"Acceptable Risk": THE CASE OF NUCLEAR POWER

Baruch Fischhoff

Abstract

The key policy question in managing hazardous technologies is often some variant of: "How safe is safe enough?" A typical response of regulatory agencies has been to lay down minimum requirements for how hazardous facilities should be built and operated, without specifying the level of safety that it is hoping to achieve. The U.S. Nuclear Regulatory Commission, charged with regulating safety in nuclear power plants, has recently tackled the safety question directly, by adopting "safety goals" that facilities must meet. The NRC's approach proves to be sophisticated in some respects, incomplete in others. More generally, it points up the inherent difficulties that exist with the concept of "acceptable risk" and with any attempt to build policy instruments around it. Lessons from the NRC case apply to other hazardous technologies, as well as to public policies unrelated to safety.

The U.S. Nuclear Regulatory Commission has recently adopted safety goals for nuclear power plants.[1] These goals express the commissioners' appraisal of what constitutes "acceptable risk" for operating plants. In principle, these overall goals could supplant the myriad of specific decisions that the commissioners currently make regarding the adequacy of hardware and procedures. It would not matter how a plant was built and run, as long as it was safe enough. For the moment, though, the standards are envisioned as providing only guidance, helping the Commission's specific decisions to be—and to be seen to be—more coherent. However, even in this modest initial role, the safety goals represent a marked change in how the Commission manages its affairs. That change could have important implications for the nuclear industry and for other industries that are subject to government safety regulation.[2]

The validity of the Commission's goals, or of any other attempt to set safety standards, depends upon the answers to three questions[3]:

Journal of Policy Analysis and Management, Vol. 2, No. 4, 559–575 (1983)

Published by John Wiley & Sons, Inc. CCC 0276-8739/83/030559-17$02.70

- Does it make sense to use safety goals to manage the technology?
- Can the particular goals that are selected be justified?
- Have the goals been made operational in a way that is faithful to the underlying safety philosophy?

The goals adopted by the NRC commissioners in January 1983 are not yet binding on the Commission. Here, however, they are interpreted literally, in the belief that the only way to understand what a standard means is to look hard at exactly what it says. In addition, an analysis of these goals offers a model for similar analyses of other standards, whether for safety or for any other public objective.

THE SAFETY GOALS

In initiating the safety goals effort, the NRC seems to have had three main concerns, each directed primarily (although not exclusively) at a different audience.

For the general public, the NRC wished to provide an explicit statement of its overall safety philosophy. The content of that statement would confirm that the public's welfare was central to the Commission's decisions. The statement would also give the public some criteria for monitoring those decisions.

For the industry, the goals promise two kinds of regulatory relief. One is reducing the number of regulatory actions. Once a plant was declared safe enough, there would be less pressure to add new safety devices or to introduce new operating procedures. Secondly, by setting the requirements in terms of goals rather than technical specifications, the industry would have the freedom to seek the least costly route to compliance.

For the Commission's own staff, the goals would clarify the role of probabilistic risk analysis (PRA) in the staff's work. PRA is the generic term for procedures that decompose a complex technical system into its various components, then assess the probability of the overall system failing by analyzing the behavior of its parts.[4] PRA has become an integral part of the design process for nuclear power and other technologies. Moreover, the results of such analyses are often consulted informally by policymakers hoping to get a feel for a system's overall safety. As we shall presently see, PRA assumes a critical role in carrying out the NRC's goals.

The NRC statement contains six interrelated goals. The first two are *qualitative* safety goals, expressing the Commission's safety philosophy: One provides that "individual members of the public should . . . bear no significant additional risk to life and health" as a result of nuclear power plant accidents.[5] The second qualitative goal asserts that "societal risks to life and health" associated with the operation of nuclear power plants should be no more than the "risks of generating electricity by viable competing technologies and should not be a significant addition to other societal risks."

Two "quantitative design objectives" attempt to translate these qualitative goals into more operational terms. One states that the "risk to an average individual in the vicinity of a nuclear power plant of prompt fatalities that might result from reactor accidents should not exceed 0.1% of the sum of prompt fatality risks resulting from other accidents to which members of the U.S. population are generally exposed." The second provides that the risk of death from cancer for people close to the plant should not exceed 0.1% of their cancer risk from all other causes.

The two remaining goals supplement the four already described. There is a "benefit-cost guideline" for evaluating safety improvements that might be used in meeting the design objectives. An improvement should normally cost less than $1,000 for each person-rem of radiation that the improvements avert.[6] (A "person-rem" is a standard measure of exposure to radiation.[7]) Finally, the "plant performance design objective" states that the probability of a large-scale core-melt in a year of reactor operation may not exceed one in 10,000.

