INL/CON-05-00424
PREPRINT

Human Reliability Analysis in the U.S. Nuclear Power Industry: A Comparison of Atomistic and Holistic Methods

Human Factors and Ergonomics Society 49th Annual Meeting – 2005

Ronald L. Boring
David I. Gertman
Jeffrey C. Joe
Julie L. Marble

September 2005

The INL is a
U.S. Department of Energy
National Laboratory
operated by
Battelle Energy Alliance

Idaho National
Laboratory

HUMAN RELIABILITY ANALYSIS IN THE U.S. NUCLEAR POWER INDUSTRY: A COMPARISON OF ATOMISTIC AND HOLISTIC METHODS

Ronald L. Boring, David I. Gertman, Jeffrey C. Joe, and Julie L. Marble
Idaho National Laboratory
Idaho Falls, Idaho 83415, USA

A variety of methods have been developed to generate human error probabilities for use in the US nuclear power industry. When actual operations data are not available, it is necessary for an analyst to estimate these probabilities. Most approaches, including THERP, ASEP, SLIM-MAUD, and SPAR-H, feature an atomistic approach to characterizing and estimating error. The atomistic approach is based on the notion that events and their causes can be decomposed and individually quantified. In contrast, in the holistic approach, such as found in ATHEANA, the analysis centers on the entire event, which is typically quantified as an indivisible whole. The distinction between atomistic and holistic approaches is important in understanding the nature of human reliability analysis quantification and the utility and shortcomings associated with each approach.

INTRODUCTION

To support safety and risk analysis for US nuclear power plants, component failure and personnel error probabilities are utilized. For events that are rare or for which there are little operating data, a series of methods has been developed to generate reliable and valid estimates of failure probabilities using expert judgment. NUREG/CR-2255 (Stillwell, Seaver, and Schwartz, 1982), NUREG/CR-2743 (Seaver and Stillwell, 1983), and NUREG/CR-3688 (Comer, Seaver, Stillwell, and Gaddy, 1984) outlined methods for using expert judgments to arrive at error probabilities. These approaches explicated paired comparison, ranking and rating, direct numerical estimation, and indirect numerical estimation techniques applied to error estimation, with a particular emphasis on aggregating the estimates from multiple experts. A number of other efforts have sought to formalize and refine available methods specifically to estimate human error [e.g., THERP (Swain and Guttmann, 1983), SLIM-MAUD (Embrey et al., 1984), ASEP (Swain, 1987), ATHEANA (US Nuclear Regulatory Commission, 2000), and SPAR-H (Gertman et al., in press)]. Examples of generalized estimation techniques beyond nuclear power plants are also available [e.g., handling high-level radioactive waste (Kotra, Lee, and Eisenberg, 1996)].

ATOMISTIC AND HOLISTIC JUDGMENT

There are two common types of quantifiable judgment processes—*atomistic* and *holistic*. Atomistic judgments involve breaking a judgment domain into constituent subcomponents. Independent judgments are made about each subcomponent, and later aggregated into a summary judgment. In risk analysis, the atomistic model of judgment is the method employed in estimating human error based on performance shaping factors (PSFs). For example, in the SPAR-H method (Gertman et al., in press), the human error probability (HEP) is the product of the influence of eight PSFs on the default or nominal error rate. Other common atomistic approaches to expert elicitation used within the US nuclear power industry include THERP (Swain and Guttmann, 1983), ASEP (Swain, 1987), and SLIM-MAUD (Embrey et al., 1984).

In contrast, in holistic judgments, a judgment about the overall event likelihood is made. Holistic judgment does not typically weigh individual contributing factors (like PSFs) but, instead, views the event and circumstances as an irreducible whole. In the holistic scaling methods advocated in NUREG/CR-2743 (Seaver and Stillwell, 1983) and NUREG/CR-3688 (Comer, Seaver, Stillwell, and Gaddy, 1984) and later in ATHEANA (US Nuclear Regulatory Commission, 2000) analysts make probability judgments about the likelihood of the event, but the analysts do not explicitly quantify the subfactors that contribute to the overall error probability. In a sense, it is argued that the sum is not the product of the parts but instead is the simultaneous interaction of all parts in the presence of plant conditions. Holistically, this interaction is considered irreducible.

