall under a broader classification. For example, the 'observation' group lists as consequents 'observation missed', 'false observation', and 'wrong identification'. Each of these consequents has a listing of antecedents that may contribute to the consequent. CREAM also makes a distinction between direct links and indirect links to antecedents. A direct link is more likely than an indirect link, but neither can be ruled out entirely. For example, 'observation missed' has several direct links, including 'equipment failure' and 'inattention'. The other general antecedents listed under the other general consequents are indirect links. For example, 'observation missed' has 'fatigue' and 'distraction' as indirect links. At a higher level, the groups are separated within the MTO triad and classify different types of problems. At first glance, the group separation in CREAM appears somewhat arbitrary, but these separations have a theoretical justification based on research showing different stages of cognitive processing (Hollnagel, 1998).

Prospective Analysis

Hollnagel's (1998) description of prospective analysis, that is, prediction of human errors and their consequences in given tasks or procedures, is more vague than that of retrospective analysis; however the method can be determined with sufficient accuracy to perform the analysis. In prospective analysis, the general direction is from antecedents to consequents, that is, the analyst chooses an antecedent and proceeds to consequents. The algorithm for prospective analysis starts the same way as for retrospective. First identify the general consequent in one of the CREAM groups; next, find all instances of it as a general antecedent among the other groups. These antecedents directly link to consequents, and hence are the direct links. Finding indirect links follow a similar procedure. From the starting CREAM group, all the other general consequents can be found in other groups as general antecedents. These indirectly link to general consequents.

CREAM SOFTWARE DEVELOPMENT

After defining the algorithms for CREAM, supporting software tools could be developed. The chosen development language was Javascript for the Firefox web browser. Javascript, coupled with HTML, offered a rapid prototyping solution for user interface design. Making the software exclusively web-based allows for people to quickly and easily try the method without a need to install additional software. Also, the Firefox browser runs on Windows, Linux, and Mac OS X, making this software tool cross-platform compatible and even more accessible to potential users.

The software has undergone several iterations in features, interfaces, and functionality, and continues to be upgraded as we gain experience in its use for various analyses in different domains. The development history of the CREAM Navigator, as well as the most recent version (v0.5.01) can be found online at http://www.ews.uiuc.edu/~serwy/cream/.

The first prototype provided a group navigator for retrospective analysis. It allowed for quick analysis, but required each steps to be recorded on paper. After a few quick tests, the paper tracking method revealed substantial problems as each branch began to overlap physically on paper after only a few levels of analysis. Subsequently, the second iteration offered a digital workspace to allow repositioning of analysis branches. The structure of the analysis now could be visualized and modified easily.

The third iteration restructured the presentation of the CREAM groups, with the later goal of integrating the digital workspace. It also had the first working implementation of prospective analysis; however, its user interface design had to be abandoned because it proved too difficult to use. The fourth iteration returned to the panel paradigm of the first iteration. It improved the presentation layout as well as offered more CREAM-related information. It displayed the group and category information, with the later goal of introducing CPC functionality.

The fifth and at the time of this writing latest version of the Navigator brought major functionality to the project. The ability to load and save analyses, modify CPC's, perform the analysis with a digital workspace, and seek help were incorporated. Figure 1 shows a screenshot of the application in the workspace. The left panel displays ways to proceed with the analysis. The workspace organizes the analysis visually, allowing repositioning of individual nodes, each with its own text box for keeping notes. A tab interface switches between the different functions and keeps major parts of the program contained within one browser window.

The CPCs window allows the analyst to select different levels for each CPC, as well note the analysis for it. The software computes the CPC triplet, a construct by Hollnagel, which translates the CPC levels to one of four operator control modes. Also, the CPC levels are weighted and summed to create a reduced performance profile. This information tells the analyst which paths are more likely and can simplify analysis by reducing the number of branch points. More features and functionality are still under active development. This software will be released under the GNU General Public License (GPL).

