NUCLEAR REGULATORY COMMISSION

BEFORE THE ATOMIC SAFETY AND LICENSING APPEAL BOARD

In the Matter of PACIFIC GAS AND ELECTRIC COMPANY (Diablo Canyon Nuclear Project, Units 1 and 2)	Docket Nos. 50-275 O.L. 50-323 O.L.

DIRECT TESTIMONY OF GEORGE APOSTOLAKIS

Q. Please state your name.

A. George Apostolakis.

Q. What is your business address?

A. 5532 Boelter Hall, University of California, Los Angeles, California 90024.

Q. What is the purpose of your testimony in this proceeding?

A. I have been asked to render my professional opinion on the applicability of probability theory, decision theory, and statistics to the verification of the design of a nuclear power plant and to evaluate the adequacy of the Independent Design Verification Program (IDVP) to insure the adequacy of the design of Diablo Canyon Nuclear Power Plant, Units 1 and 2. Specifically, my testimony pertains to contentions 1 and 7.

I.

QUALIFICATIONS

Q. What is your present position?

A. I am a Professor in the School of Engineering and Applied Science at the University of California, Los Angeles, where I have taught since July 1974. I am a member of the faculty of the Mechanical, Aeronautical, and Nuclear Engineering Department.

Q. Please summarize your education.

A. I hold a Ph.D. in Engineering Science and Applied Mathematics and an M.S. in Engineering Science, both from the California Institute of Technology. I also hold a diploma in Electrical Engineering from the National Technical University, Athens, Greece.

Q. Are you a member of any professional organizations?

A. I am a member of the American Nuclear Society and the Society of Risk Analysis. I am a past recipient of the Mark Mills Award from the American Nuclear Society.

Q. Please summarize your work experience in the fields of risk assessment and nuclear engineering.

A. For the past ten years, I have been continuously engaged in research in risk assessment, including the conduct of probabilistic risk analyses for nuclear power plants; probability theory, decision theory, and statistics; reliability analyses; and nuclear engineering.

Since 1977, I have served as a consultant to Pickard, Lowe and Garrick, Inc., where I participated in probabilistic risk analyses of the Oyster Creek, Zion, and Indian Point nuclear generating stations; I also served for Pickard, Lowe and Garrick on the technical review board for the Seabrook Probabilistic Safety Study. For the past three years, I have also served as a consultant to the Bechtel Power Corporation on probabilistic risk assessment. In the past I have served as a member of the Peer Review Panel for the Load Combination Program of the Lawrence Livermore National Laboratory, as a consultant to the Seismic Safety Margins Research Program of Lawrence Livermore National Laboratory, as a consultant on risk methodology for geologic disposal of radioactive waste for the Sandia National Laboratories, and as a member of a research review group for the Probabilistic Analysis Staff of the U.S. Nuclear Regulatory Commission.

My research work at UCLA has been both theoretical and applied. I have conducted research on the foundations and methods of probabilistic risk analysis, on data analysis, on fire risk analysis, and the general area of risk-benefit. I have developed and taught two courses on probabilistic risk analysis. I have also taught courses in nuclear engineering as well as basic engineering courses.

Q. Do you regularly publish in the professional literature?

A. Yes. I have edited one book and contributed to another on risk analysis. I have published numerous articles on probabilistic risk assessment, nuclear engineering, and related matters. I also serve as a reviewer for <u>Nuclear Safety</u>, <u>Nuclear Science and Engineering</u>, <u>Nuclear Technology</u>, <u>IEEE Transactions on Reliability</u>, <u>AIChE Journal</u>, <u>Risk Analysis</u>, and <u>Reliability Engineering</u>. The list of my publications has been submitted separately in my affidavit of qualifications.

/

II.

PROBABILITIES AND STATISTICS

Q. What do you mean by statistical inference?

A. Statistical inference is the process by which evidence is incorporated in our body of knowledge. This body of knowledge is, in general, expressed by probabilistic statements.