These goals are to be evaluated over a two-year trial period, during which all prior existing regulations will remain in place. At the end of the trial period, the Commission will decide whether to incorporate the safety goals in its regulatory procedures and, if so, how existing regulations will be adjusted.

DO SAFETY GOALS MAKE SENSE?

Until the NRC launched its new policy, it managed safety by choosing between different designs for nuclear power plants and their operation. Through those decisions, the NRC attempted to identify, at a given time, the design offering the best mixture of costs, risks, and benefits. Those choices were subject to revision as superior designs became available. Because they involve tradeoffs, each such choice is a political act, reflecting someone's values regarding the relative importance of costs, risks, and benefits.

The safety goals replace this case-by-case decision-making with a set of uniform standards that pass judgment on individual plants. All may be found acceptable or none may be. Although setting the goals is a political act, their application should be entirely technical. Engineers will assess whether plants are in compliance, after the policymakers have decided what compliance means. The NRC, therefore, has begun to move from case-by-case decision-making toward standard setting. Elsewhere, I have pointed out that the relative superiority of standards over decisions can be determined in any situation by seeing how well a series of explicit conditions apply to it.[8]

The power of the regulatory agency is limited, so that it is not empowered to make the big policy decisions. The key energy questions facing the United States are: How much? and What kinds? However, neither the NRC nor any other regulatory body is empowered to make those big decisions. The best that the Commission can do is to manage the technology within its jurisdiction. For that purpose, a standard will do just fine, functioning as a sort of consumer labeling: This product has passed these tests.

No choice between technologies is required. Just as the U.S. need not choose a single best energy source, the NRC need not choose a single best design for reactors or their components. Indeed, in both cases, diversity is probably desirable. Thus, the NRC can and should avoid making the effort of ranking all options in order to choose the best one. From this perspective, too, a standard is all that is needed.

Predictability is desired. As long as the Commission prescribes specific hardware and procedures, the industry has difficulty predicting how, or even when, the demands on it will change. One reason for the unpredictability in the current process is that new technical information constantly appears, suggesting the need to review old decisions. A second reason is that the Commission's capacity for consistent analysis is strained by the volume and variety of the decisions it must make. If the new safety goals do displace the existing technical requirements, they may provide a framework that makes the Commission's actions highly predictable. In the short run, though, because they supplement rather than replace current regulations, the goals will just complicate matters.

Regulators hope to shape future options. The safety goals should promote innovation in reactor engineering, by allowing designers to find the most cost-effective way to meet them. Conversely, they should avoid the premature standardization that often comes with rigid specification of how plants must be built and operated. Whether these potential advantages of safety goals will be realized depends on how effectively PRA can be applied. In practice, designers might not even consider innovations that are hard to evaluate with PRA, such as better training or pay for plant personnel.

Competing technologies fall into the same jurisdiction. When all alternative technologies face the same constraints, competitors are on an equal footing. The safety goals, however, apply only to nuclear power—and not to other means of generating or conserving energy. To the extent that the safety goals help the nuclear industry to achieve compliance in a cost-effective way, they should improve its competitiveness. However, it is unclear how the burden imposed by the goals compares with the burden imposed on other energy technologies, by regulatory agencies or market forces. Requiring that nuclear power risks should be no more than those associated with "viable competing technologies" is an attempt to achieve equivalence indirectly by measurement against a category that lies outside the NRC's jurisdiction. As discussed below, ambiguities about the meaning of that phrase leave the success of this attempt unclear.

The facilities that are subject to the goals are similar. Whatever the NRC's success may be in drawing other energy technologies into its web, the safety goals will at least be applied to all nuclear power plants. Despite their differences in design, most nuclear plants produce electricity with roughly the same cost per unit, reliability, and capital requirements. Accordingly, they are sufficiently similar to one another that it is possible to create a standard that treats all about equally. Thus, a single standard could apply roughly the same safety philosophy to all.

Issuing an explicit policy statement is attractive, independent of its impact on safety. Until the safety goals have operational consequences, they will "merely" enunciate the NRC's safety philosophy. However, if such a gesture increases public and industry confidence in the regulatory system, it would be a legitimate use of standard setting. These benefits would evaporate, though, if the philosophy is unpopular or the goals have no clear impact.

Value issues are too contentious to be resolved explicitly. Although the safety goals deal implicitly with many issues in which values play a major role, they wholly avoid some and deal with others indirectly. They do not, for example, provide an expression of the Commission's views on the value of human life or on the distributional effects of nuclear power. Standards can be used to avoid such issues, ones for which there is no societal consensus, or which the agency's mandate does not allow it to address. In such polarized arenas as the nuclear debate, the warring parties may be able to agree on a compromise standard, even when they cannot agree on fundamental principles.