EVALUATION OF THE CREAM SOFTWARE

The ability to quickly click through tables greatly helps in evaluating the CREAM method. In our preliminary tests of the technique and the software tool, analyzing an event proved to be relatively easy. By identifying the physical manifestation of the error, CREAM suggests several general possibilities for the error, and these possibilities can be quickly explored. We used several case studies to test and evaluate the CREAM Navigator software. For our first case study, Hollnagel (1998) offered a NYC subway accident from June 5, 1995 as an example for retrospective analysis (see Figure 1).

We also interviewed operators and engineers at the Abbott Power Plant on the UIUC campus for further case studies. Abbott has several dual fuel boilers, operating on either natural gas or fuel oil, and several boilers that burn coal only. The main product of the plant is steam, which is used for a variety of purposes on the campus. Steam is also used to generate electricity by steam turbines and supplied for the campus. We analyze a coal boiler start-up procedure and two recent incidents at Abbott to evaluate the CREAM Navigator software

The New York City Subway Accident from Hollnagel (1998)

On June 5, 1995, a subway train crashed into the rear end of another train at a speed of about 14-18 mph. Reports said that the motorman missed the red warning light, but the train had automatic emergency brakes which should have prevented the crash. The physical manifestation of the error was the collision, which was an error regarding distance. As shown in Figure 1, we selected ‘distance/magnitude’ and entered a brief explanation in the new node. By selecting the new node in the workspace, the list of direct and indirect antecedents updates in the left-hand column of the Navigator interface.