Q. How is evidence incorporated in our body of knowledge?

A. I view this question in the context of the Bayesian (or Subjectivistic) Theory of Probability. According to this theory, we always have some degree of knowledge of any uncertain event of interest. Bayesian Theory asserts that our degree of knowledge can be expressed in terms of probabilities. As information becomes available, we modify our state of knowledge; that is, we revise our probabilities. This modification is done in a consistent manner, using Bayes' Theorem.

Q. What do you mean by "evidence"?

A. "Evidence" can be any kind of information. This includes what is commonly referred to as "statistical evidence" as well as such qualitative information as opinions of people, scholarly literature, the results of experiments, etc.

Q. What does the term "statistical evidence" mean?

A. For present purposes, I use the term "statistical evidence" to refer to information concerning the frequency with which a given attribute is observed in a specified population. This would include how many redheads we find in a given group of people, the number of times a coin turns up heads in a sequence of tosses, the proportion of American families within a given income bracket, and so on.

Q. What is the relationship between frequencies and probabilities?

A. Frequencies are observable quantities in a given sample or population. Often we express a frequency as a proportion of a sample or a population. Probabilities, on the other hand, are not observable. They are numerical measures of degrees of belief. In other words, frequencies are objective facts and probabilities are subjective beliefs.

Q. What is the distinction between probability theory and statistics?

A. Statistics is part of probability theory. Probability theory is a set of rules that, if obeyed, guarantee coherence. Statistics is that part of probability theory that deals with the coherent use of evidence.

Q. What do you mean by "coherent"?

A. Human beings dealing intuitively with uncertainty have been found to make inconsistent and unreliable use of the information at their disposal. Probability theory, or, more generally, decision theory, requires them to make their reasoning process, their assumptions, and their use of information consistent with certain principles of rational behavior. This makes the decision process explicit and visible.

/

Q. What is the virtue of making the process explicit and visible?

A. Probabilities are inherently subjective, as are decisions made under uncertainty, leading to differences of opinion among people. By making the process explicit and visible, we allow people holding different opinions, and third parties observing the differences, to approach resolution of the differences on a reasoned basis.

Q. What is the nature of the differences in opinion among people?

A. People differ in their assessments of probabilities. They also differ in their assessments of the costs and benefits of different consequences of decisions.

Q. What are the reasons for different probability assessments?

A. Different decision makers may have different states of knowledge. In addition, there is evidence that human beings have great difficulty expressing their knowledge in terms of probabilities.

There is a substantial body of evidence indicating that people perform poorly in assessing probabilities, that is, in dealing coherently with a body of incomplete evidence. For example, Slovic, Fischhoff, and Lichtenstein, in their article "Facts and Fears: Understanding Perceived Risk" (published in Societal Risk Assessment, R.C. Schwing and W.A. Albers, Jr., Editors, Plenum Press, 1980), state, on the basis of their own experiments and research and those of others, that people tend to deny uncertainty, misjudge risks, and express unwarranted confidence in their judgments. The same authors show that expert assessments are also susceptible to biases, particularly underestimation of risks.

Kaplan, Garrick, Duphily, and I found similar evidence of expert underestimation of failure rates in a study we did of the performance of several components of a nuclear plant. We found, somewhat to our surprise, that the statistical evidence of failures at that plant indicated substantially higher failure rates that the experts had predicted. (Apostolakis, Kaplan, Garrick and Duphily, "Data Specialization for Plant Specific Risk Studies," Nuclear Engineering and Design, 56:321-329 (1980).)

For rare events the difficulties people have assessing probabilities can lead to dramatically different opinions. Of course, this is one area where statistical evidence can be most useful. Bayes' Theorem tells us that when statistical evidence is strong, the prior beliefs (i.e., beliefs prior to obtaining the statistical evidence) become unimportant and the probability assessments are controlled by this evidence, that is, they are independent of the assessor. All this, of course, assumes that different assessors interpret the evidence in the same way, something that is not always true.