The resources available for decision-making are limited. The number and diversity of the NRC's decisions strain not only the agency's resources, but also those of the other parties involved. For each decision, industry representatives, environmentalists, legislators, local residents, and others must identify their own interests and fight to have them considered. The development of safety goals offers an opportunity to address the central issues in a concentrated, public forum. Moreover, that debate deals with the safety philosophy directly, not through the filter of engineering specifications.

The decision-making process is not important. Application of the safety goals would fall entirely in the domain of technical experts, using analyses and computer codes that are impenetrable to all but the most determined outsiders. By contrast, the current system allows for continuing involvement of anyone interested in nuclear power. This involvement provides those opportunities to educate themselves, mobilize opinion, observe their opponents, and monitor the Commission. Given the complexity of nuclear power and nuclear politics, the loss of these opportunities may be a substantial price to pay for the efficiency and predictability of standards.

Awkward applications can be avoided. An inherent problem with standards is their potential rigidity. For example, if interpreted literally, standards may require massive alterations of technologies that are only slightly out of line or may block them altogether. The safety goals are deliberately designed to avoid some awkward situations: The very approach is designed to avoid fixation on a single solution to problems. The benefit-cost guideline ensures that safety will not be bought at any price. The provisional character of the goals allows them to be adjusted in the trial period. One great potential cause for embarrassment is that the NRC may be unable to apply the safety goals it has promulgated. That will depend on the ability of the PRA approach to deliver the numbers needed to assess compliance. Although PRA has demonstrated its ability to assess the relative reliability of competing designs, assessing the absolute reliability of each of them may be another matter. Worse yet, there is the danger that using PRA to prove compliance will erode its credibility as a design tool. With or without justification, observers may note that the aggressive ferreting out of potential risks—a process that is needed for good design—conflicts with the desire to show that the design is in compliance. The difficulties that external reviewers will have with PRAs could compound these fears.

Applying these various criteria to the NRC's present effort, my conclusion is that, on balance, the use of goals does seem appropriate in the case of nuclear safety regulation. Properly designed goals can simplify the regulatory process, focus political discussions, clarify industry's obligations, and improve the allocation of safety dollars. In order to fulfill this promise, it is necessary (but not sufficient) to ensure that the standard-setting process does, in fact, resolve the political issues. It is also necessary to clarify the relationship of the goals to the existing regulatory processes.

One obstacle to the goals' success is their ambiguity on many key issues. In part, this reflects the need for additional work. In part, it reflects a conscious decision to leave the goals flexible, so as to avoid awkward complications and learn from experience. However, without some rigidity, the goals will prove meaningless. A second obstacle is the possibility that PRA cannot deliver the absolute risk estimates needed to evaluate compliance and, hence, make the rules practicable. These two obstacles occupy central places in the next two portions of the analysis, which consider how the goals are chosen and how they are meant to work.

HOW CAN THE COMMISSION'S SAFETY PHILOSOPHY BE JUSTIFIED?

Despite their name, the safety goals are not just about safety. Rather, they are the outward expression of a philosophy that, in some way, balances (or chooses to ignore) all of the varied consequences associated with nuclear power (and perhaps the consequences of alternative energy technologies as well). In choosing its approach to the problem of safety, the NRC might have taken any one of several different tacks, each reflecting a somewhat different philosophy.[9]

Formal Analysis

The Commission might have decided to adopt whatever goals proved most attractive after all possible goals have been evaluated by cost-benefit analysis, decision analysis, or some related technique.[10] Such formal analysis begins with an assessment of the expected consequences of adopting each possible safety goal. Those consequences would include both the good and the bad, and would be weighted by their probabilities. NRC would choose the goal with the highest expected worth. That choice could be justified by the inherent attractiveness of analysis: It is open, explicit, and systematic. The expected net worth criterion ensures the most efficient allocation of resources. Sophisticated techniques are available to help achieve it.

Unfortunately, however, analysis cannot always deliver on this potential. Moreover, this particular problem contains many of the features that are most awkward for formal analysis. To begin with, there would be disagreement about what options to evaluate; for example, should the Commission even consider goals that the industry could not meet for technical or economic reasons? There would also be disagreement over what consequences were relevant; for example, should the Commission consider the effect a given standard has on productivity or innovation? There would be disagreement, too, over how to compare diverse consequences, such as changes in reactor reliability and in the probability of premature death. It is one of the strengths of formal analysis that it lays bare issues such as these that might be glossed over by less systematic procedures. Nonetheless, that exposure may spark added dissent from those who dislike the conclusions of the analysis.