DISADVANTAGES OF EACH APPROACH

It is important to note that each method of expert estimation has documented shortcomings. Atomistic estimation, for example, is known to fail due to insufficient information for completing the atomistic rubric or due to clerical or procedural errors in completing the atomistic forms or worksheets (Hammond, Hamm, Grassia, and Pearson, 1987). Moreover, there is serious difficulty in designing a valid and comprehensive atomistic rubric. A poorly designed atomistic rubric will hamper efforts to arrive at a meaningful representation and quantification of the problem space. Designing a solid atomistic scoring rubric is exacting and time consuming. For example, the SPAR-H worksheets (Gertman et al., in press) represent ten years and three full iterations in terms of development history. This development time was necessary to produce a comprehensive list of PSFs, map the relationship between these PSFs and HEPs, and make adjustments based on analyst feedback and peer reviews. Other atomistic approaches to expert elicitation in the nuclear power arena feature comparably long gestation periods.

Holistic estimation features a similar array of shortcomings. One primary concern is the level of expertise required to perform holistic estimation. Because holistic estimation allows the analyst to evaluate according to their own criteria and impressions, this method is not well suited for novice analysts. The use of novices in holistic estimation results in very inconsistent ratings, as novices may use ad hoc or inoperative judgment processes (Madigan and Brosamer, 1991). A related concern with holistic estimation is that it commonly enlists selective information about the object of investigation. Ettenson, Shanteau, and Krogstad (1987) show that expert analysts tend to focus on selective information about a problem space, to the detriment of other information. For example, one analyst may focus on how the HEP will fit within the plant model construction (i.e., what configurations have been modeled or what subcomponents may have been incorporated into a system, etc.), while another analyst may give more weight to his or her impression of plant conditions as opposed to model representations of those conditions. While this selectivity allows analysts to focus on the most crucial information, it may also hinder the analyst from considering other contributing factors. The holistic method does not explicitly require the analyst to make a broad sweep of the problem space. This means that the analyst, especially the novice analyst, may omit important information when considering the likelihood of an event. The open-ended evaluation criteria of holistic estimation are a significant contributor to the holistic method's low inter-rater reliability reported earlier in this paper.

UNIVARIATE AND MULTIVARIATE JUDGMENTS

An important concept in measurement in the physical sciences centers on the multidimensionality of measurement for any given object. As Kyburg (1984, p. 17) notes:

> *Measurement is often characterized as the assignment of numbers to objects (or processes). Thus we may assign one number to a steel rod to reflect its length, another to indicate its mass, yet another to correspond to its electrical resistance, and so on. It is thus natural to view a quantity as a function whose domain is the set of things that quantity may characterize, and whose range is included in the set of real numbers.*

Any object or event has a multitude of magnitude dimensions in which it may be measured. While in many cases these magnitude dimensions may be orthogonal, they are often interrelated.

The psychological analog to measurement in the physical sciences is psychological scaling, including error estimation in human reliability analysis (HRA) methods. Psychological scaling has univariate and multivariate components. If a single factor contributes to an event, then the judgment of that event is univariate. If a combination of factors contributes to an event, then the judgment of that event is multivariate.