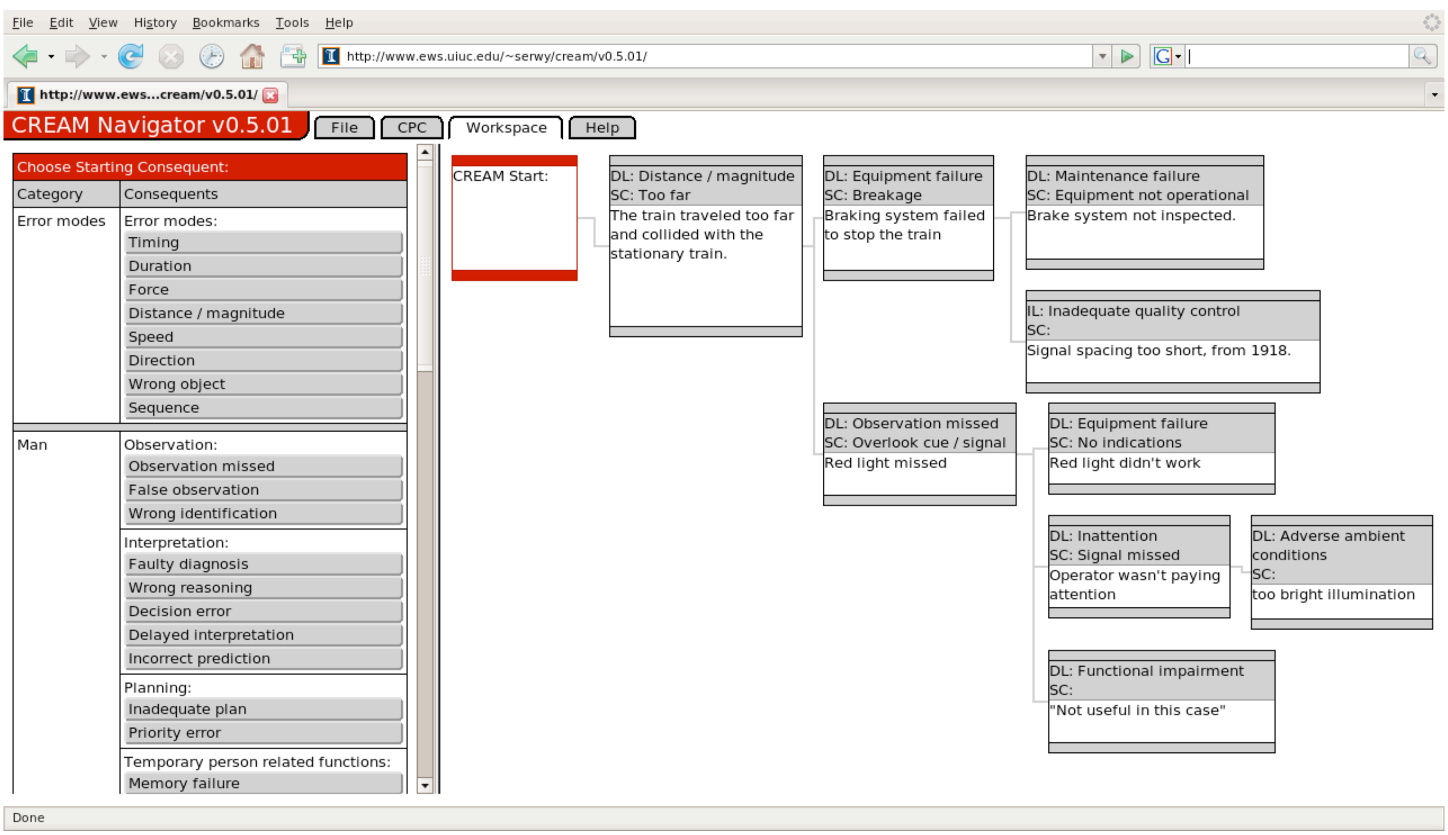

Figure 1. The example of the 1995 New York subway accident provided by Hollnagel (1998) and analyzed by the CREAM Navigator software tool. Clicking on the nodes in the workspace in the right-hand side of the screen updates the relevant tables shown in the column at left, allowing the analyst to explore several possibilities and identify some which may otherwise not come readily to mind. The workspace can be saved for subsequent reviews or changes. The CREAM Navigator also has a text output of the analysis in the right, which may be printed or archived.

Of the suggested ways to proceed, 'Equipment failure', as with the brake system, could explain the failure to stop. From here, a direct link to 'maintenance failure' and an indirect link to 'inadequate quality control' could explain why the brake system failed. Both of these nodes belong to the Organizational group are stopping points since there are no general antecedents to follow. CREAM, as originally presented, intended to identify problems in man-machine interaction, not uncover organizational shortcomings (Hollnagel, 1998). However the CREAM groups can be extended to cover organizational issues.

Returning to the 'distance/magnitude' node, another direct link to 'observation missed', as with the motorman missing the red light, could explain the crash. From here, Hollnagel (1998) mentioned a direct link to 'functional impairment', but explained that it implied the motorman incompetent, which was not the case. The rest of the branches, as shown in Figure 1, should be self-explanatory. It turned out that the root cause of this accident was 'inadequate quality control', an indirect link from 'equipment failure'.

Prospective Analysis of a Coal-Fired Boiler Start-Up

The following case studies were obtained through operator interviews at the Abbott power plant on the UIUC campus. For prospective analysis we chose a coal boiler start-up procedure. The CREAM analysis began with a detailed task analysis of boiler start-up, which takes about nine hours with a boiler supervisor overseeing the entire process. The boiler is checked to ensure that all access doors are closed, 1-1.5 inches of coal is loaded onto the furnace floor grating, wood is loaded on top of the coal and accelerant is added for faster start-up. The fire is transferred to the coal, which has a self-feeding loop system. Temperature is monitored until it reaches the desired level and system is balanced so the equilibrium is reached between the water injected into the furnace and the heat produced by the boiler. At this point the operator can put the boiler in an automatic mode.

Next, context was described by defining CPCs. The resulting CPC triplet showed 1 improved, 6 non-significant, and 2 reduced CPCs, placing the operators in a tactical control mode, and sums of reduced performances pointing the likely paths for analysis were Technology (4), Organization (3), and Man (1). Note that these numbers were not part of original CREAM method, but added in the software to see whether any of the MTO categories could be ruled out as not likely.