III.

DESIGN ERRORS

Q. Has there been any formal research done on the frequency and significance of design errors in nuclear power plants?

A. Yes. Three studies are particularly pertinent here:

/

(1) J. R. Taylor, "A Study of Failure Causes Based on U.S. Power Reactor Abnormal Occurrence Reports," in Reliability of Nuclear Power Plants (Proceedings of a Symposium, Innsbruck, April 14-18, 1975), pp. 119-130, Unipub, Inc., N.Y., 1975. Taylor studied Abnormal Occurrence Reports (now known as Licensee Event Reports (LERs)) submitted to the Atomic Energy Commission and found that a large proportion of the failures in U.S. plants involved design, installation, and operation errors, with an unexpectedly large proportion of the incidents involving multiple failures. Of 490 failures, he classified 36 percent as being due to design errors. The largest single cause of design errors was found to be unforeseen conditions.

(2) T. M. Hsieh and D. Okrent, "On Design Errors and System Degradation in Seismic Safety," in Transactions of the 4th International Conference on Structural Mechanics in Reactor Technology, San Francisco, Calif., August 15-19, 1977, T. A. Jaeger and B. A. Boley (Eds.), Vol. K, Paper K9/4, Commission of European Communities, Luxembourg, 1977. Hsieh and Okrent investigated the possible number and influence of seismic-related design errors by examining the historical record of such errors for a specific reactor. Their estimates of the core melt frequency were substantially higher than those of the Reactor Safety Study (WASH-1400), which had not taken into account the possibility of design errors.

(3) P. Moieni, G. Apostolakis, and G. E. Cummings, "On Random and Systematic Failures," Reliability Engineering, 2:199-219 (1981). We analyzed the LERs for two power reactors plus 100 design errors compiled by Oak Ridge National Laboratory. We found that 18 percent of all licensee events at one of the two reactors and 13 percent at the other were due to design errors. We found that the most common design error was the failure to foresee environmental conditions. That design error alone accounted for nearly as many LERs as all operational procedure errors.

It is important to keep in mind that these results are based on each group of researchers' definitions of the term "design error" and on their interpretation of the events reported. Despite these reservations, there is a great deal of useful information in these studies. For example, they show that design errors are a more frequent cause of failures in nuclear power plants than has been widely assumed.

Q. What are the typical causes of design errors in nuclear power plants?

A. The cited studies indicate that major causes appear to be unforeseen environmental conditions, specification errors, and wrong analyses.

Q. Do these studies show that design errors are inevitable or widespread in commercial reactors?

A. Not necessarily. Each of these studies has examined previously identified operational failures and classified them in various ways. There is no evidence from which one could conclude how representative the plants experiencing these events are of all commercial U.S. reactors. I know of no study of how frequent design errors are in general and of what their impact on the margin of safety is.

So while these studies show that design errors are a more significant factor in plant failures than was previously thought, they do not tell us how frequent and how important to safety such errors are.

Q. Is there any basis for evaluating the safety significance of the design errors described in the literature?

A. One must be very careful about the meaning of the term "safety significance." If by that we mean actually causing injuries to the public, then none of the errors were safety significant. But if we are speaking about an error having the potential for such harm under possible conditions that were not actually experienced before the error was detected, then it is more difficult to dismiss any error as not being safety significant.

I think that the most meaningful way to investigate these issues is based on the reduction in the presumed margin of safety. The only way I know to practically evaluate the safety significance of an error in these terms is to conduct a probabilistic risk assessment. This enables one to test the sensitivity of a given facility to designated system and component failures. In my experience, PRAs sometimes reveal failure paths not perceived by knowledgeable engineers involved in the design of the plant. Furthermore, the potential of multiple failures of redundant components due to design errors cannot be fully assessed without a PRA.