Most reportable events in the nuclear power industry feature a combination of contributing factors. This is, in part, due to deliberate safeguards and redundancies in plant processes, which minimize the chance that a single failure will escalate to become an off-normal event. For example, a maintenance electrician may accidentally reverse the polarity of a switch during installation. This error may be due to the electrician's fatigue at the end of the work shift. However, due to prescribed post maintenance checking, the electrician catches the error before it affects plant operations. If another contributing factor is added to the situation, the likelihood of the error leading to a reportable event increases. For example, if the procedures for installing the switch fail to specify post maintenance testing, the error may compound to affect plant operations adversely. The likelihood of error further increases as additional contributing factors are added. Fatigue and poor procedures might be joined by poor lighting at the switchboard where the switch is being installed. This poor lighting might make it difficult for the electrician to see the color of the wires being installed. Thus, ergonomics would escalate the probability of the error and the failure to correct it before it is put into service.

Table 1 depicts considerations for using atomistic or holistic analysis for univariate and multivariate problems. In an atomistic approach to HRA, the analyst would classify each of the contributing factors to the event. The above example suggests that fitness for duty was low, that the switchboard featured poor maintenance ergonomics, and that there were issues with the procedures used for electrical maintenance. PSFs corresponding to these contributing factors would be flagged, and the HEP would be computed accordingly. The PSF rubric or checklist forces the analyst to consider a variety of factors and therefore minimizes the chance of excluding important contributing factors. But, if the list of PSFs is incomplete or fails to match the actual circumstances of the event, the analyst may overlook an important contributor or may need to use a supplemental holistic approach to model the complete circumstances of the event.

Table 1. Table depicting considerations for using atomistic or holistic judgment strategies for univariate and multivariate problem spaces for an event.

		Judgment Strategy	
		Atomistic	**Holistic**
Problem Space	**Univariate**	*Works well if one of the items on rubric/ checklist matches the cause of the event.*	*Works well if analyst avoids extraneous factors.*
	Multivariate	*Rubric/checklist helps analyst focus on relevant contributors to multivariate events.*	*Prone to inclusion of extraneous factors or scaling biases for multivariate events.*

In a holistic approach, the analyst synthesizes contributors to an event to determine the appropriate error probability. The open-ended nature of the holistic approach affords the analyst considerable flexibility in considering unusual contributing factors that may not be included in an atomistic checklist. Because the holistic approach does not necessarily provide guidance to zero in on common contributing factors, the holistic analyst is more likely than the atomistic analyst to consider factors extraneous to the event outcome.

Further, the holistic approach typically fails to provide clear guidelines for aggregating multivariate contributors to an event. Without a formal procedure, the aggregation of multivariate contributors may exhibit large inconsistencies within an individual analyst and across multiple analysts.

HUMAN RELIABILITY ANALYSIS METHODS

In the following sections, we explore a number of HRA methods developed in support of safety and risk analysis at US nuclear power plants. The discussion provides a brief introduction to each method as well as contextual information regarding the methods' incorporation of atomistic or holistic approaches.

THERP

The Technique for Human Error Rate Prediction (THERP; Swain and Guttmann, 1983) was developed beginning in the 1970s. THERP assumes that types of errors, such as misreading or omitting an instructional step, occur at predictable, constant rates. THERP decomposes a task into subtasks for which errors can be predicted and then applies probabilities of successful completion to each subtask.

THERP begins by decomposing the task of interest using event trees. Tasks are represented as a tree of task steps, which can be successfully completed or be unsuccessful due to error. The error branches of the tree are usually left undeveloped, resulting in a tree having a single success path with errors at each of the steps represented by undeveloped leaf nodes. Subtasks are analyzed at a very low level (e.g., misread procedure step), and an associated error probability is assigned to each failure branch. The probability of task failure is then the product of these assigned values. Because of its focus on task decomposition, THERP is best considered an atomistic approach to HRA.

The tables of error rates used by THERP were composed from meta-analysis of many psychological research studies. They have further evolved through a combination of statistical data and expert judgment. These data are presumed to be accurate within an order of magnitude. In addition to task descriptions, THERP includes a few PSFs such as stress and time available, which are collected to modify probabilities. These probabilities are then used to compute the probability of success for the overall task. However, the emphasis in THERP is on analysis of the task, rather than on assessing the impact of the context on human performance.