According to Hollnagel (1998), predictive analysis should consist of iterative processes that include quantitative analysis that narrows scenarios to highly likely factors to achieve the final result. As quantitative analysis was not possible to be performed in this case, unlimited number of scenarios could be imagined making this analysis very complex. However, each antecedent was evaluated, and possible unsafe scenarios discussed down to the second and the third layer. Although it was not appropriate to examine the antecendents to this depth at this time because analysts are supposed to stop extending a node when they find the same antecedents that appeared in the higher layers of the analysis, it was an intuitive path to follow using the CREAM Navigator.

The outcome of the MTO analysis is clearly compiled in the Navigator but it might be useful to restate it in the workspace view. The students also found it confusing to have the Error Modes as a possible starting point in the prospective analysis (Hollnagel allows for this as a way to start and end analysis with a direct cause for an error) and expressed a desire to have explicit rules for stopping points for analysis along each branch clearly displayed. Means for quantitative analysis also remain to be incorporated into the Navigator.

Retrospective Analysis of a Pump Mix-Up

In the particular analyzed, an operator switched load from one turbine to another; he was also supposed to turn on one feedwater (FW) pump and one circulation (CIRC) pump for the new turbine. However, he actually turned on two FW Pumps, one for the new turbine and another for a turbine that was down for maintenance, leaving the new turbine without CIRC pump. The FW pump supplies water to turbine unit and controls steam/water levels. The CIRC pump circulates cooling water throughout the turbine system and ensures that the turbine is cooled; without it, the turbine would overheat and shut down, possibly causing extensive damage and power loss.

The CREAM analysis of this incident started with definition of CPCs, yielding three reduced ones: (1) Adequacy of organization, due to lack of communication between control room personnel and floor operator; (2) Adequacy of MMI and organizational support, due to cluttered system display with multiple similar windows, which made it highly probable to overlook the pump number and description appearing on top of the window; and (3) Adequacy of training and experience; training is done by word of mouth and standardized operator training program exists.

The control mode was found to be tactical. We found three main error modes through CREAM analysis, with direct and indirect links and specific antecedents:

(1) Error Mode: Wrong Object, Similar Object; operator started a second FW pump instead of a CIRC pump. The menus for different pumps are almost identical in appearance;

- (a) Direct Link: Wrong Identification, Partial Identification; operator incorrectly identified the feed pump menu as the circulation pump menu. He likely jumped to the wrong conclusion with incomplete information from not looking at the label closely;
- (b) Direct Link: Observation Missed, Overlooked Measurement; operator was following a sequence of tasks by clicking 'start' for different pumps and missed the fact that he turned on the wrong pump because he overlooked the pump menu label information;
- (c) Direct Link: Inattention, Signal Missed; operator did not attend to the pump menu label.
    - (i) Indirect Link–Specific Antecedent: Boredom; we assume monotonous tasks lead to boredom, and consequently, to the operator paying less attention.

(2) Error Mode: False observation; False recognition; operator mistook FW pump #9 menu for CIRC pump #8 menu.

- (a) Indirect Link–Specific Antecedent: Habit, Expectancy;

the task was monotonous, and the operator was in the habit of just selecting pumps and clicking 'start';

(b) Indirect Link: Faulty Diagnosis, Wrong Diagnosis; the operator thought he turned on a FW pump and a CIRC pump when he actually turned on two FW pumps.

(3) Error Mode: Ambiguous Information; Coding Mismatch; the coding scheme is the same for all pump menus regardless of pump type of number.

(a) Direct Link: Design Failure, Inadequate MMI; the interface does not help an operator distinguish between menus for the different pump types or numbers.

The 'likely paths for analysis' did not seem to correspond well with actual CREAM analysis. Although the CPC results were 1 for Man, 4 for Technology, and 3 for Organization, the 'Man' category dominated the analysis by having many more links and possible classifications that overlap with the other categories.