Q. In the probabilistic risk assessments with which you are familiar, how have design errors been treated?

A. Design errors have been treated only indirectly. By this I mean that, while something is usually done, the analysis is not as rigorous as other parts of PRAs are. For example, Appendix X to the Reactor Safety Study (WASH-1400, NUREG 75/014, October 1975) is entitled "Design Adequacy." The study team felt that they needed additional assurance that certain components would function as intended under severe conditions. Part of the reason for this was that the failure-rate distributions did not reflect experience with such environments. The design adequacy assessment was performed by the Franklin Institute Research Laboratories, which checked a sample of components, systems and structures. They found only minor problems, e.g., errors in assumptions used to calculate stresses and inadequate tests. The consequence of these errors was assessed to be a reduction in the safety margin.

In more recent PRAs, like those for the Zion and Indian Point nuclear power plants, the issue of design errors was in the minds of the analysts when they quantified their judgment, so that very low values for failure rates were avoided. Design errors were part of the "other" category of failure causes, which means, causes not explicitly quantified. The notion of the "other" category has been proposed by Kaplan and Garrick (see Risk Analysis, vol. 1, p. 11, 1981), who were among the principal investigators performing these PRAs.

IV.

VERIFICATION OF DESIGN USING PROBABILITY THEORY

Q. Do you know of any case where the adequacy of a nuclear power plant's design was demonstrated using sampling?

A. No. There have been the studies of design errors I described above. But to the best of my knowledge, no nuclear power plant has ever been licensed using a sampling verification program as a substitute for a quality assurance program that was found to be inadequate.

Q. What is the significance of the decision to verify the design by sampling?

A. Ordinarily, licensing decisions are framed in deterministic terms, i.e., does the plant design comply with the NRC criteria? A relatively straightforward answer to this question could be obtained by checking the entire design and fixing any errors found. If one decides to verify the design by sampling less than 100 percent of the design, then one transfers the problem into the realm of probabilities, i.e., one is assessing the probability of an affirmative answer to the original question regarding compliance with the NRC criteria. In other words, one is no longer asking the deterministic question, "Does the design meet the licensing criteria?" Instead, one is asking, "What is the probability that the design meets the licensing criteria?" Or, more precisely, one is asking, "What is the probability that there are no deviations from the criteria in the existing design?"

The nature of the problem has now been considerably changed. One is now explicitly accepting the possibility of a deviation from the licensing criteria remaining undetected.

Q. Can statistical techniques make a contribution to a program to verify the design of a nuclear power plant?

A. Yes, given my earlier discussion of statistics as part of probability theory. Once the decision has been made to characterize the problem in probabilistic terms, statistical techniques enable us to make full use of the information that we have available and furnishes the discipline and guidance that insures we are using the data properly.

Q. How do statistical techniques do so?

A. These methods can provide guidance to the decision maker concerning both the qualitative aspects of the problem (e.g., what kinds of errors have been made, what can be done about them, etc.) and the quantitative aspects (e.g., how likely errors of a certain type are, how many errors remain undetected, etc.)

In this way, probability theory and statistics further the goal of making the analysis and evaluation explicit and visible.

Q. Is it possible to estimate the frequency of design errors in a nuclear power plant using statistical techniques?

A. Yes. Again, one has to be very careful with one's terminology. Because there is no general definition of "design errors," a definition would have to be established at the outset of the study. The definition would have to correspond to the purpose of the study and be precise enough to permit consistent classification of observations. These requirements are not substantially different from the requirements for any engineering study, whether or not statistics are used.

Assuming, however, that we are working with well-defined events, like selecting the wrong design pressure, we could, then, consider the universe of such selections and apply random sampling to estimate the frequency of such errors.

Q. What is a "random sample"?

A. A random sample of a population is one in which each element of the population has an equal chance of being drawn for the sample.