ASEP

The Accident Sequence Evaluation Program (ASEP; Swain, 1987) was designed as a simplified alternative to THERP. ASEP's unique features include a detailed screening procedure for pre- and post accident tasks, separate HEPs for pre- and post-accident tasks, inclusion of recovery factors, consideration of the role of diagnosis in error and recovery, tables to account for the influence of available time on the error probability, a simplified three-level account of dependency, and use of tables and software for providing uncertainty bounds. The distinction between THERP and ASEP is characterized by the handling of the HEP. Whereas THERP requires the analyst to calculate the HEP, ASEP provides predefined HEP values (i.e., 3E-2 for pre-accident tasks and 1E-1 for post-accident tasks). The use of predefined HEPs decreases the accuracy of the analysis to the benefit of the ease and simplicity of completing the analysis. This tradeoff results in ASEP having slightly more conservative error estimations than THERP.

The ASEP analyst is able to model the influence of PSFs, and the screening procedure encourages careful consideration of all contributing factors. ASEP essentially works by categorizing errors according to tabular definitions, thus ensuring analyst consistency inasmuch as analysts are able to agree on the categorization. Beyond open-ended PSFs, the ASEP procedure explicitly accounts for time, procedures, immediate response, and stress.

Because of the level of proceduralized detail provided in the ASEP method, it can safely be counted as an atomistic method of HRA. The format of ASEP is not a rubric or worksheet but rather a checklist procedure. The systematic approach to analyzing events by decomposing them into aspects of an analysis is characteristically atomistic. Additionally, the use of lookup tables for calculations eliminates any ambiguity in the assignment of probabilities to events or in the calculation of the overall error probability.

SLIM-MAUD

The Success Likelihood Index Method-Multiattribute Utility Decomposition (SLIM-MAUD; Embrey et al., 1984) was developed for the US Nuclear Regulatory Commission by British researchers as a means of automating some of the mechanics of the earlier Success Likelihood Indexing Method (SLIM). SLIM-MAUD makes use of paired-comparison methods. Assumptions are that similar tasks are grouped for analysis and upper bound and lower bound anchor point HEPs are determined for that class of tasks. PSFs are also rated for their importance and quality. Analysts' expert opinions are used in selecting the bounding HEPs (which are usually taken from THERP or time-reliability curves), in declaring the tasks to be similar enough to be compared, and in assessing the PSFs. The strength of the method lies in the paired-comparison approach, which increases the reliability of the analysis. SLIM-MAUD comes with a set of suggested PSFs, although analysts are free to generate their own. Because of its reliance on PSFs, the method is considered atomistic.

ATHEANA

A Technique for Human Error Analysis (ATHEANA; US Nuclear Regulatory Commission, 2000) was born out of the desire to capture errors of commission and plant context in analysis, two facets of human error that were not accounted for in first-generation HRA methods such as THERP, ASEP, and SLIM-MAUD. Human Failure Events (HFEs) represent basic events that can be modeled in PRA event and fault trees. HFEs are shaped by unsafe actions (UAs) and error-forcing contexts (EFCs), which are modeled separately.

ATHEANA is comprised of ten systematic steps designed to identify the HFE and the contributing factors and context that shaped the event and hindered recovery, and to quantify the error distributions for the HFE. ATHEANA draws on a group of experts who, through a facilitator, identify as many event contributors as possible. This panel of experts then assigns probabilities to the events by selecting the upper and lower bounds of the uncertainty distribution. The individual expert estimations are combined to produce a composite error or failure probability.