CONCLUSIONS

The simplicity of our CREAM software readily allowed for novices such as undergraduate students enrolled in the 'Human Error' class to analyze events in much detail. There is certainly room for improvement in the interface as well as the CREAM tables that underlie the entire system. The suggested antecedents are for a general domain, but they were readily applicable to the NYC train crash example as well as to the Abbott cases. CREAM can thus be a very powerful and easy to use error analysis tool. However, this requires that the tables remain useful in their structure. Hollnagel tried to be as general as possible in creating these tables. By performing more analysis of different errors with CREAM, the value of the method can be assessed and further refinements made. The CREAM Navigator software tool should considerably lower the threshold for undertaking such efforts.

In our evaluations of the CREAM method and the Navigator software, where undergraduate students served as analysts and interviewed expert operators for domain-specific information about the procedures and incidents, it was clearly demonstrated that domain expertise must the foundation for any useful analysis of complex systems. Although the CREAM Navigator was adequately simple to use with minimal training, the discrepancy in domain expertise—and associated communication problems—between the Abbott operators and the students prevented the latter from applying CREAM to its full potential in the cases described above. Therefore, it is clear that further development of both the CREAM method and its software implementations must consider expert operators as the primary users, that is, people who have the necessarily intimate understanding of the systems they interact with but who lack expertise and need much guidance and support in the HRA process.

This conclusion allows for drafting of specific design guidelines for future upgrades of the CREAM Navigator. First, the software must have tightly integrated help function that will guide the aforementioned domain experts—who cannot be expected to have expertise in psychology, HRA, and human factors practice—to apply their systems knowledge in analysis of procedures and incidents. Second, the output from the CREAM analysis must be formatted in a manner that allows for actionable directions for plant managers and engineers to address the inadequacies and antecedents of errors discovered by the analysis. Finally, both the CREAM method and the Navigator software must better facilitate quantitative analyses. Clear procedures for quantitative CREAM analysis are lacking at this time, with Hollnagel (1998) and Fujita and Hollangel (2004) offering different alternatives for quantification of human error probabilities. After all, it is here where the computational prowess of computers can be put to best use, and our hope is that the CREAM Navigator will provide for a useful testbed for research, development, and testing of quantitative methods within the CREAM method.

ACKNOWLEDGMENTS

We wish to thank the following students in the AVI 447/ PSYC 457 *Human Error* course in the spring semester 2007 for performing the CREAM analyses on selected incidents: Stephen Blanchette, Patrick Burke, Scott Butler, Courtney Crump, Christopher Fischetti, Scott Grams, Evan Hunt, Harrison Kaplan, James Keller, Matthew Kump, Garrett Leffelman, Matthew Marcin, Eric Mucha, Vincent Murphy, Seth Redenbo, Ahsan Rizvi, Charles Saloga, Robert Shavers, Anthony Smith, Maria Stadnicka, Takahiro Suzuki, Robert Swartz, Jennifer Tsai, Jacob Wagner, Matthew Wood, and Marius Zymantas.

Many thanks are also due to the engineers and operators at the Abbott power plant, especially Mike Brewer and Gary Watson, for offering their time and expert knowledge to the students performing the analyses.

REFERENCES

Fujita, Y., & Hollnagel, E. (2004). Failures without errors: quantification of context in HRA. *Reliability Engineering and System Safety, 83*, 145–151.

Hollnagel, E. (1998) *Cognitive Reliability and Error Analysis Method*. New York: Elsevier Science Inc.

Kirwan B. (1998). Human error identification techniques for risk assessment of high risk systems—Part 1: Review and evaluation of techniques. *Applied Ergonomics, 29*(3), 157-77.

Kolaczkowski, A., Forester, J., Lois, E., & Cooper, S. (2005). *Good Practices for Implementing Human Reliability Analysis: Final Report* (NUREG-1792). Washington, DC: U.S. Nuclear Regulatory Commission.

Swain, A. D., & Guttman, H. E. (1980). *Handbook of human reliability analysis with emphasis on nuclear power plant applications* (NUREG/CR-1278) Washington, DC: U.S. Nuclear Regulatory Commission.