Q. What is "judgmental sampling"?

A. This is not a term I had encountered before my involvement in this case. I gather from the IDVP materials I have read that the IDVP uses this term to refer to the process of selecting elements from the population by using engineering judgment.

Q. Are both kinds of sampling used in statistical analysis?

A. There are places for the use of informed judgment, including engineering expertise, in a statistical study. For example, judgment is used to formulate hypotheses. However, once a

/

population is identified for study, samples are drawn from the population randomly.

Q. Why?

A. In statistical terms, any sample that is not drawn randomly is suspect of biases. Once one departs from random selection, the danger exists that the selection mechanism contains a bias, presumably unintended, that will lead to an unrepresentative sample and results that cannot validly be generalized to the population from which the sample was drawn.

Q. Can you state a pertinent example?

A. There are many well known examples of biased samples rendering invalid results. One of the best known is the Presidential preference poll taken by the Literary Digest before the 1936 election. Over two million respondents to the poll showed a preference for Landon over Roosevelt by a 57% to 43% margin. In the election, President Roosevelt got 62% of the vote.

Any time one departs from random sampling one hazards similar errors. For example, it has been stated that the IDVP sampled the Diablo Canyon design work emphasizing complex designs on the assumption that those were the designs where errors were most likely to be found. However, it is entirely possible that the managers who oversaw the design work recognized the complex problems and assigned them to the most competent engineers and designers. If so, sampling in this way could underrepresent the work of those people most likely to make errors.

Q. Are you saying that what the IDVP calls judgmental sampling has no place in a design verification program?

A. No. If one has information leading one to suspect the location or type of errors, that information should be exploited. But I do not believe that a sample drawn non-randomly can validly be used to generalize about the frequency of errors in the unsampled portion of the population.

V.

EVALUATION OF THE IDVP

Q. What have you reviewed concerning the Diablo Canyon Independent Design Verification Program?

A. Parts of the Phase II Program Management Plan, the IDVP Final Report, NUREG-0675 (Safety Evaluation Report, Supplement 18), the IDVP Program Management Plan for Phase II, Interim Technical Reports 1, 8, 34, and 35, and certain depositions and interrogatory answers.

Q. What is your understanding of how the IDVP sought to verify the adequacy of the non-seismic design?

A. Three systems were selected (the auxiliary feedwater system, the control room ventilation and pressurization system and the safety-related portions of the 4160-V electrical distribution system). I am told that the IDVP verified completely the design of these systems in Unit 1. The IDVP examined the design of these systems and identified errors. It grouped these errors into classes according to whether or

/

not the errors caused criteria or operating limits to be exceeded.

The IDVP then sought to group some of these errors into "generic concerns." Five generic concerns were raised and all systems where these could apply were verified. No other samples were taken.

On the basis of this examination, the IDVP drew conclusions about the adequacy of the overall design of Unit 1, including the systems not sampled.

Q. In your opinion, did the IDVP proceed in an appropriate way?

A. It is not clear to me why they chose to sample and use probabilistic arguments rather than a full deterministic review. Given, however, that they decided to sample, the available statistical methods, particularly random sampling, that would justify extrapolation of their findings to parts of the plant not sampled, have not been used.

Q. In your opinion, was the IDVP's judgment concerning the five generic concerns sound?

A. I do not have enough information to judge. I do recognize that issues like this involve extensive use of judgment. Therefore, different analysts may classify errors in many different ways. Nevertheless, I find the presentation of the IDVP's classification unconvincing.

For example, the selection of system design pressure, temperature, and differential pressure across valves is identified as a generic concern. I can see a more general concern being the selection of system design parameters, which would also include other variables, such as stress, enthalpy, humidity, etc. Since the literature I cited above suggests that incorrect selection of design parameters in general is a common source of errors, I find no adequate justification for limiting this generic concern to incorrect selection of pressures, temperatures, and differential pressures across valves.