ATHEANA is something of a hybrid method of expert elicitation, but mostly resembles a holistic approach. The ATHEANA analysis is open-ended in terms of identifying UAs and EFCs. It does not use an atomistic, pre-defined list of PSFs for characterizing the HFE. Instead, it encourages analysts to explore the event from multiple angles to arrive at a thorough set of UAs and EFCs. This explication of event contributors is similar to atomistic methods, but the subsequent quantification of these contributors is holistic. Typically, atomistic methods of HRA use base error rates that are modified negatively or positively by PSFs. In ATHEANA, no formal guidance is given regarding how the event contributors are used to shape the overall event probability. While guidance is given for defining the scale to be used in the uncertainty distribution, the actual combination of contributors into a single distribution is left to the analysts' discretion.

SPAR-H

The Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H; Gertman et al., in press) was developed by the Idaho National Laboratory as an easy-to-use approach for representing uncertainty information for use in analysis with the probabilistic risk assessment plant models employing Systems Analysis Programs for Hands-on Integrated Reliability Evaluation (SAPHIRE) software (Smith et al., 2000). SPAR-H does not explicitly distinguish among skill-, rule- and knowledge-based behaviors, nor does it specifically rely on the distinction between errors of omission and commission. Rather, it distinguishes slips, lapses, and mistakes as the underlying taxonomy of error types. Because of its assessment of the effects of context on individual components of the task via the use of standardized PSFs, SPAR-H may be considered an atomistic approach to HRA.

SPAR-H combines information processing with stimulus-response to model human performance. SPAR-H also acknowledges the role of environmental factors upon diagnosis and action. For example, during evaluation of PSFs, analysts note whether interactions might be difficult to analyze due to misleading indicators, complexity, time-dependent aspects, and the effects of combinations of unavailable or faulted equipment.

SPAR-H incorporates eight summary operational factors, or PSFs, associated with nuclear power-plant operation. These operational factors can be directly associated with the model of human performance. The eight PSFs are *Stress/Stressors*, *Available Time*, *Complexity*, *Experience/Training*, *Procedures*, *Ergonomics/HMI*, *Fitness for Duty*, and *Work Processes*. These PSFs were derived from reviews of psychological research on human error. In addition, the development of SPAR-H has included significant time devoted to determining the underlying mathematical distributions of these PSFs in terms of their contribution to human error.

The SPAR-H method divides tasks performed by personnel into two components, the processing component (termed *diagnosis*) and the response component (termed *action*). SPAR-H can be used by HRA experts, as well as by nuclear power plant operations experts without an HRA background to screen task steps for the potential for error.

Information gleaned from the initial assessment of the scenario of interest is then used by the analyst to fill out the SPAR-H worksheets. By using the initial assessment of the scenario of interest plus the analyst's understanding of the degree to which the PSFs would have been present in the scenario, SPAR-H walks the analyst through quantification of the potential for human error. The SPAR-H documentation provides definitions of the PSFs as well as examples to allow an analyst who is not expert in HRA to consistently evaluate various scenarios.

SUMMARY

As can be seen from the previous examples, most formal HRA methods currently utilized in the US nuclear power industry take an atomistic approach. The exception is ATHEANA (US Nuclear Regulatory Commission, 2000), which is primarily holistic. Among the atomistic approaches, all but the SPAR-H method (Gertman et al., in press) use a flexible definition of PSFs. In these atomistic approaches, the process of extracting PSFs for each event in an analysis complicates the methods and introduces subjectivity in factor selection that also characterizes holistic approaches. However, proponents of holistic approaches would argue that the use of fixed factors in SPAR-H makes the method inflexible to the meet the demands of all analyses.

The purpose of this article is not to rate the appropriateness of atomistic vs. holistic approaches to HRA. Instead, this article chronicles the use of HRA methods in the US nuclear industry, beginning with first generation atomistic methods such as THERP, ASEP, and SLIM-MAUD, and emerging recently with second generation methods such as the holistic ATHEANA and the atomistic SPAR-H. The balance of atomistic and holistic methods is achieved with these second generation approaches, affording the analyst greater flexibility in arriving at HEP values.