There are also problems arising from analysis itself. The foremost of those in the present context is neglecting equity issues. Nuclear power is controversial in part because the people who bear its risks are not always the people who reap its benefits. In such cases, the criterion of identifying "the greatest good for the greatest number" which underlies most analysis is unsatisfactory. A second inherent limit is that the formal analysis of decisions, like the engineering analysis of reactor designs, is a job for experts. As a result, there is no opportunity for the sort of public involvement needed to create and maintain an informed electorate. Moreover, unless all the parties involved can afford (or are provided with) independent analysts, some fear of hidden bias is likely.

To date, the Commission has declined to use formal analysis in setting the safety goals. Possibly, some (or even all) parties are doing independent analyses on the side, to see how alternative safety goals would affect them. Asking "What's in it for me?" is much more in keeping with the hedonistic purposes for which analysis was developed than is the formulation of public policy.

Professional Judgment

The greatest source of safety standards is the pooled judgment of technical experts. How technical experts reach such judgments varies widely. There are the unwritten norms of professional conduct, voluntary codes such as those promulgated by the American National Standards Institute, and binding rules issued by regulatory bodies. The putative advantage of relying on professionals' experience is that it leads to practical solutions that manage to balance different interests in a reasonable way.

Most of the NRC's current rules and regulations have been set on the basis of professional judgments. The continued functioning of the nuclear industry without major catastrophe or bankruptcy suggests that it has had some success. However, the safety goal problem is one that stretches professional judgment to its limits and threatens the viability of its product.

One problem is that for so complex a technology as nuclear power, experience provides rather vague lessons regarding the effect of any standard. Using professional judgment, it would be hard to establish that the plants were too safe (hence too costly), and hard to attribute an appropriate share of their safety to any particular standard. A second problem is that the nuclear industry's professionals have few opportunities to communicate with the public, one of the principal parties whose interests must be considered. Like other complex, expert-intensive industries, nuclear power concentrates its people, both geographically and socially. The general public is viewed primarily in adversarial contexts, or through the filter of the mass media.[11] A third problem is the conflict of interest facing professionals whose personal fortunes depend upon the continued viability of an industry. Can they create standards that may slow the industry's growth or stop it completely? The fourth problem is that each professional is specialized in some particular phase of the power plant's operations, and none has any special expertise regarding the safety level for power plants as a whole. Indeed, one of the main factors that led to promulgation of the safety goals was the feeling that the standards developed for each separate portion of the plant did not add up to a coherent safety philosophy.

Professionals must be part of the standard-setting process if its product is to be practical. However, it is hard for them to see the whole problem or to entertain all solutions. Despite that lack, the Commission seems to have been influenced by the safety goals proposed to it by different professionals.[12] The fact that these professionals agreed with one another may reflect their shared world view more than the convergence of independent sources of evidence.

Political Process When the views of experts are not altogether trusted, the obvious alternative is to invite the public to participate in the standard-setting process. To some extent, the NRC has done this by holding hearings and convening workshops that involve diverse participants. The commissioners have tried to synthesize these views, acting as surrogates for the elected officials who reconcile public views through bona fide political processes. If they have been successful, then the goals should gain the acceptance afforded the products of actual political processes. All those who were involved in the goals' formation, either directly or through representatives, and who feel that their views have been accommodated should form a cadre of active supporters. The goals should also be sounder for having weathered the critiques of different parties, each trying to eliminate weaknesses that are prejudicial to their own interests. It must be realized, however, that the strength of political processes lies in avoiding grossly wrong solutions rather than in identifying the right one.

Time will tell how successful the Commission has been. One thing that has increased its chances for success is that in developing the views of interested parties, it made extensive use of workshops in addition to the usual mandated hearings. The repeated meetings allowed the parties to see if their views were heard (and not just aired), to focus the discussion on value issues (rather than letting it drift to technical topics that are intellectually safer), to interact with the other parties (rather than relying on stereotypes), and to search for compromises. Furthermore, these public forums dealt with issues about which most participants had stable, articulated opinions. Many of the lay people present had worried as much about the value side of nuclear power as technical experts have worried about the fact side.

As a regulatory body, the NRC can do no more than simulate political processes. The quasipolitical solution that it has produced faces several threats: from those who feel that they were not represented; from those who feel that their representatives were present, but ignored; from those who feel that participants misrepresented their constituency (such as antinuclear spokespeople who claim to represent the general public, or nuclear spokespeople who claim to represent American industry); from those who feel that they lacked the resources to develop their case adequately. Any one of these claims can be argued plausibly by some interested group.

Revealed Preferences The NRC stated two of its goals in qualitative terms: to avoid significant additions to the risks to which society is already exposed; and to avoid risks that exceed those of competing electricity-generating technologies. These two goals incorporate a recurrent industry proposal,[13] namely, that nuclear power should be governed by the same safety philosophy as is applied to other technologies. Although the idea of a consistent safety philosophy seems quite reasonable, these goals entail two controversial (if common) assumptions.