As a second example, it is stated on page 6.3.4-2 of the IDVP Final Report that three EOIs (8001, 963 and 1069) involve the misapplication of computer programs. Because there was no commonality between the programs involved in EOI 8001 and the other pair, and because the types of errors were different, a generic concern was not identified. It may be reasonable, however, to identify "misapplication of computer codes" as a generic concern.

Q. What is the significance of the fact that the IDVP found what it called "random errors," that is, errors that were not covered by the five generic concerns?

A. If the three sampled systems were really representative of the unsampled systems, this implies that there are similar errors remaining to be found in the unsampled parts of the plant. On the other hand, if the three systems are unrepresentative, we have almost no information about the unsampled elements of the design and no basis for confidence in the adequacy of the design.

Q. Is the safety significance of the errors uncovered relevant?

A. It depends on what the issue is. If the issue is whether the plant's design meets licensing requirements, safety significance of the design errors is not relevant.

If the issue is the safety of the plant, then safety significance of errors is obviously relevant, but, as I stated earlier, the only way I know to perform such an evaluation is in the context of a PRA.

Q. In your opinion, does the IDVP's work provide a basis for estimating the number of as yet undetected design errors?

A. No. The failure to use random sampling techniques makes a reliable extrapolation impossible and creates the suspicion that there may be errors whose types are not known yet. Furthermore, the same lack of random sampling does not allow the estimation of error frequencies or absolute numbers. The design of the IDVP was not amenable to providing a basis for estimating frequencies.

Q. Does the IDVP provide a basis for concluding that the rate of undetected errors is acceptable?

A. No. To decide that a given rate of errors is acceptable, one must know two things: what the rate of errors remaining in the plant is and what rate is acceptable. For the reasons I have just given, one cannot get from the IDVP's work an estimate of the rate of remaining errors at Diablo Canyon. And nowhere have I seen anyone attempt to set and justify an acceptable rate. The decision that I identified earlier, namely, to recast the problem in probabilistic terms has created the need to have a criterion for acceptability. The issue of an acceptable rate of design errors has not been studied and resolved.

Q. Could one not attempt to set a rate that provides reasonable assurance of safety?

A. The term "reasonable assurance" is not defined. This term is usually used in NRC regulatory matters to refer to the level of assurance sought in setting the design criteria. Thus, we say that the criteria, if met, will provide a reasonable assurance of safety. It would be a significant departure to talk about a reasonable assurance that the criteria are even met. Then one is talking about a reasonable assurance of meeting license criteria that, if met, would provide a reasonable assurance that the plant is safe. This is a novel notion, the implications of which are not obvious.

Q. What can be said about the adequacy of Diablo Canyon Unit 2 from the verification program for Unit 1?

A. I have already said that the findings of the IDVP in Unit 1 cannot be generalized to the portions of Unit 1 not examined. That is obviously true of Unit 2, for which the IDVP does not have a sample at all.

Q. Do we know whether the rates and distribution of errors in the two units are the same?

A. No. We know of certain similarities and certain differences between the two units. To be able to say anything about the error rates in the two units, random samples would be needed from both units.

/

Q. What can now be done to achieve confidence in the design of Diablo Canyon?

A. As a first step, the decision to cast the problem in probabilistic terms should be fully understood. Given the decision to verify by sampling, the objectives of the study and the decision criteria should be explicitly stated, and the populations should be defined. Random samples should be drawn to determine the nature and frequency of the errors. This would permit one to draw valid conclusions about the design as a whole.

VI.

CONCLUSION

Q. How would you summarize your evaluation of the IDVP's work?

A. In general, it appears that a great deal of good engineering work has been done. In my opinion, the greatest weakness of the IDVP effort has been its failure to recognize the implications of the decision to cast the verification program in probabilistic terms and its failure to use the principles and methods appropriate to a probabilistic analysis. These shortcomings are particularly manifested in the lack of explicit and visible decision rules and the failure to use random samples.