The comparison between atomistic and holistic approaches is informative, because it allows the analyst to tailor his or her estimation approach to best meet the needs of the event under investigation. Yet, this article remains merely a starting point for further exploration. The actual practice of analysts needs to be considered in order to determine what lessons can be learned and applied across both atomistic and holistic domains in the form of best practices. Moreover, a comparison still remains to be made between atomistic and holistic approaches in the international nuclear power industry. The implementation of atomistic and holistic approaches to HRA is an international phenomenon, which requires further exploration to assess the relative merits of each approach. The present article is not an all-inclusive catalog of atomistic and holistic HRA methods. It is, however, a framework to which additional analyses of HRA methods can be appended.

DISCLAIMER

The views expressed in this paper do not necessarily represent the views of Idaho National Laboratory, the US Department of Energy, Battelle Energy Alliance, or its customers.

REFERENCES

Comer, M.K., Seaver, D.A., Stillwell, W.G., & Gaddy, C.D. (1984). *Generating Human Reliability Estimates Using Expert Elicitation, Volume 1. Main Report*, NUREG/CR-3688. Washingtion, DC: US Nuclear Regulatory Commission.

Embrey, D.E., Humphreys, P., Rosa, E.A., Kirwan, B., & Rea, K. (1984). *SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Unstructured Expert Judgment, Volume I: Overview of SLIM-MAUD*, NUREG/CR-3518. Washington, DC: US Nuclear Regulatory Commission.

Ettenson, R., Shanteau, J., & Krogstad, J. (1987). Expert judgment: Is more information better? *Psychological Reports*, *60*, 227-238.

Gertman, D., Blackman, H., Marble, J., Byers, J., Haney, L., & Smith, C. (in press). The SPAR-H Human Reliability Analysis Method. NUREG/CR-in press. Washington, DC: US Nuclear Regulatory Commission.

Hammond, K.R., Hamm, R.M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, *17*, 753-770.

Kotra, J.P., Lee, M.P., Eisenberg, N.A., & DeWispelare, A.R. (1996). *Branch Technical Position on the Use of Expert Elicitation in the High-Level Radioactive Waste Program*, NUREG-1563. Washington, DC: US Nuclear Regulatory Commission.

Kyburg, H.E. (1984). *Theory and Measurement*. Cambridge, UK: Cambridge University Press.

Madigan, R.J., & Brosamer, J.J. (1991). Holistic grading of written work in introductory psychology: Reliability, validity, and efficiency. *Teaching of Psychology*, *18*, 91-94.

Seaver, D.A., & Stillwell, W.G. (1983). *Procedures for Using Expert Judgment to Estimate Human Error Probabilities in Nuclear Power Plant Operations*, NUREG/CR-2743. Washington, DC: US Nuclear Regulatory Commission.

Smith, C.L., Wood, S.T., Kvarfordt, K.J., McCabe, P.H., Fowler, R.D., Hoffman, C.L., Russell, K.D., & Lois, E. (2000). *Testing, Verifying, and Validating SAPHIRE Versions 6.0 and 7.0*, NUREG/CR-6688. Washington, DC: US Nuclear Regulatory Commission.

Stillwell, W.G., Seaver, D.A., & Schwartz, J.P. (1982). *Expert Estimation of Human Error Probabilities in Nuclear Power Plant Operations: A Review of Probability Assessment and Scaling*, NUREG/CR-2255. Washington, DC: US Nuclear Regulatory Commission.

Swain, A.D. (1987). *Accident Sequence Evaluation Program Human Reliability Analysis Procedure*, NUREG/CR-4772. Washington, DC: US Nuclear Regulatory Commission.

Swain, A.D., & Guttmann, H.E. (1983). *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*, NUREG/CR-1278. Washington, DC: US Nuclear Regulatory Commission.

US Nuclear Regulatory Commission. (2000). *Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA)*, NUREG-1624, Rev. 1. Washington, DC: US Nuclear Regulatory Commission.