One of these assumptions is that the risks associated with other technologies are socially acceptable, meaning that they reveal society's risk-bearing preferences. One reason for doubting that assumption, however, is the continued strong public support for stricter environmental regulations.[14] Another reason is that the public knows very little about the benefits and costs of many technologies, hence could not have given informed consent to them. A third is that some technologies have been in place so long that the choice of technology bears no relation to today's public preferences. A fourth is that many technologies function under conditions (e.g., near-monopoly) that leave them insensitive to public tastes.

Another controversial assumption is that risk levels alone capture the public's views on risk bearing. In practice, however, the level of risk seldom is an accurate reflection of public preferences, even when measured against economic benefit. The technologies that society adopts depend upon the choices it is offered. A safety philosophy can express itself in the choice of relatively risky technologies if there are no alternatives; but if there are cheap ways to reduce risk (or little risk to begin with), the same philosophy can lead to a choice of relatively safe technologies. Without a detailed analysis of these factors, it is hard to infer any safety philosophy (or "acceptable level of risk") from existing practices.

Some of these problems are avoided when the safety of nuclear power plants is compared not with all other technologies, but with "competing energy technologies." In comparisons of this sort, all risk can be expressed in terms of a common unit of benefit (electricity generated or saved). Such comparisons also have the advantage that if nuclear power proves to be the inferior choice, the superior one is explicitly identified. Nonetheless, it is still unclear how the diverse risks of the different energy technologies can be compared. How, for instance, does one equate the low probability of a high number of fatal cancers from nuclear power with the virtual certainty of a smaller number of accidental deaths from coal or solar energy?

Still another application of the revealed preferences approach is to assume that an acceptable level of risk for new plants can be determined by measuring them against the riskiness of licensed plants; this is an approach advanced by the major industry organization in the nuclear power field.[15] That proposal rests on the assumption that the policy followed by the commissioners in the past is socially acceptable today. That assumption is belied by the controversy surrounding nuclear power (and, indeed, by the safety goals themselves, which are a symptom of that controversy). The approach also assumes that the risks revealed by PRAs conducted today are identical to the risks perceived by the commissioners in making their licensing decisions in the past. However, it is in the nature of PRA that it not only changes the perception of risks, but changes the risks themselves by identifying needed changes in design.[16]

In general, the NRC's approach to setting the safety goals has been eclectic, drawing on a number of different perspectives. Although the Commission has not used formal analysis publicly, its actions are constrained by the private analyses of those who seek to influence its decisions. To the extent that the Commission's new safety goals articulate the philosophy implicit in the Commission's existing requirements, the safety goals express the professional judgment that shaped those requirements. The Commission's willingness to expose itself to a series of workshops in which the public participated reflects a willingness to be influenced by an essentially political process. And the comparison with other technologies implies some notion of revealed preferences.

Because the different methods lead to different conclusions, the choice of a method is a political act. To guide itself in that choice, the Commission would do well to ensure that the full potential (and limitations) of all these approaches is recognized. It might usefully conduct a rudimentary formal analysis of the goals' consequences, solicit opinions from professionals who have no vested interests to defend, hold public workshops to consider experience with the goals during the two-year trial period, and look harder at whether anything can be learned from the record of the past.

DO THE GOALS REFLECT THE SAFETY PHILOSOPHY?

Once the NRC has settled on its safety philosophy, it must still be translated into operational form. A poor translation may lead to a technology whose costs and risks are much different than what the Commission intended. An incomplete translation may leave too much discretion to those who apply the goals. Elsewhere I have identified the questions that need to be addressed in achieving a faithful translation from safety philosophy to operational standards.[17] These questions fall under four main heads: the range of the standard's application, the directness of the standard's application, the measurement of a technology's compliance, and the methods for enforcing compliance. Evaluating (or constructing) a standard requires a detailed technical analysis of all of these questions. The following describes how some of them reveal the meaning (or ambiguity) of the safety goals.

The Range of Application

In defining the range in which standards such as those of the NRC are to apply, the first question to be answered is: What technologies fall under its jurisdiction? On this point, the NRC's safety standards are very clear. The standards apply to all nuclear power plants in the United States. No special consideration is given to age, design, profitability, or any other factor that might affect a plant's ability to satisfy the goals; nor are exemptions possible on the grounds that plants have already met some other set of NRC regulations. Less clear is the range of competing technologies that are supposed to provide one basis for comparison. For example, does that comparison group include the technologies entailed in energy conservation or solar power? Nor is it clear whether the comparison is with current technologies or with such technologies ten years hence, when the nuclear plants that are now being planned will be operating.

The second application question is: How are those technologies to be defined? The Commission has also taken a clear position regarding the spatial definition of the technology being regulated. The goals apply to each individual power plant considered separately. Alternatively, the NRC might have applied the goals to groups of plants or to the industry as a whole. Grouping plants would allow safer plants to compensate for riskier ones; but it would also allow a few unsafe plants to take the industry down with them.

Regarding the temporal definition of the technology, the NRC's goals are clear, but controversial: Only the risks of accidents at the electricity-generating stage of the nuclear fuel cycle are covered; risks that might exist before or after that stage are not. This definition suggests that the risks of accident at the electricity-generating stage alone are being balanced against all benefits accruing from nuclear power. If other risks are substantial, then the operational goals allow more risk than is implied by the safety philosophy.

The application of the benefit-cost guidelines for safety improvements is precise in some respects, ambiguous in others. These guidelines clearly apply to all improvements that could be used to meet the overall safety goals. Risks and benefits are to be computed over the lifetime of the plant, using 1983 dollars adjusted for the general rate of inflation. What is not clear is what constitutes a "safety improvement." Ineffective changes may be adopted if they are lumped together with more effective ones. Effective changes may be neglected if they are separated from necessary concurrent changes (e.g., changing maintenance procedures, instructing personnel in those changes, paying performance bonuses to maintenance personnel). For the time being, the definition of "improvement" is left to the good sense and good will of those who apply the standard.

Directness of the Standards

Safety standards vary according to how directly they relate to the technology being judged. The most direct are design standards, specifying how a facility should be built and operated. The Commission has traditionally made use of such standards as its principal mode of operation. The great strength of design standards is in providing clear, predictable criteria for evaluating compliance; their weakness is in obscuring the amount of risk allowed. As a result, it is unclear what safety philosophy they express and why alternative designs are rejected.

By contrast, the Commission's two overall safety goals are defined in terms of ultimate consequences. Such standards are attractive because they deal with the aspect of safety performance that really interests most people: human health effects. They are unattractive because it is hard to establish compliance. Looking at the technology as planned or as operating, it is difficult to know how much death and illness it will produce. Looking at the health effects, such as cancer, it is difficult to link them with any particular source.

The benefit-cost and core meltdown guidelines avoid some of the problems and forfeit some of the benefits of standards that are based on ultimate health consequences. Without retreating all the way back to the regulation of the technology itself, these guidelines are nevertheless linked less directly to the safety objective. The reference to released radiation in the benefit-cost standard refers to an intermediate outcome that could have health effects. The core meltdown guideline regulates an initiating event that could lead to such an intermediate outcome.

By using standards that vary in their immediacy, the Commission hoped to balance the attractions and drawbacks of different approaches. The standards that are based on ultimate health consequences relate to events so infrequent that even a lifetime of safe operation could not demonstrate compliance; PRA is needed to assess those small risks. PRA's task is simplified as the standard moves closer to the technology. For example, with the core-melt guideline, there is no need to model containment probabilities, evacuation plans, or dose–response relationships. The practicality of this more proximal standard has its dangers, however; such standards may lead to a decline in the emphasis on achieving the ultimate health goals and may concentrate the Commission's attention once again on equipment characteristics. That shift in emphasis could also reduce interest in measures that would mitigate the effects of accidents after they occurred.

Mode of Measurement Once it has been decided what to measure in order to assess compliance, a scheme is needed for how to measure it. Here, in particular, attention to detail is essential to creating a useful standard. A sampling of these issues follows, as they arise in the design of the safety goals.[18]

A first question is how many features of the technology to measure. Whereas existing design standards looked at many aspects of the technology and its operation, the Commission's philosophy considers tradeoffs among three summary characteristics: costs, risks, and benefits. The goals themselves, however, deal only with risks. Presumably, benefits and costs are so well known and so stable that they require no monitoring. If this is not the case, measuring risks is not enough. Possibly more controversial is ignoring the benefits to individuals. Whether individuals bear a "significant additional risk" (in the words of the first overall goal) should depend on the benefit they gain. The Commission hopes to make the individuals feel that the risks that they bear personally are negligible, and hence require no compensating benefit. The risks that the Commission's standards refer to are prompt fatalities and cancer to people near the plants due to accidents. But if other health effects are large, such as birth defects, or stress-related illness due to concern about reactor accidents, then the risks are being underestimated.

Whenever several safety or health features are measured, a rule is needed to determine how the various results are to be used in the final judgment of whether there is compliance. Must all features be in compliance or can more-than-adequate performance in some respects compensate for inadequate performance in others? The safety goals appear to be noncompensatory; all must be met for a plant to be satisfactory, as must all of the design standards. Thus, the Commission has adopted the most stringent rule possible.

Whenever the features of a technology vary over time or space, some scheme is needed for characterizing the technology. Recognizing the variability in plants, the NRC's goals call for separate analyses of each. That specification can only be faulted on grounds of expense and the uncertainty that will cloud the industry until all of the analyses have been completed.[19] Recognizing that the riskiness of plants increases whenever problems arise and decreases whenever they are solved, the Commission specifies that the average risk during the period of a year will be its reference measure. If the same risk limits were applied to a shorter period of observation, such as a day or a week, then plants would be more likely to lose their licenses. One dimension of variability that seems to be ignored is the changes occurring in a plant over its life span. In the course of a 30–40-year life span, a plant might perform quite differently while breaking in, wearing down, or responding to shifting regulatory pressures.

A strong point of the safety goals is that they recognize that a PRA cannot accommodate all available information. For example, a part may have a distribution of failure rates, or experts may disagree about the likelihood of an operator error. PRAs are to use "best estimate or mean-value analyses [which] should include an understandable presentation of the magnitude and nature of the uncertainties."[20] Thus, the plant as a whole would be characterized by a single "best estimate" of its risk. An alternative approach would be to sample different values for these components in order to create a distribution of overall risk estimates.

Less clear is what should be done with evidence that does not lend itself to incorporation in PRA models. Human behavior falls into this category. Its omission could lead to the overestimation or underestimation of risk, depending upon whether the human factor creates or solves more problems.

Methods of Enforcement

Finally, a meaningful standard must be linked to action. Although the NRC's safety goals specify how compliance is to be measured, they do not specify what should be done with the results. In part, this is due to the deliberately tentative nature of the safety goals. In part, it is due to uncertainty about whether the goals are meant to have teeth. One reason for that uncertainty is the possibility that the goals may prove ambiguous in critical respects. Any ambiguity in a standard reduces its enforceability by allowing those to whom it is applied to argue for interpretations that are congenial to their interests. As shown above, the goals are specific in many, but by no means all, details. Perhaps more significant than the existence of ambiguities is the lack of a clear mechanism for resolving them. If interpretative problems are worked out in the course of applying the goals, then outsiders will fear that some cozy arrangements have been made between the Commission staff and the industry. If, at the other extreme, all questions were referred back to the commissioners, the process would prove cumbersome and costly. Given the tentative character of the goals, it may be enough to identify the ambiguities that arise in the two-year trial period and to expose them to the workshop and hearings process that preceded the adoption of the goals.

Although NRC's statement of the goals acknowledges the uncertainty of the data going into the PRAs, it does not provide any guidance for the administrators who must act, without waiting for certitude. The interpretation given to uncertain evidence should depend on where the burden of proof lies: whether plants should be required to prove their compliance, or administrators to prove noncompliance. That in turn depends on whether the NRC prefers to expose itself to the error that goes with closing plants that should be operating, or to the error that goes with keeping plants open that should be closed. If its staff is not to set social policy, then the commissioners must lay down some enforcement rule, such as: "A plant is in compliance if there is at least an X% chance that its risk is below the allowable maximum."

A final context in which the staff needs guidance in coping with uncertainty is how to respond to new evidence showing that a plant that was licensed in the past (on the basis of what was known then) was, in fact, too risky (on the basis of what is known now). If approval can be rescinded in such cases then there will be a large measure of residual unpredictability in the application of the standards. If it cannot be rescinded, then operators who are aware that their plants are risky will prefer fast and sloppy PRAs in the hope that they can slip by.

However the goals and evidence may be interpreted, the NRC staff will be faced with making some summary assessment of whether a plant is in compliance. Regarding the rigor with which that conclusion will be enforced, the NRC says only that "the necessary action should be taken within a time commensurate with the increased risk involved."[22] The reason for this vagueness may be found in an introductory statement which notes that "... current regulatory methods are believed to meet the basic statutory requirement that there be adequate protection of the public. Nevertheless, current practice could be improved to provide a better means for enunciating NRC decisions and for testing the need for an adequacy of regulatory requirements." Thus, the safety goals themselves may have purely heuristic value. Indeed, the term "goal" suggests an aspiration more than a requirement. Nowhere is it said that a plant could actually be shut down for noncompliance.

CONCLUSION

The safety goals for nuclear power announced by NRC represent a significant reorientation of its regulatory policy. The articulation of such goals could serve a salutary purpose in the Commission's regulatory framework. The goals could respond to the public's need for a clear statement of the Commission's philosophy, the industry's need for more predictable and flexible demands, and the staff's need for a way to accommodate PRA in the regulatory process. However, in their current advisory role, the goals may become just another encumbrance in an overburdened regulatory process.

The Commission has adopted an eclectic approach in laying the basis for its goals and attendant standards. It has listened to a variety of individuals and has incorporated arguments generated by a variety of methods. The eclecticism enhances the goals' political acceptability, at the expense of reducing their coherence. Thus, although some of the provisions incorporated in the NRC's goals and standards are attractive in themselves, the logical interrelationships between the various provisions are unclear. Moreover, the NRC is silent on critical aspects of its new goals and standards, notably on the issue of enforcement.

As an example of how safety standards should be framed, developed, and applied, the NRC exercise represents an important and informative case study. Simple lessons cannot be derived easily from the case. But it does serve to highlight some of the obvious gaps and critical ambiguities that tend to imperil the effectiveness of regulatory efforts of this kind.

I would like to thank S. A. Carnes, T. E. Cole, E. D. Copenhaver, H. Inhaber, H. Kunreuther, G. Morgan, D. C. Parzyck, P. Rathbun, P. Slovic, K. A. Solomon, S. Watson, and two anonymous reviewers for their many helpful comments on previous drafts.

BARUCH FISCHHOFF is a Research Associate with Decision Research, A Branch of Perceptronics, 1201 Oak, Eugene, Oregon.

NOTES

1. U.S. Nuclear Regulatory Commission, *Safety goals for nuclear power plants: A discussion paper*, NUREG-0880 (Washington, DC: The Commission, 1982).
2. For example, the Federal Aviation Administration has recently followed the NRC's lead by using safety goals in licensing the 767.
3. For a more detailed exposition of this approach, see Fischhoff, B.,"A theory of safety standards," *Management Science*, in press.
4. Greene, A. E., and Bourne, A. J., *Reliability technology* (New York: Wiley-Interscience, 1972); U.S. Nuclear Regulatory Commission, *PRA procedures guide*, NUREG/CR-2300 (Washington, DC: The Commission, 1981); Veseley, W. E., Goldberg, F. F., Roberts, N. H., and Haasl, D.F., *Fault tree handbook*, NUREG-04492 (Washington, DC: U.S. Nuclear Regulatory Commission, 1981).
5. Actually, the Commission says that "while this policy statement includes the risks of normal operation as well as accidents, the Commission believes that risks from routine emissions are small and therefore does not believe that they need to be routinely analyzed on a case-by-case basis in order to demonstrate conformance with the safety goals." Hence, the goals are treated here as dealing with accident risks.
6. W. C. Wood provides an intriguing account of the possible origins and implications of this figure: "Putting a price on radiation," *Journal of Policy Analysis and Management, 2* (1982): 291–295.
7. Screening potential ways to reach compliance is a modest role for this rule. In earlier drafts, the benefit-cost guideline specified when improvements should be adopted for plants that are already in compliance with the overall design objectives. The role of deciding when to adopt extra safety improvements is now left to the Commission's existing backfitting regulation (10 CFR 50, 109).
8. Fischhoff, *op. cit.*
9. A more complete discussion of the approaches can be found in Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L., and Keeney, R. L., *Acceptable risk* (New York: Cambridge University Press, 1981).
10. See, for instance, Keeney, R. L., *Issues in evaluating alternative standards* (San Francisco: Woodward-Clyde Associates, 1981).
11. See, for example, Fischhoff, B., Slovic, P., and Lichtenstein, S., "Lay foibles and expert fables in judgments about risk," *American Statistician, 36* (1982): 240–255.
12. See Appendix 2 to U.S. Nuclear Regulatory Commission, 1982, *op. cit.*
13. Atomic Industrial Forum, *A proposed approach to the establishment and*

use of quantitative safety goals in the nuclear regulatory process (New York: The Forum, 1981).

14. "A call for tougher—not weaker—antipollution laws," *Business Week*, January 27, 1983, p. 82.
15. Atomic Industrial Forum, *Committee on reactor licensing and safety statement on licensing reform* (New York: The Forum, 1976).
16. "Nuclear risk update," *Engineering News-Record*, July 15, 1982, pp. 35–36.
17. Fischhoff, *op. cit.*
18. Fischhoff, *op. cit.*, provides further detail.
19. Although one might wonder what happens until all of the analyses have been completed. If the best managed plants are also the first analyzed, then riskier plants may get an undue lease on life. The converse is true if the most troublesome plants are analyzed first.
20. U.S. NRC, 1982, *op. cit.*, p. 33.
21. A puzzling statement regarding refinement is that the core-melt "guideline will need to be revised as new knowledge and understanding of core performance under degraded cooling conditions are acquired" *(ibid.*, p. xiv). Although the way a plant is evaluated should be sensitive to changes in knowledge, the standard itself is a political statement which should be sensitive only to changes in tastes.
22. U.S. NRC, 1982, *op. cit.*